Vienna 2015

ABSTRACT BOOK

CONTENTS

PLENARY TALKS 3

FITCH SYMPOSIUM 7

1 Beyond the Equilibrium Paradigm: The role of temporal processes in population genetics and evolution 15

2 Untangling information, noise, and phylogenetic reconstruction in genome scale data 53

3 The biological impact of transposable elements 102

4 Evolution and Ecology of microbial communities 137

5 Open Symposium 192

7 Origins and evolution of molecular innovation 396

8 The horizontal component of microbial evolution 465

9 Inference of demography and selection under violations of the Kingman coalescent assumptions 498

10 Speciation genomics 507

11 PopGen in space! Theory and inference in spatial population genetics 563

12 Micro-evo-devo: using natural variation to explain the how and why of phenotypic evolution 614

15 The evolution of alternative splicing 676

16 Exploring the consequences of ancient and contemporary gene flow 695

17 Genomics of sex bias: Addressing questions with or without genomes 733

18 Within- and between-host viral evolution 757

19 Novel functional approaches to evolutionary genomics 792

20 Evolution of molecular pathways and networks: Molecular evolution meets systems biology 807

21 Fungal evolutionary genomics: Unravelling mysteries from the Forgotten Kingdom 851

22 Adaptive and non-adaptive evolution of gene expression and regulation 875

23 Reframing the demography vs. selection debate using 21st century models and data 971

24 Short Tandem Repeats in the post-genomics age: Accurate typing, variability, evolution, and function 1004

25 Ancient genomes: A Time machine for investigating natural selection 1016

26 Genomic and epigenomic evolution of sex chromosomes: Broad patterns and intriguing cases 1054

27 The origins of multicellularity under the light of functional genomics 1076

28 The golden age of Archaea: unveiling the diversity and evolution of the third Domain of Life 1087

POSTER INDEX 1101

SPEAKER AND AUTHOR INDEX 1102

PLENARY TALKS

Pl.1

How natural selection may connect evolutionary biology to physics

Joseph Felsenstein
University of Washington, Seattle, Washington, USA

In this talk I emphasize the centrality of natural selection in evolutionary biology, and the way it connects evolutionary biology to physics. Not the ubiquity of natural selection, as I am not a panselectionist, but the necessity of natural selection if we are to explain how species can be as well-adapted as they are. I will explain how Leslie Orgel's concept of specified information makes this argument. Although that concept has been picked up and emphasized by advocates of Intelligent Design as an argument against the power of natural selection, the invalidity of their argument does not invalidate Orgel's original insight. Furthermore, natural selection appears essential to giving a quantitative account of the evolution of flows of energy through biological systems. Although biologists are averse to hazy and grandiose claims that we can describe of evolution in thermodynamic terms, we ought to make the attempt. Such attempts seem to have started with someone born here in Vienna, Ludwig Boltzmann. It seems that extremely simplified models will be the way to start such an effort. I made such an attempt in 1978, in my least-cited paper, and I will describe that in hopes of harvesting a few more citations. In my model one can predict to what extent an evolving ecosystem can incorporate energy which is on its way to dispersing. (I wish to reassure potential audience members that no actual molecular sequences were harmed, or even considered, in this work).

Pl.2

De novo evolution of genes

Diethard Tautz
Max-Planck Institute for Evolutionary Biology, Plön, Germany

The notion that new genes (and proteins) evolve via duplication and divergence mechanisms has been around since the beginning of the last century. In 1977 Francois Jacob wrote his famous paper on "tinkering" where he considered specifically the question of whether new genes could alternatively emerge *de novo* out of non-coding DNA. He concluded with the often cited statement: "... creation of entirely new nucleotide sequences could not be of any importance in the production of new information." Of course, he acknowledged that genes and proteins must have come from somewhere, but he relegated this to a pre-biotic phase: "The really creative part of biochemistry must have occurred very early". On the other hand, the genome projects of the past decades have turned out a large number of "orphan" genes that appear to have no similarity to any previously known genes. This has raised the question whether new genes and proteins can arise *de novo* after all. There is now increasingly good evidence that this is indeed not only possible, but that *de novo* evolution is actually an extremely active process in any genome. I will discuss this evidence in my talk and address the question of what this means for our understanding of the evolutionary role of the non-coding part of the genome.

Pl.3

Genomics of Rapid Adaptation

Dmitri Petrov
Department of Biology, Stanford University, California, USA

tba

Pl.4

The biology of repetitive DNA: adaptation, epigenetic silencing and aging

Doris Bachtrog
Department of Integrative Biology Center for Theoretical Evolutionary Genomics, University of California, Berkeley, USA

Repetitive DNA is a major component of many eukaryotic genomes, yet its general role in evolution is controversial. In my talk, I will discuss three different consequences of repetitive DNA to organismal fitness, using empirical data from Drosophila. Mobile repetitive elements can contribute to the adaptive rewiring of regulatory networks, by supplying ready-to-use regulatory elements to many locations in the genome, and non-allelic gene conversion can spread beneficial mutations across different insertion sites. Furthermore, host genomes have evolved mechanisms to silence repetitive DNA through epigenetic modifications, and spreading of these epigenetic marks to nearby genes can silence their gene expression. Finally, silencing marks at repetitive DNA can get lost during an individuals life span, and contribute to aging.

FITCH SYMPOSIUM

F.1

Inferring population structure across space and time

Gideon Bradburd [1], Peter Ralph[2], Graham Coop[1]
[1] UC Davis, Davis, CA, USA, [2] USC, Los Angeles, CA, USA

Population genetic structure can be continuous, the result of isolation by distance or drift acting through time, or more discrete, the result of population replacement, or geographic, physical, or behavioral barriers to gene flow between populations. The continuous decay of relatedness with temporal and geographic distance is a natural null model against which hypotheses of more discrete population structure can be tested. Here, we present a method in which we model the decay of relatedness within a population as a continuous function of isolation across space and through time, and capture more discrete structure by allowing sampled individuals to choose membership across multiple populations. Specifically, we model populations as independent spatiotemporal Gaussian processes from which individuals draw different amounts of ancestry. We develop a method to estimate the assignments of the ancestry proportions for each individual, as well as the parameters that govern the spatiotemporal decay of relatedness in each population.

This model explicitly addresses the "clines vs. clusters" problem in quantifying population structure by jointly accommodating both continual and discrete patterns of differentiation. The model also naturally captures population replacement, a phenomenon for which there is substantial evidence in humans from archaeological evidence and ancient DNA. We demonstrate the utility of this approach using a combination of ancient and modern human individuals sampled throughout Europe.

F.2

Watching the dynamics of molecular evolution over 60,000 generations

Benjamin Good [1], Michael McDonald[1], Richard Lenski[2], Michael Desai[1]
[1] *Harvard University, Cambridge, MA, USA,* [2] *Michigan State University, East Lansing, MI, USA*

Predicting the long-term dynamics of evolution in a fixed environment is a fundamental problem in evolutionary biology. A quantitative characterization of this simple scenario is vital for understanding more complex topics like the architecture of the fitness landscape, the role of environmental change, or the formation and maintenance of ecology. But despite extensive theoretical work, progress is limited by the paucity of direct empirical data with which to test these evolutionary models. Here, we help to fill this void by constructing a high-resolution "molecular fossil record" documenting 60,000 generations of evolution in experimental populations of *E. coli*. We find that, while fitness gains have dramatically slowed through time, molecular evolution proceeds at a rapid pace even after tens of thousands of generations, with rampant interference between competing beneficial mutations. Time-resolved mutation trajectories also reveal previously unobserved cases of competitive coexistence, which are stable in the short term but subject to evolutionary perturbation on longer timescales. These findings highlight the use of genetic data to uncover aspects of adaptation that are difficult to observe phenotypically, and they suggest that long-term adaptation to a constant environment may be more complex and dynamic than is often supposed.

F.3

Evolution of dosage compensation in the dioecious plant *Silene latifolia*

Aline Muyle [1], Niklaus Zemp[2], Cécile Fruchard[1], Alex Widmer[2], Gabriel Marais[1]
[1] Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France, [2] ETH Zurich, Institute of Integrative Biology, Zürich, Switzerland

p { margin-bottom: 0.25cm; direction: ltr; color: rgb(0, 0, 10); line-height: 120%; text-align: left; }p.western { font-family: "Liberation Serif",serif; font-size: 12pt; }p.cjk { font-family: "Droid Sans Fallback"; font-size: 12pt; }p.ctl { font-family: "FreeSans"; font-size: 12pt; }a:link {It is not clear how similar the evolutionary patterns of plant and animal sex chromosomes are. In particular, the existence of dosage compensation in plants is controversial. Moreover, very few sex chromosome sequences are available in plants. Here, we developed a model-based probabilistic method to infer sex-linkage from RNA-seq data of a family with much greater confidence than in previous work. It allowed us to identify ~1000 X/Y gene pairs and ~300 X-hemizygous genes (without Y copy) in the plant species *Silene latifolia*. We used an outgroup without sex chromosomes, *Silene vulgaris*, to study the evolution of X and Y-allele expression. Our data support the existence of dosage compensation in *S. latifolia*. We found evidence that, in males, expression of the X has increased when Y expression decreased, but this increase in X expression, surprisingly, also happened in females. Dosage compensation is stronger in highly expressed genes, and in flower buds and seedlings compared to rosette leaves. X-hemizygous genes are rarely dosage compensated, suggesting that selection of dosage compensation is weaker for those genes, explaining the discrepancy with previous studies. We also found some evidence for X-inactivation in this plant. Overall, our results support Ohno's view on dosage compensation – X over-expression in both sexes and X-inactivation in females – in a plant lineage.

F.4

Human epigenomic variation is driven by historical and recent changes in habitat and lifestyle

Fagny Maud [1,2], Patin Etienne[1], MacIsaac Julia[3], Rotival Maxime[1], Timothée Flutre[4], Meaghan Jones[3], Quach Hélène[1], Harmant Christine[1], McEwen Lisa[3], Froment Alain[5], Evelyne Heyer[6], Antoine Gessain[7], Edouard Betsem[8], Patrick Mouguiama-Daouda[9], Jean-Marie Hombert[10], George Perry[11], Luis Barreiro[12], Michael Kobor[3], Lluis Quintana-Murci[1]

[1] Institut Pasteur - CNRS URA3012, Paris, France, [2] Université Pierre et Marie Curie, Cellule Pasteur UPMC, Paris, France, [3] University of British Columbia, Vancouver, BC, Canada, [4] INRA - UMR AGAP, Montpellier, France, [5] MNHM - IRD - UMR208, Paris, France, [6] niversité Paris Diderot - MNHN - UMR7206, Paris, France, [7] Institut Pasteur, Paris, France, [8] University of Yaoundé I - Faculty of Medicine and Biomedical Sciences, Yaoundé, Cameroon, [9] Université Omar Bongo - LCC, Libreville, Gabon, [10] Université Lumière-Lyon 2 - CNRS UMR 5596, Lyon, France, [11] Pennsylvania State University - Departments of Anthropology and Biology, University Park, PA, USA, [12] Université de Montréal - Centre de Recherche CHU Sainte-Justine, Montréal, Canada

The epigenomic landscape of the human genome, in particular DNA methylation, is increasingly recognised as an important driver of phenotypic diversity. However, the relative impacts of DNA sequence variation and temporal changes in lifestyle and ecological habitat on the epigenome remain unknown. Here, we generated whole-blood genome-wide DNA methylation and genetic profiles for 352 individuals from 5 populations of African rainforest hunter-gatherers and sedentary farmers differing in their present or historical lifestyles and habitats. We found that historical and current differences in lifestyle and habitat have similarly profound impacts on the global methylome. However, the biological functions affected and the mechanisms underlying DNA methylation variation differed strongly. Methylation variation between populations living in different habitats but sharing the same historical lifestyle and genetic background mostly involved genes related to immune system functions. Conversely, methylation variation between populations that share the same current environment but differ in their historical lifestyle and genetic background involved genes related to developmental processes. Importantly, methylation variation due to historical differences was strongly enriched in associations with nearby SNPs (meQTLs), contrary to variation due to current differences in environment, which was depleted in such associations. Moreover, meQTLs associated with historical differences in lifestyle explain a larger part of the interindividual variance in methylation levels than expected, and have been privileged targets of natural selection. Together, these findings provide new insight into the contribution of epigenetic modifications and DNA sequence variation to the adaptation of human populations to changes in lifestyle and environment over different time scales.

F.5

Insights into recombination and sex chromosome evolution from whole-genome sequencing of platypus

Hilary Martin [1], Elizabeth Batty[1], Julie Hussin[1], Portia Westall[2], Tasman Daish[3], Rory Bowden[1], Tom Grant[4], Frank Grutzner[3], Jaime Gongora[2], Peter Donnelly[1,5]

[1] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, [2] Faculty of Veterinary Science, University of Sydney, Sydney, Australia, [3] School of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia, [4] University of New South Wales, Sydney, Australia, [5] Department of Statistics, University of Oxford, Oxford, UK

As a member of the most basal mammalian group, the platypus is ideal for studying the evolution of different biological processes. We have sequenced 58 platypus samples from 14 river systems from across the species' range between 16°S and 43°S in eastern Australia. This is the first whole-genome sequencing project to look at diversity in this species. One of the most interesting aspects of platypus biology is the sex chromosome system: there are five X and five Y chromosomes which, in male meiosis, form a multivalent chain connected by nine pseudoautosomal regions (PARs). The Y chromosomes are missing from the reference genome, which was made from a female. However, we can recover Y sequences by *de novo* assembly of male-specific reads and by examining sequence divergences where they map to the reference. Through the latter, we find evidence for Y-specific sequences similar to their gametologs on the Xs. These regions appear to range from a few hundred base pairs to over 100kb and to localize at PAR boundaries. Some seem to be segregating between or within river systems, possibly suggesting that the divergence of X and Y chromosomes is ongoing. In addition, we are investigating the population structure and history of the species, and estimating fine-scale recombination rates. We cannot find clear evidence that platypus have *PRDM9*, the product of which controls hotspot placement in primates and mice, and we anticipate that this study will shed light on the evolution of the important process of meiotic recombination in mammals.

F.6

Functional genetic analysis of stickleback craniofacial evolution

Priscilla Erickson, Phillip Cleves, Jiyeon Baek, Craig Miller
University of California-Berkeley, Berkeley, California, USA

Understanding the genetic and developmental basis of morphological diversity is a primary goal of evolutionary biology. The genomic resources and phenotypic diversity of the threespine stickleback system provide an excellent model to study the genetic basis of morphological evolution. Marine and freshwater sticklebacks are adapted to different diets and have striking differences in the food-processing structures of the branchial skeleton: freshwater fish have more teeth and longer branchial bones due to both early and late developmental differences. Overlapping quantitative trait loci (QTL) on chromosome 21 control these skeletal differences in multiple freshwater populations, suggesting either a single pleiotropic locus or closely linked loci control these traits. The tooth number QTL maps to a small genomic region containing a cis-regulatory allele of *Bmp6,* with the freshwater allele having reduced expression. A short, 190 bp upstream enhancer is required for *Bmp6* expression in teeth. While a TALEN-generated coding mutation in *Bmp6* reduces tooth number in heterozygotes and is lethal in homozygotes, fish heterozygous or homozygous for a loss-of-function mutation in the enhancer are viable and have more teeth, mirroring the evolved phenotype. However, recombinant chromosomes exclude the coding region of *Bmp6* from the bone length QTL. Ongoing studies are functionally testing the roles of *Bmp6*, and a nearby developmental regulatory gene, *TFAP2a*, in bone length evolution using loss- and gain-of-function approaches. Combined, these results suggest multiple tightly linked loci control different skeletal adaptations and that *cis*-regulatory changes can produce evolved phenotypes while avoiding deleterious pleiotropic effects.

F.7

The dynamics of adaptive and antigenic evolution of influenza B virus

Pinky Langat [1,2], Simon Watson[1], Jayna Raghwani[3], Gytis Dudas[4], Colin Russell[2], Rod Daniels[5], Andrew Rambaut[4], Oliver Pybus[3], John McCauley[5], Paul Kellam[1]

[1] Wellcome Trust Sanger Institute, Hinxton, UK, [2] University of Cambridge, Cambridge, UK, [3] University of Oxford, Oxford, UK, [4] University of Edinburgh, Edinburgh, UK, [5] MRC National Institute for Medical Research, London, UK

Influenza B virus (IBV) causes nearly one third of the global influenza burden, accounting for significant morbidity, hospitalizations, and epidemics in many parts of the world. Despite this, the evolutionary dynamics of the two co-circulating phylogenetically- and antigenically-distinct lineages, B/Victoria and B/Yamagata, are not well understood, particularly compared to influenza A. A history of low cross-reactivity between the predominant circulating virus and the chosen vaccine strain each year illustrates the difficulty in vaccine strain selection for IBV. Therefore a better understanding of the complex relationship between viral genotype and antigenic phenotype is necessary. Here we present a comprehensive analysis of 2,165 complete genomes sampled globally between 1987 and 2014, including 425 isolates newly-sequenced in this study. We use Bayesian approaches to estimate evolutionary rates of the viral genes and compare patterns of genealogical diversity between the two lineages across the world. By quantifying the rates of adaptive evolution across the genome, we observe patterns of adaptation in the major glycoproteins over time. Additionally, by integrating these phylogenies with antigenic data (haemagglutination inhibition data for 737 isolates), we identify key substitutions that result in major antigenic drift for each IBV lineage. Overall, our study provides insight into how interactions of influenza B viruses at the epidemiological scale shape processes of molecular and antigenic evolution.

F.8

Epistasis Shaped the Evolutionary Sequence Space of an Ancestral Transcription Factor and its DNA Regulatory Elements

Alesia McKeown [1], Dave Anderson[1], Joseph Thornton[1,2]
[1] *University of Oregon, Eugene, OR, USA,* [2] *University of Chicago, Chicago, IL, USA*

Gene expression is regulated by interactions between transcription factors (TFs) and the DNA regulatory elements (REs) to which they bind. We sought to understand how the genetic and biochemical determinants of TF/RE binding shaped the function and evolution of the TF/RE complex. We measured binding affinity of all combinations of TF/RE genotypes across the sequence space defining a historical transition in RE recognition by a biologically important TF. We statistically identified the main and epistatic effects of sequence states within and between ancestral TFs and their RE targets, mapped these functions onto sequence space, and characterized the implications for the evolution of regulatory function under various evolutionary scenarios.

We found that epistasis within and between the TF and RE strongly shaped the complex's sequence space and the evolutionary trajectories through it. Epistasis between nucleotides in the RE played a key role in determining the DNA-specificity of each TF. Additional inter-molecular epistasis between the TF and RE permitted historical replacements in the TF to change the set of RE sequences specifically recognized by the evolving TF. Mapping of these interactions onto sequence space revealed that only a small fraction of mutational pathways across the space could have produced a TF with novel RE specificity without traversing nonfunctional intermediates. All of these pathways involved passing through a promiscuous intermediate TF whose specificity was then narrowed. Additional paths existed by which both the ancestral TF and RE target could drift through sequence space together, maintaining their specific interaction, even as both molecules diverged.

1 Beyond the Equilibrium Paradigm: The role of temporal processes in population genetics and evolution

1.1

Adaptive immunity and predictability of evolution in commensal gut bacteria

João Barroso-Batista, Jocelyne Demengeot, Isabel Gordo
Instituto Gulbenkian de Ciência, Oeiras, Lisbon, Portugal

A co-evolutionary process between the host immune system and the microbes that inhabit its gut is thought to be key in shaping the astonishing diversity of the mammalian gut microbiome. A corollary of this hypothesis is that the adaptive immune system will, directly or indirectly, influence the evolution of any given commensal species. We have followed the evolution of a commensal strain of *Escherichia coli* when colonizing healthy or immuno-compromised Rag2$^{-/-}$ mice, which lack T and B cells. We show that the pace of bacterial adaptation depends on the presence of an adaptive immune system and is lower in immuno-compromised animals. This is not due to differences in mutation rate or generation time. Natural selection is found to be the main mechanism responsible for the differences in the evolutionary dynamics of *E. coli* between host types. Emerging mutations are found to exhibit strong beneficial effects in healthy hosts but substantial antagonistic pleiotropy in immuno-deficient mice. The genetic basis of *E. coli* adaptation to colonize the gut is dependent on host genetic background. Thus, the adaptive immune system influences both the *tempo* and the *mode* of *E. coli* adaptation to the mouse gut. The results further suggest that the adaptive immune system enhances the predictability of adaptive evolution of bacteria comprising the microbiota.

1.2

Characterising the within-host fitness landscape of the influenza virus using time-resolved sequence data

Chris Illingworth
University of Cambridge, Cambridge, UK

The influenza virus is capable of rapid evolution in response to selective pressure, with the potential for significant evolutionary change within the timescale of a single infection. We here describe a method for inferring the primary components of the fitness landscape acting upon an influenza virus during within-host growth, based upon time-resolved short-read sequence data; our method uses the partial haplotype data captured by paired-end reads to account for linkage disequilibrium between alleles. Applying our method to data from an evolutionary experiment conducted in ferrets using a reassortant H5N1 virus, we characterise the fitness landscape of the virus, identifying alleles showing evidence of evolving under selection, and epistatic effects.

1.3

Evolutionary dynamics of an optimal adaptive immune response

Tobias Lenz
Max Planck Institute for Evolutionary Biology, Ploen, Germany

The adaptive immune system in vertebrates has evolved to fight novel and coevolving parasites and pathogens. However, the most optimal immune response requires a delicate balance of maximizing recognition and inhibition of invading pathogens while minimizing damage to own tissue by the activated immune machinery. The major histocompatibility complex (MHC) plays a key role in this process. MHC molecules present self and non-self antigenic peptides to immune effector cells and depending on which peptides are presented they can trigger either pathogen resistance or autoimmunity. This qualitative and quantitative balance is also reflected at the genetic level; while MHC genes are among the most polymorphic genes at the population level, each individual carries only a small subset of this diversity. Here we first present population genetic data from non-human species (fish, mammals) supporting pathogen-mediated selection for increased individual MHC diversity. We then show new results from large-scale autoimmune datasets in humans, indicating contrasting selection through an elevated autoimmune disease risk for more diverse MHC genotypes. This novel evidence supports theoretical models of antagonistic selective pressures and may explain the observed limited individual genetic diversity in the MHC. Eventually, we show data from natural populations suggesting that the optimal level of individual MHC diversity may be dynamic, responding to changes in the relative strength of pathogen-mediated selection in time and space. These insights are relevant in the face of climatic changes, likely affecting pathogen communities, but also in the context of the increasing prevalence of many autoimmune diseases in human populations.

1.4

Risk and reward of high mutation rate: why large populations favor mutators while small populations inhibit them.

Scott Wylie [1], Yevgeniy Raynes[1], Paul Sniegowski[2], Daniel Weinreich[1]
[1] Brown University, Providence, RI, USA, [2] University of Pennsylvania, Philadelphia, PA, USA

The error rate of genomic duplication is a fundamental evolutionary parameter influenced by enzymes for DNA proofreading and repair, which can themselves be altered by mutations. Some allelic variants of those enzymes, known as mutators, increase the genomic rate of mutation and play a major role in laboratory evolution, pathogens, and cancer. Even when mutators have no direct effect on survival and reproduction, mutation rate can evolve under *indirect selection*: Particularly in asexual populations, mutators become physically/genetically linked to beneficial and deleterious mutations that occur elsewhere in the genome.

We report the novel and striking phenomenon that the *sign* of selection of mutators depends on population size (N). Specifically, in Wright-Fisher competitions with non-mutators, mutators "win" in large populations yet "lose" in small populations. Using diffusion equations, we calculate the fixation probability of rare mutators in terms of beneficial/deleterious mutation rates, the strength of beneficial mutations, and N. Our solutions imply a simple yet powerful risk-reward principle linking N with the payout of selective sweeps. We generalize our findings for Wright-Fisher populations by studying mutator fixation in temporally and spatially structured populations, where population size is not conceptually clear-cut. We discuss our theoretical predictions in terms of previous and ongoing laboratory experiments with yeast and *E. coli.* Finally, we discuss the applicability of our risk-reward principle to other cases of indirect selection, such as the evolution of recombination and time-dependent selection. `

1.5

Genomic variation across space and time in *Drosophila simulans* and *D. melanogaster*

Heather Machado [1], Alan Bergland[1], Emily Behrman[2], Katherine O'Brien[3], Paul Schmidt[2], Dmitri Petrov[1]

[1] Stanford University, Stanford, CA, USA, [2] University of Pennsylvania, Philadelphia, PA, USA, [3] University of Nebraska–Lincoln, Lincoln, NE, USA

One of the most exciting horizons for the study of evolution is the mass measurement of genomic changes in natural populations across long stretches of space and time. Here we present a unique dataset examining how genomic variation in *Drosophila simulans* and *D. melanogaster* varies over a ~2000 km latitudinal cline and across multiple years. By using both spatial and temporal data we are able to identify the genetic variation that consistently increases or decreases across the cline. We observe that *D. melanogaster* shows a strong pattern of isolation by distance, which is highly conserved across time. In *D. simulans* the opposite is true: there is only a weak pattern of isolation by distance and this pattern fluctuates over time, suggesting a key role for migration. These results demonstrate the power of incorporating spatial and temporal data to better understand the relative contributions of adaptive and demographic processes to genetic variation.

1.6

Recent evolution of the mutation rate and spectrum in Europeans

Kelley Harris
UC Berkeley, Berkeley, CA, USA

As humans dispersed out of Africa, they adapted to new environmental challenges including changes in exposure to mutagenic solar radiation. This raises the possibility that different populations experienced different selective pressures affecting genome integrity. Prior work has uncovered some evidence for local adaption of eQTLs that regulate the DNA damage response, as well as indications that the human mutation rate per year may have changed by 2-fold since we shared a common ancestor with chimpanzees. Here, I present evidence that the rate of a particular mutation type has recently increased in the European population, rising in frequency by over 50% during the 40,000-80,000 years since Europeans began diverging from Asians. A comparison of single nucleotide polymorphisms (SNPs) private to Africa, Asia, and Europe in the 1000 Genomes data reveals that private European variation is enriched for the transition 5'-TCC-3'→5'-TTC-3'. Although it is not clear whether UV played a causal role in the changing the European mutational spectrum, 5'-TCC-3'→5'- TTC-3' is known to be the most common somatic mutation present in melanoma skin cancers, as well as the mutation most frequently induced in vitro by UV. Regardless of its causality, this change indicates that DNA replication fidelity has not remained stable even since the origin of modern humans and might have changed numerous times during our recent evolutionary history.

1.7

Beyond 2/3--1/3: time-dependence of the signature of sex-biased admixture on the X chromosome

Amy Goldberg, Noah A Rosenberg
Stanford University, CA, USA

Sex-biased demography, in which parameters governing migration and population size differ between males and females, has been empirically and analytically studied through comparisons of X chromosomes, which are inherited sex-specifically, and the autosomes, which are not. For the case of sex-biased admixture, empirical studies often examine the mean ancestry for markers on the X chromosome in relation to the autosomes. A simple framework suggests that the mean X--chromosomal admixture fraction can be represented as a linear combination of female and male admixture parameters, with coefficient 2/3 for the female parameter and 1/3 for the male parameter. Extending a mechanistic admixture model to accommodate the X chromosome, we demonstrate that this prediction holds only as a long-term limit for an admixture process that takes place at a single point in time. For general admixture histories, we derive recursive expressions for the time-dependent behavior of X--chromosomal admixture. For a model with constant admixture over time, we provide analytical results for the mean X--chromosomal admixture fraction, comparing admixture on the female and male X chromosomes in an admixed population to corresponding values on the autosomes. We reanalyze African-American genetic data to estimate the sex-specific contributions to African-Americans from African and European source populations. Surprisingly, we demonstrate that by explicitly accounting for time-dependence in the X--chromosomal signature of sex bias, a large range of potential contributions are compatible with the observed excess of African ancestry on the X chromosome compared to the autosomes, including scenarios without male-biased contributions from Europe.

1.8

Efficient inference of time-varying population demography and locus-specific mutation rates from large-sample genomic variation data

Anand Bhaskar [1,2], Rachel Wang[2], Yun S. Song[2]

[1] Stanford University, Stanford, CA, USA, [2] University of California, Berkeley, Berkeley, CA, USA

Human exome sequencing datasets now routinely involve tens of thousands of individuals sequenced at thousands of genomic regions. Most genomic variation observed in such large samples are in fact private to single individuals or families. Such an excess of rare variants is the result of recent rapid super-exponential population growth that has resulted in the population variation being highly out of equilibrium.

In order to study the effect of temporal demographic processes on population genetic variation, in this work, we develop an efficient method to infer recent historical population demography using genomic variation data from large samples of tens to hundreds of thousands of individuals. In particular, we develop a very efficient algorithm to infer piecewise-exponential models of the historical effective population size using the distribution of allele frequencies in large samples. Our method is based on theoretical results about the allele frequency distribution under time-varying demographies, and leverages powerful tools like convex optimization and automatic differentiation. Besides being orders of magnitude faster than previous numerical and simulation-based demographic inference methods, our method can also accurately estimate locus-specific mutation rates. We perform extensive validation of our method on simulated data and show that it can accurately infer multiple recent epochs of rapid exponential growth, a signal which is difficult to pick up with small sample sizes. Applying our method to a recent large sample exome sequencing dataset of over 11,000 individuals of European descent, we find a rapid 250-fold growth in the effective population size over the last 10,000 years.

1.9

Demographic-aware inference of the strength of purifying selection based on haplotype patterns

Diego Ortega Del Vecchyo [1], Kirk E. Lohmueller[1], John Novembre[2]

[1] University of California, Los Angeles, Los Angeles, CA, USA, [2] University of Chicago, Chicago, IL, USA

The extent of purifying selection in a population plays a key role in determining the amount of genetic and possibly phenotypic variation in a population. Recent genome sequencing studies with large sample sizes in humans capture a vast amount of putatively deleterious rare variants, providing an important source of information to analyze how selection impacts those variants. To estimate the strength of purifying selection, we have developed a novel likelihood-based method that uses the lengths of pairwise haplotype identity by state and by descent among rare-variant-carrying haplotypes. Our method conditions on the present-day frequency of the allele and is based on theory that predicts that alleles under purifying selection are on average younger than neutral alleles and should have higher average levels of haplotype identity among variant carriers. We developed a computational framework to obtain the probability distribution of the lengths of pairwise haplotype identity given a certain selection coefficient and demographic scenario. It performs two integrations: one over all possible allele frequency trajectories given a certain selection coefficient using a fast importance-sampling algorithm and another integration over all pairwise coalescent times given a certain allele frequency trajectory. We find the maximum likelihood estimate of the selection coefficient given a set of pairwise haplotypic identity lengths using a grid-search. Simulations indicate that our method provides unbiased estimates of selection in arbitrary demographic scenarios including bottlenecks and recent population growth. We estimate the selection coefficient of variants under different functional categories in a large genomic dataset of individuals from the Netherlands.

1.10

Background selection with non-equilibrium demographic models

Ryan Hernandez
UCSF, San Francisco, CA, USA

Background selection refers to the reduction of neutral genetic variation due to the purging of linked deleterious alleles. This evolutionary force is known to alter the landscape of genetic variation across a genome, and to complicate some scans for positive selection. Under some conditions, background selection can also impact the frequency distribution of linked neutral alleles. Such an effect is thought to contribute to biased inference of demographic models. Here we use simulations to quantify this effect, and set the stage for background selection-adjusted demographic inference. We find that under simple 2-epoch demographic models, the magnitude of the size change is more impacted than the timing of the population size change, but under complex demographic models the patterns are less intuitive. Using detailed, time-course simulations, we attempt to resolve this issue. While it is known that instantaneous population size changes lead to a roughly linear change in diversity, we show that background selection alters the rate at which diversity changes. In particular, background selection increases the rate at which diversity changes, which leads to fascinating time-course dynamics. These simulations give insights into how to model these dynamics, which we use to interpret the demographic history of humans and other species.

1.11

Analysing Evolve and Resequence Experiments

Srdjan Sarikas, Katarina Bodova, Nick Barton
IST Austria, Klosterneuburg, Austria

In “evolve and resequence” experiments, which start from a limited number of initial haplotypes, selected loci are typically identified by looking for marker alleles that rapidly increase in frequency. We argue that one should instead estimate the frequencies of the underlying haplotypes: over the short timescale of these experiments, each initial haplotype is only broken up into long blocks, each carrying large numbers of markers. We propose an algorithm for reconstructing haplotype frequencies and recombination breakpoints that integrates information from allele frequencies across marker loci, and so gives a much clearer picture of the evolutionary dynamics. We examine the general properties of the algorithm and assess its sensitivity to sequencing noise. Finally, we illustrate its potential using synthetic data from simulated evolve and resequence experiments, using parameters typical for Drosophila.

1.12

Using time-resolved genetic data to monitor evolving populations

Ville Mustonen
Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

Populations can evolve to adapt to external challenges. Of particular interest are adaptive processes in cancer or in bacterial, parasitic and viral infections, enabling the populations to escape from selective pressures exerted by drugs or the immune system.

To study the dynamics of such adaptation we experimentally evolved heterogeneous populations of budding yeast (*Saccharomyces cerevisiae*) under selective chemotherapy conditions. We followed the evolution of these populations over the course of 200 generations using whole-genome sequencing of the bulk population. The allele frequency patterns observed across time reveal different modes of adaptation: initially, selection acts on standing variation at segregating sites, later replaced by the emergence of macroscopic subclones. A set of *de novo* mutations within these subclones shows a validated resistant phenotype. The subclones also show genomic hallmarks of an active loss of heterozygosity dynamics that enables the resistance mutations to become homozygous. This is a collaboration project with the groups of Jonas Warringer and Gianni Liti.

The inference of the subclonal evolution from the data described above is underpinned by our probabilistic inference algorithm, cloneHD, which we use to learn the number of emerging subclones, their population fractions and their genotype posterior probabilities. A similar inference problem arises in the evolution of cancer where intra-tumour heterogeneity can underpin the emergence of resistance. Therefore, the ability to track subclonal dynamics and changes in clonal composition can inform therapy. We theoretically link the ability to monitor subclonal evolution to the development of novel therapy paradigms.

1.13

A model of protein evolution within fitness landscape changing with time

Dinara Usmanova [1,3], Inna Povolotskaya[1,2], Fyodor Kondrashov[1,2]

[1] Centre for Genomics Regulation, Barcelona, Spain, [2] Universitat Pompeu Fabra, Barcelona, Spain, [3] Moscow Institute of Physics and Technology, Moscow, Russia

Each amino acid in protein interacts with others. Thus fitness contribution of specific amino acid in particular site depends on the whole genetic background. This background changes over time resulting in change of allele fitness. In other words selection acting against particular alleles is not constant. We developed methods of analysis of long-term protein evolution which allow us to observe patterns of this altering selection. Then we formulate a covarion-like model of protein evolution, which describes this process mathematically. The model tracks not only the evolution of sequence but also the evolution of its local fitness landscape. In more details we allow fitness contribution of specific amino acid in particular site switch from being acceptable to being deleterious and vice versa. We calculated the rate of this switches for approximately 100 bacterial genes and 10000 vertebrates' genes. It appears that fitness landscape changes very fast: on average 5 switches between allowed and blocked states occur on the same timeframe as a single amino acid substitution.

1.14

The fate of a mutation in a fluctuating environment

Ivana Cvijovic, Benjamin Good, Michael Desai
Harvard University, Cambridge, MA, USA

Evolutionary tradeoffs are common and environmental variation can dramatically alter the selection pressures that act on new mutations. Yet, the influence of changing environmental conditions on the evolutionary process remains poorly understood. This makes it difficult to answer even the simplest questions: How beneficial on average an allele has to be in order to be efficiently maintained in a population? How does this depend on the rate and regularity of environmental change? We have investigated the fate of a new mutation in an environment that fluctuates between two conditions. We find that there is a broad parameter regime where the fate of the mutation cannot be determined by its time-averaged selective effect, even when it experiences many environmental epochs before fixing or going extinct. Instead, environmental variability dramatically reduces the efficiency of selection, rendering it unable to distinguish between mutations that are substantially beneficial and substantially deleterious on average. Temporal fluctuations can also dramatically increase fixation probabilities, often making the details of these fluctuations more important than the average selection pressures acting on each new mutation. For example, mutations that result in a tradeoff between conditions but are strongly deleterious on average can nevertheless be more likely to fix than mutations that are always neutral or beneficial. These effects can have important implications for patterns of molecular evolution in variable environments, and they suggest that it may often be difficult for populations to maintain specialist traits, even when their loss leads to a decline in time-averaged fitness.

1.15

Fluctuations in traveling wave models of asexual adapation: large scale oscillations in fitness variance

Lukas Geyrhofer [1], Oskar Hallatschek[2,1]
[1] *MPI for Dynamics and Self-Organization, Goettingen, Germany,* [2] *UC Berkeley, Berkeley, USA*

Adaptation takes place if populations accumulate beneficial mutations over the course of generations.
Recently, evolutionary experiments with microbes exposed an intricate dynamics: beneficial mutations appear continuously and many of them are present simultaneously in different clones.
In this dynamical regime --- termed clonal competition --- mutations interfere on their way to fixation or extinction.

Traveling wave models gained considerable attention as a possible description of these dynamics found in experiments.
In such models the population is represented as a density in fitness space, that changes via mutation and selection.
Moreover, a stochastic component is crucial for very fit clones that are still low in number:
even the fittest clones can go extint due to genetic drift.
For a sufficiently large influx of beneficial mutations the bell-shaped density moves to higher fitness as soliton-like traveling wave:
the population adapts.

By now many aspects of this behaviour, based on averaged quantities, are well understood.
However, much less is known about fluctuations in the adaptation process.
Based on a novel approach to obtain exactly solvable dynamical equations, we investigate the fluctuation behaviour in traveling wave models.
In particular, we are interested in the time-dependent behaviour of diversity, represented by fitness variance.
For parameters in the clonal competition regime, we observe emergent oscillations in fitness variance, which correspond to recurrent sweeps within the population.
Here, we present theoretical results backed by extensive simulation evidence to describe and explain this phenomenon.

1.16

The Evolutionarily Stable Distribution of Fitness Effects

Daniel P. Rice, Benjamin H. Good, Michael M. Desai
Harvard University, Cambride, MA, USA

The distribution of fitness effects of new mutations (the DFE) is a key parameter in determining the course of evolution. This fact has motivated extensive efforts to measure the DFE or to predict it from first principles. However, just as the DFE determines the course of evolution, the evolutionary process itself constrains the DFE. Here, we analyze a simple model of genome evolution in a constant environment in which natural selection drives the population toward a dynamic steady state where beneficial and deleterious substitutions balance. The distribution of fitness effects at this steady state is stable under further evolution, and provides a natural null expectation for the DFE in a population that has evolved in a constant environment for a long time. We calculate how the shape of the evolutionarily stable DFE depends on the underlying population genetic parameters. We show that, in the absence of epistasis, the ratio of beneficial to deleterious mutations of a given fitness effect obeys a simple relationship independent of population genetic details. Finally, we analyze how the stable DFE changes in the presence of a simple form of diminishing returns epistasis

328A

A hidden source of meiotic drive - allelic bias affecting human noncrossover but not crossover recombination

Linda Odenthal-Hesse [1,2], Ingrid Berg[3,2], Amelia Veselis[2], Alec Jeffreys[2], Celia May[2]
[1] Max Planck Institute for Evolutionary Biology, Ploen, Germany, [2] University of Leicester, Leicester, UK, [3] University of Stockholm, Stockholm, Sweden

Meiotic recombination contributes to DNA diversity and occurs at discrete loci called recombination hotspots. Largescale crossovers alter the haplotypic context of variants, but not their population frequencey. This is instead achieved by gene conversion events.

Gene conversions are either associated with crossover or result from nonreciprocal exchanges events, as noncrossovers (NCOs).

Biased gene conversion can result from biased initiation. Recombination is initiated on one of the homologous chromosomes via DNA double-stranded breaks, which are then repaired using the uncut homologue as templete. Consequently variants from the uncut homologue are transmitted more frequently, distorting the expected 50:50 ratio.

Transmission distortion was first observed in crossovers, but NCOs are also affected. However very little is known about NCOs and whether additional biases exist. We simultaneoulsy isolated both types of recombinants from sperm DNA at six highly active human hotspots and then looked for disparity in allelic transmission.

We not only found additional incidences of transmission distortion affecting both crossover and NCOs, but also a novel form of biased transmission exclusive to the NCO class of recombinants at two of the hotspots.

Here NCO events show strong allelic bias at a heterozygous AT/GC SNP in each hotspot, with GC alleles transmitted into 70% of NCO recombinants. This form of biased gene conversion constitutes a previously undetected and powerful source of meiotic drive, and for the first time provides direct evidence for the well-established correlation between high GC content and recombination in mammalian genomes.

329B

Modeling polygenic selection driving fast adaptation

Kavita Jain [1], Wolfgang Stephan[2]

[1] J Nehru Centre for Advanced Scientific Research, Bangalore, Karnatake, India, [2] Ludwig-Maximilians University of Munich,, Planegg-Martinsried, Germany

We consider an infinitely large population under stabilising selection and mutation in which the allelic effects determining a polygenic trait vary between loci. We obtain analytical expressions for the stationary genetic variance as a function of the distribution of effects, mutation rate and selection coefficient. We also study the dynamics of the allele frequencies, focussing on short-term evolution of the phenotypic mean as it approaches the optimum after an environmental change. We find that when most effects are
small, the genetic variance does not change appreciably during adaptation, and the time until the phenotypic mean reaches the optimum is short if the number of loci is large. However, when most effects are large, the change of the variance during the adaptive process cannot be neglected. In this case, the short-term dynamics may be described by that of a single locus of large effect. Our results may be used to understand polygenic selection driving rapid adaptation.

330C

OUTCROSSING IN *C. ELEGANS* MODELED USING COMPUTER SIMULATION OF DIGITAL "DNA" POPULATIONS

David GREEN, Chris MASON
University of Otago, Dunedin, New Zealand

Sexual reproduction incurs a cost of males in higher organisms that should result in its marginalization by clonal reproduction. This does not occur, but there is no agreement as to why. Here, we model outcrossing in *C. elegans* computationally using populations of *n*-bit genotypes of variable length. These are subject to negative frequency-dependent selection (NFDS). Models use a Wright-Fisher approach with discrete generations and a range of mutation rates. At low rates of environmental mutation, host populations can track environments with essentially no lag load, and clones and hermaphrodites compete reproductively on approximately equal terms. As the rate of environmental change increases, host lag loads emerge and where substantial, hermaphrodites respond faster to changes than clones and outcompete them. At even higher external mutation rates, reproductive interference arises in both clonal and hermaphroditic lines, and tracking of external change is thereby impaired in both. We find there is a threshold interference level above which males are stably maintained in hermaphrodite populations. The cost of males is offset, in these circumstances, by the increased efficiency with which alleles made obsolescent by NFDS are removed in pan-mictic populations, but not clonal ones. Sex increases mean population adaptation of hermaphrodites, paradoxically by reducing the genotypic variation otherwise present in the absence of males. The model successfully simulates outcrossing behavior described in a recent study of the nematode, *Caenorhabditis elegans*, subject to NFDS by the pathogen *Serratia marcescens*.

331D

The role of fluctuating selection in explaining observed genetic variation under stabilizing selection models

Sebastian Novak, Nick Barton
IST Austria, Klosterneuburg, Austria

Natural environments are never constant but vary in time and space. Furthermore, from the point of view of a single locus, there is background selection on multiple other loci, possibly for a plethora of traits that are influenced pleiotropically. To what extent can this combination of external and internal factors be described by a scheme of randomly fluctuating selection? What pattern of fluctuations can we expect to be experienced by a single locus? Continuing from there, we can venture forth trying to explain the observed levels of variation in natural populations that could so far not be reproduced convincingly by quantitative trait models of constant selection. Can fluctuating selection models unveil the missing variability?

332A

Using Gaussian process models to detect selection in time series from D. simulans experimental evolution studies

Agnes Jonas [1,2], Hande Topa[3], Antti Honkela[4], Viola Nolte[1], Christian Schlötterer[1], Carolin Kosiol[1]

[1] Institut für Populationsgenetik, Veterinärmedizinische Universität, Wien, Austria, [2] Vienna Graduate School of Population Genetics, Wien, Austria, [3] Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, Espoo, Finland, [4] Helsinki Institute for Information Technology HIIT, Department of Computer Science, University Helsinki, Helsinki, Finland

Recent advances in sequencing technologies have made it possible to observe evolutionary trajectories in more detail than ever. It has become feasible not only to sequence the last generation of a population at the end of long-term selection treatments but to monitor experimentally genome evolution at intermediate generations. Recently, experimental evolution has been combined with whole-genome resequencing in sexually reproducing multicellular organisms such as Drosophila. The resulting time-resolved data represent genome-wide evolutionary trajectories that we model with a novel Gaussian Process (GP) approach. Previous attempts to analyze data from evolve and resequencing experiments have been limited because they could not take advantage of the full time series data and replicate experiments.

Our beta-binomial Gaussian process (BBGP) model identifies SNPs with significant frequency change over time that is consistent across replicates. One major advantage of our approach is that we explicitly incorporate sequencing noise such as coverage differences with a beta-binomial model. On simulated data the GP-based methods outperform frequently used pairwise statistical tests in average precision for finding selected alleles.

We present results from applying our approach on real data from Drosophila simulans experimental evolution for temperature adaptation. We show on preliminary results that the proposed method is able to find classical signatures of selection on ACE and Timeout genes, and hold great promises to incorporate genome-wide spatial structure finding regions under selection.

333B

Fungal *CYP51* paralogues: temporal processes in the fate of duplicate genes

Nichola Hawkins, Bart Fraaije
Rothamsted Research, Harpenden, Hertfordshire, UK

Studies into the fate of duplicate genes have generally focussed on the likelihood of emergence, loss or neofunctionalisation immediately following duplication, assuming equilibrium thereafter. We track the fate of *CYP51* paralogues in ascomycete fungi, where initial duplication and neofunctionalisation has been followed by multiple subsequent losses over 400 million years. *CYP51* encodes the target site of the azole fungicides, and species which have retained the inducible *CYP51A* in addition to the constitutive *CYP51B* paralogue are intrinsically less sensitive to azoles. In *Rhynchosporium commune*, an important pathogen of barley, isolates pre-dating the introduction of azole fungicides have only *CYP51B*, whereas more recent isolates also have*CYP51A*. Phylogenetic analysis indicates that *R. commune CYP51A* diverged from *R. commune CYP51B* around the origin of the filamentous ascomycetes, long before anthropogenic fungicide use. *CYP51A* was then almost lost from the *R. commune* population, but when selective pressure changed due to the introduction of azole fungicides, *CYP51A* was selected from standing variation. We have used archived plant material from Hoosfield, one of the Rothamsted Classical Experiments, running since 1852. Through pyrosequencing, we have followed changes in paralogue frequency over time, showing the re-emergence of *CYP51A* coinciding with selection for azole resistance. Further work will look at the frequency of other genes and alleles associated with fungicide resistance in plant pathogens in samples from the Rothamsted Classical Experiments.

334C

Patterns of amino acid evolution in the mitochondrial genomes of sexual and asexual lineages of a New Zealand snail

Joel Sharbrough [1], Meagan Luse[1], Jeffery L. Boore[2], John M. Logsdon[1], Maurine Neiman[1]
[1] University of Iowa, Iowa City, IA, USA, [2] University of California-Berkeley, Berkeley, CA, USA

Understanding how mutations affect phenotype is of paramount importance in evolutionary biology but remains poorly understood. Because protein sequence represents a primary measure of phenotypic function, we can generate important inferences about genotype-phenotype connections by studying protein sequence evolution. Related sexual and asexual taxa provide a particularly powerful setting in which to study such connections because asexuals are expected to experience less effective selection than sexuals. We addressed these fundamental evolutionary questions by comparing radical nonsynonymous evolution (*i.e.*, amino acid changes between biochemical groups) to conservative nonsynonymous evolution (*i.e.*, amino acid changes within biochemical groups) in the mitochondrial genomes of eight sexual and 23 asexual lineages of *Potamopyrgus antipodarum*, a freshwater snail in which ecologically similar and closely related sexual and asexual lineages coexist. We show that the rate of conservative nonsynonymous evolution is significantly higher than radical nonsynonymous evolution and that conservative changes are fixed significantly more often than radical changes. This result provides compelling evidence that conservative changes are generally much less harmful than radical changes and that differences in the strength of selection against these two mutational types are evident even at relatively short time scales. Asexual lineages harbor more radical polymorphisms and retain these polymorphisms at higher frequencies than sexual lineages, consistent with a scenario wherein asexuals experience reduced efficacy of purifying selection. The relatively short time frame at which radical polymorphisms appear to be accumulating in asexuals suggests that deleterious mutation accumulation could happen rapidly enough to influence the maintenance of sex in natural populations.

335D

Genomic positional effects on mutation patterns in Staphylococcus aureus.

Sion Bayliss, Ed Feil, Harry Thorpe
Bath University, Bath, UK

Staphylococcus aureus is responsible for a significant public health burden. The genomes of 1822 clinical isolates of S. aureus, collected from UK hospitals between 2001-2012, were sequenced as part of a UKCRC consortium investigating the transmission and microevolution of this pathogen. These data provide unprecedented detail into mutational processes impacting across the S. aureus genome over very short time scales. We first considered the genomic distribution of nonsense and frameshift mutations (together referred to as “disruptive mutations”). As expected, we note an enrichment of disruptive mutations in genes encoding proteins associated with the cell surface and an avoidance of mutations in genes under strong selective constraint, such as ribosomal proteins. Genes assigned as “essential” on the basis of transposon knockout data are significantly less likely to experience disruptive mutations in nature, although a small number of so-called essential genes appear to have been pseudogenised in a small number of genomes. Surprisingly, we noted an enrichment of disruptive mutations around the origin of replication which cannot be explained by changes in base composition or coding density. Moreover, this enrichment is not apparent for synonymous or nonsynonymous mutations. We sought to explain this effect by examining mutation biases across the genome. Examination of four-fold degenerate sites revealed a significant enrichment of G>A and C>T transitions, and an overall increase in mutation rate, around the terminus of replication. Whilst these patterns remain largely unexplained, they point to a central role for DNA replication in shaping mutational processes across the genome.

336A

Models for strong seasonally fluctuating selection at many loci

Meike J. Wittmann, Alan O. Bergland, Marcus W. Feldman, Dmitri A. Petrov
Stanford University, Stanford, California, USA

Recently, strong seasonal allele-frequency fluctuations (~10%) were reported for hundreds of SNPs throughout the genome in a temperate population of *Drosophila melanogaster*. Fluctuations were consistent over several years and many of the SNPs appear to be ancient balanced polymorphisms. Classical population genetics theory is not sufficient to explain such observations. Although temporally fluctuating selection is known to maintain single-locus polymorphism if heterozygotes have the highest geometric mean fitness over time, it is unclear whether this marginal overdominance is sufficient to maintain diversity at many loci. In fact, it has been debated for a long time whether there is a limit to the number of polymorphisms that can be maintained by heterozygote advantage. Here we use individual-based simulations and deterministic stability analyses to explore the ecological and evolutionary conditions under which strong seasonally fluctuating selection can promote long-term stable polymorphism and predictable oscillations at many unlinked loci. We find that stability and fluctuations are extremely sensitive to the mapping between multi-locus genotype and fitness. For many plausible scenarios where fitness depends on a quantity that is additive within and between loci ("equivalence models"), polymorphism and fluctuations are either absent or occur at few loci, even if there would be strong marginal overdominance at single loci. By contrast, under a broad class of non-equivalent models where heterozygous loci contribute more than proportionally to seasonal adaptation, we obtain stable polymorphism and strong allele-frequency fluctuations at hundreds of loci, especially if there are multiple generations per season.

337B

An exact method for reconstructing variable past population size from distributions of coalescent times

Lucie Gattepaille, Mattias Jakobsson
Evolutionary Biology Center, Uppsala, Sweden

Species and populations rarely evolve under the simple assumptions of most models of population genetics. Here, we focus on the problem of inferring variable past population size using polymorphism data from a sample of any size, in a population otherwise behaving under all other assumptions of the Wright-Fisher model. In our novel method, we provide the exact analytical solution of the past population size from the distributions of coalescent times. We show that population size can still be accurately estimated when using the empirical instead of the true distributions. We study the robustness of the reconstruction method to a range of different conditions: sample size, number of loci, recombination rate, mutation rate. Finally we demonstrate on human sequence data from the 1000 genomes project how this analytical result can be used simply to uncover the main features of human past population sizes.

338C

Time-dependent patterns in molecular rate estimates - the prevalence, causes and implications

Martyna Molak
Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland

Molecular clock hypothesis, which postulates a fairly constant rate of change in DNA sequences over time, allows inferences about evolutionary history of organisms. Basing on the observable genetic differences and non-genetic temporal information, a tempo of molecular evolution can be estimated. This in turn allows estimation of the timing of various evolutionary events such as species divergences and putting these events in a chronological context together with various geological, environmental and other evolutionary events.
Since the time when the molecular clock hypothesis was first proposed by Emile Zuckerkandl and Linus Pauling in 1962 it has been recognized that substitution rates differ between lineages and genes and the approaches to the application of molecular clock have been changed accordingly. Additionally rate estimates have been shown to depend on the timescale over which they are estimated. This time-dependence of rates has raisen an intense discussion as well as many misconceptions.
Time-dependence of rate estimates has potentially great implications for the studies of evolutionary timescales and can lead to erroneous interpretations of phylogenetic inference studies. In my talk I would like to present results of my and Dr Simon Ho's recent project on the prevalence and persistence of time-dependent patterns in rate estimates for mitochondrial markers in metazoans. I will present evidence for the ubiquity of time-dependent patterns. I will also attempt to explain where these patterns come from, what impact they might have on phylogenetic inference and advise how to avoid timescale estimation biases caused by time-dependent patterns.

339D

Reliable fitness estimation of known haplotypes using pool-sequencing data

Nicolas Rode [1], Vincent Ranwez[2], Laurène Gay[1]
[1] *INRA, Montpellier, France,* [2] *Supagro, Montpellier, France*

Fitness is a central concept in evolutionary biology, yet its empirical measurement is notoriously difficult, especially in natural populations. In highly-selfing and asexual species, GBS techniques in combination with pool-sequencing represent a promising cost-effective method to estimate fitness as genotypic frequency change over time.

Estimating fitness accurately requires accounting for different types of variation (e.g. drift, experimental sampling and sequencing stochasticity).

In a first step, we estimated variation due to DNA extraction and sequencing by performing pool-sequencing of 10 different genotypes of the highly selfing plant *Medicago truncatula*. We quantified the effects of fitness differences between genotypes (using balanced and highly unbalanced mixes) and genotypic polymorphism (using either genotypes from the same or from different populations).

In a second step, we performed simulations with different levels of drift and sampling error to investigate their relative impacts on fitness estimation.

We discuss the relative advantages and drawbacks of this approach and show preliminary results from a long-term local adaption study in Southern France.

340A

Human population substructure and population size changes in the Japanese Archipelago

Timothy Jinam, Isaac Babarinde, Naruya Saitou
National Institute of Genetics, Mishima, Japan

The genetic diversity of modern human populations in the Japanese Archipelago was shaped during the Paleolithic (>16 thousand years ago, kya), Jomon (16-3 kya) and Yayoi (3-1.7 kya) periods. Using genome-wide SNP data generated using Affymetrix 6.0 genechip, we examined the population structure and admixture model in the Ainu, Ryukyuans and Japanese Mainlanders. We found that by using Ainu as descendants of Jomon people and continental Asians (Han Chinese, Koreans) as descendants of Yayoi people, the proportion of Jomon genetic component in the Mainlanders was ~18% and ~28% in Ryukyuans. The time since admixture for Mainlanders ranged from 55-58 generations ago, and 43-44 generations ago for the Ryukyuans. We also identified highly differentiated loci in the Ainu, which include the *EDAR* gene responsible for tooth and hair morphology. Population size estimates using complete mitochondrial DNA in Japanese Mainlanders shows a population growth that corresponds to the Yayoi period. These results are in agreement with historical records and confirm the dual-structure model for the origins of the Japanese people.

341B

An evolutionary study of the *RNF213* gene associated with Moyamoya disease in the people of the Japanese archipelago.

Kae Koganebuchi [1], Shigeki Nakagome[2], Shuhei Mano[3], Naoya Ishizaki[4], Shoji Kawamura[4], Ryosuke Kimura[5], Hajime Ishida[5], Keiichiro Joh[6], Hidenobu Soejima[6], Kazuma Fujimoto[6], Kimitoshi Sato[7], Yoshie Yasui[7], Toshihiro Kumabe[1,7], Kiyotaka Fujii[1,7], Tokiho Akiyama[8], Motoyuki Ogawa[1,7], Tsunehiko Hanihara[1,7], Hiroki Oota[1,7]

[1] Graduate School of Medical Sciences, Kitasato University, Kanagawa, Japan, [2] Department of Human Genetics, University of Chicago, Illinois, Japan, [3] The Institute of Statistical Mathematics, Tokyo, Japan, [4] Graduate School of Frontier Science, The University of Tokyo, Chiba, Japan, [5] Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan, [6] Graduate School of Medicine, Saga University, Saga, Japan, [7] School of Medicine, Kitasato University, Kanagawa, Japan, [8] School of Advanced Sciences, The Graduate University for Advanced Studies, Kanagawa, Japan

Moyamoya disease is characterized by vascular stenosis and development of the vessel network around the stenosis. Recently, two studies have reported a risk mutation, Arginine to Lysine at amino acid position 4810 (R4810K), of the *RNF213* gene (138Kb) as shown by GWAS and a pedigree analysis. The previous study shows that the odds ratio of Japanese cases (338.9) is higher than those of Chinese (135.6) and Koreans (14.7). The risk allele frequency in Japanese the cases (48.1%) is higher than those in the other East Asians (Chinese: 12.5%, Koreans: 39.5%). This mutation isn't found outside of East Asia, suggesting it would have been born after the divergence between Europeans and East Asians. Therefore, it's interesting to investigate R4810K of the *RNF213* gene in Japanese.

In our preliminary analyses, we sequenced Japanese cases (N=32) and controls (N=90) (47 Ryukyu islanders; 43 northern Kyushu) PCR-products by the Sanger method for four exons where amino acid changes had been reported in Japanese. The sequence showed that the genetic diversity of these regions was lower than the average for autosomes in the other non-Africans. Linkage disequilibrium (LD) analysis based on HapMap Phase3 database showed higher LD around R4810K (50Kb) in non-Africans than in Africans. These suggest that the region around R4810K is conserved by purifying selection.

342C

Evolution of a single-stranded DNA phage under an elevated mutation rate

Alexander Wilcox, Benjamin Dickins
Nottingham Trent University, Nottingham, UK

Mutation is the ultimate source of genetic variation on which evolutionary processes act. As a phenotype affected by the efficiency of proof-reading polymerases or DNA repair enzymes, the mutation rate is also plausibly a selected characteristic of organisms with large populations. A supply of beneficial mutations is required for adaptation to changing environments, but the accumulation of deleterious mutations can reduce the mean fitness of a population. If the mutation rate is sufficiently high this may lead to population extinction and, if realisable, lethal mutagenesis may be deployed as an effective strategy against various types of virus. To the best of our knowledge this has only been demonstrated in RNA viruses, and it is unclear whether this is a feasible strategy for DNA viruses with their lower base mutation rates.

The single-stranded DNA phage ΦX174 does not encode its own polymerase, so its mutation rate is dependent on the polymerase of its host, *Escherichia coli* C. Using mutant alleles of *dnaQ*, which encodes the error-correcting ε subunit of *E. coli* DNA polymerase III, we have produced environments that greatly increase the mutation rate of ΦX174 without altering the phage itself or introducing mutagenic chemicals that can inactivate phage growth .

We present data from lines of ΦX174 evolved by serial passaging under different mutation rates and we discuss the effects of varying the mutation rate on the fitness and genetic composition of populations of ssDNA organisms.

343D

Evolutionary and network properties of human disease genes

Nino Spataro, Elena Bosch Fusté
Institut de Biologia Evolutiva (IBE), Barcelona, Spain

What evolutionary forces shape the genetic architecture of disease genes? Do disease and non-disease genes have different network properties? Do the same evolutionary pressures affect Mendelian and complex human disorder genes? What is the role of coding and regulatory sequences on disease gene evolution?
The answers to these questions could shed light on understanding human genetic disorders and help to predict which types of allele variants and allele frequencies contribute to disease risk, with relevant implication on mapping strategies for future genetic studies. We have analyzed and compared sequencing data from the 1000 Genomes project as well as expression data and the network properties of human Mendelian genes (from the hOMIM database), genes associated to complex disorders (from the GWAS catalog database), essential genes (considering the orthologs human gene of mouse essential genes detected by knock-out experiments) and of the remaining human protein coding genes in Ensembl.
Our results show that disease genes have specific evolutionary profiles and protein network properties when compared to non-disease genes. Moreover, those genes that are both causal for Mendelian diseases and risk factors for complex traits of medical relevance tend to show particular protein network and expression features, higher effect sizes and specific evolutionary pressures on their coding and regulatory sequences. The observed pattern on this subset of disease genes suggests that purifying selection could not be the only force shaping the properties of this subgroup and that the classic scenario of mutation-selection balance may be inappropriate for some genes underlying human genetic disorders.

344A

Mitochondrial Hotspots: Evidence of Widespread Selection at Synonymous Sites?

Meggan Alston [1], Will Goodall-Copestake[2], Sílvia Pérez-Espona[3], Sarah Bradburn[1], Carina Marón[1], Jon Seger[1]

[1] Department of Biology, University of Utah, Salt Lake City, Utah, USA, [2] British Antarctic Survey, Natural Environmental Research Council, Cambridge, UK, [3] Department of Life Sciences, Anglia Ruskin University, Cambridge, UK

Multiple substitutions often occur at individual synonymous sites in large samples of conspecific mitochondrial DNA sequences, with a few sites having unexpectedly large numbers relative to the mean. Elevated site-specific mutation rates have been proposed as an explanation for this pattern of "hot spots." Selection could also simultaneously increase the frequencies of many lineages carrying independent occurrences of a given mutation, especially in a large population with abundant standing variation. An incipient soft sweep of this kind would be hard to distinguish from an elevated site-specific mutation rate. Here we analyze site-specific substitution frequencies in some large data sets for the cox1 genes of several species of marine invertebrates (whale lice and krill), and smaller data sets for several other genes. "Hot" synonymous sites are found in all of the data sets, but they occur in different places, even in closely related species (as seen in the nad2 sequences of primates by Galtier et al. 2006). Coalescent simulations suggest that such patterns could be caused by selection, possibly induced by complex epistatic effects of synonymous substitutions on rates of translation (Chevance et al. 2014), which could be sensitive to varying environmental factors such as temperature.

345B

Ancient, but not recent, population declines have had a genetic impact on alpine yellow-bellied toad populations, suggesting potential for complete recovery.

Andrea Benazzo [1], Luca Cornetti[2], Sean Hoban[3], Cristiano Vernesi[2], Giorgio Bertorelle[1]
[1] University of Ferrara, Ferrara, Italy, [2] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy, [3] National Institute for Mathematical and Biological Synthesis (NIMBioS), University of Tennessee, Knoxville, Tennessee, USA

When low genetic variation and genetic bottlenecks are found in recently declining populations, it is common to assume that the demographic decline is the cause of the genetic pattern, implying that the population/species is trapped in an extinction vortex. However, the genetic pattern may be related to ancient events in species history and not (yet) to recent processes, supporting the view that complete recovery is possible and thus avoiding excessive alarmism. Here we show that this probably the case for several declining population of *Bombina variegata* in the Alps, by using a set of statistical methods, including likelihood methods and Approximated Bayesian Computation, applied to 200 individual genotypes obtained from 11 nuclear microsatellites. Inbreeding within populations was high, and the effective sizes in the last few generations, as estimated from the random association among markers, never exceeded few dozen of individuals. Our most important result is that several analyses converge in suggesting that genetic variation was shaped in all groups by a 7- to 45-fold demographic decline, which occurred between a few hundred and few thousand years ago. Remarkably, only weak evidence supports recent genetic impact related to human activities. We believe that the alpine *B. variegata* populations should be monitored and protected to stop their recent decline and to prevent local extinctions, with highest priority given to two genetically isolated populations. Nonetheless, current genetic variation pattern, being mostly shaped in earlier times, suggests that complete recovery can be achieved.

346C

Drift model selection in time-series evolutionary analysis

Nuno Nene [1], Ville Mustonen[2], Christopher Illingworth[1]

[1] University of Cambridge, Cambridge, UK, [2] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

The latest developments in sequencing technologies and methodologies such as evolve and re-sequence have allowed for the collection of richer time-series datasets. As a result, finer maps between phenotypic changes and the underlying genetic variation occurring during experimental evolution have been generated. Typical strategies based, for example, on diffusion approximations to the Wright-Fisher model, have enabled the development of efficient statistics tailored to the identification of targets of selection via changes in allele frequency. Yet, the assumption behind the actual drift model has not been questioned. In fact, under various circumstances the diffusion approximation may not be distinguishable from the purported Wright-Fisher approach. In order to address this question, we compare different drift assumptions in experimental evolution time-series analysis. We evaluate the performance of multiple noise models, including exact Wright-Fisher sampling, in estimating N and selection coefficient sizes.

347D

Gene evolution in dependence on position on human chromosomes

Dorota Mackiewicz
Department of Genomics, Biotechnology Faculty, University of Wroclaw, Wroclaw, Poland

Meiotic recombination is the main process in driving the evolution of genomes by generating new combinations of alleles and also in ensuring the faithful segregation of homologous chromosomes during gametogenesis. Genomic analyses of this process have shown an uneven distribution of recombination events along the chromosomes in many eukaryotic genomes including the human genome. The mean recombination rate is higher in the sub-telomeric regions than in the middle parts of chromosomes. Computer simulations have demonstrated that for this variable distribution can be responsible a different strategy of genes' evolution dependent on their location on chromosome. The strategy of complementation is favored in the center, whereas the purifying selection occurs near the end of chromosomes. Consequently, genes can form clusters preferentially in the central parts of chromosomes, in the regions that emerge as "recombination deserts". This effect should influence the evolution of genes in dependence on their location along chromosomes. Comparisons of human genes with orthologs from the set of different genomes that successively earlier separated from the common ancestor have shown that the probability of finding homologous genes is higher for genes located near the center of chromosome than at the end. The genes in the sub-telomeric regions were also characterized by a greater divergence. Thus, chromosomes with the uneven distribution of recombination rate can lost homology faster in the sub-telomeric regions than in the central parts. The main reason of this process can be a purifying selection eliminating defective alleles.

348A

A temporal perspective on the interplay of demography and selection on deleterious variation in humans

Evan Koch [1], John Novembre[2]

[1] Department of Ecolofy and Evolution, University of Chicago, Chicago, IL, USA, [2] Department of Human Genetics, University of Chicago, Chicago, IL, USA

A large portion of low-frequency human genetic variation has been predicted as deleterious for fitness by a variety of methods. Since the amount and distribution of deleterious mutations harbored by a population depends on size, and human populations vary greatly in their recent demographic history, a number of studies have sought to describe differences in deleterious variation between human populations. We provide theoretical expectations for these differences, using a diffusion model to get a greater level of detail than previous studies. We focus on African and European population trajectories, and examine heterozygosity, homozygosity, and derived allele burden at a fine scale over selection and time. These patterns are compared to those in simplified demographic scenarios. Consistent with basic population genetic expectations, we predict that the historically reduced effective population size in OOA populations has and continues to cause an accumulation of deleterious mutations, but that this rate is low. We examine the concordance between theory and data on this point. Equilibrium theory is provided to show how the proportion of deleterious variants in a sample is affected by selection, drift, and mutation, and its dependence on past demographic events and sample size is demonstrated in nonequilibrium cases. To investigate some of the more subtle patterns in our theoretical results, we analyze sequence data from the Exome Aggregation Consortium, the large sample size of which allows us to look more precisely at strongly deleterious variation.

L 2B

Analysis of synteny and rearrangements in cheetah genome based on multiple sequence alignment

Ksenia Krasheninnikova[1] , Mark Diekhans[2], Pavel Dobrynin[1] , Gaik Tamazian[1] , Guojie Zhang [3] , John Bhak [4], Stephen O'Brien [1]

[1] Dobzhansky Center for Genome Bioinformatics, St Petersburg, Russia, [2] Center for Biomolecular Science and Engineering, University of California Santa Cruz, USA, [3] Personal Genomics Institute, Genome Research Foundation, Republic of Korea, [4] BGI-Shenzhen, China

Accurate multiple sequence alignment provides a profound data for comparative species analysis. We performed multiple species alignment of Felis Catus, Panthera Tigris Altaica, Panthera Leo, Acinonyx Jubatus, and Canis Familiaris assemblies in order to investigate the synteny and rearrangements occurred between domestic cat and cheetah genomes. Cheetah genome was covered in alignment more than for 91% by the other Felidae species and for 96% by synteny blocks with the domestic cat genome. We found out 19 putative large-scale rearrangements between cat and cheetah based on homologous synteny blocks. The discovered divergence rates agrees with the known evolutionary distances among species. The obtained results can be used for assemblies improvements and further biological analysis of homologous and breakpoint regions.

.

2 Untangling information, noise, and phylogenetic reconstruction in genome scale data

2.1

Analytical and computational challenges in coalescent-based species tree estimation

Tandy Warnow [1], Siavash Mirarab[2], Md. Shamsuzzoha Bayzid[2], Bastien Boussau[3]
[1] University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, [2] University of Texas at Austin, Austin, Texas, USA, [3] University of Lyons, Lyons, France, France

Estimating the Tree of Life will likely involve a two-step procedure, where in the first step trees are estimated on many genes, and then the gene trees are combined into a tree on all the taxa. However, the true gene trees may not agree with with the species tree due to biological processes such as deep coalescence, gene duplication and loss, and horizontal gene transfer. Statistically consistent methods based on the multi-species coalescent model have been developed to estimate species trees in the presence of incomplete lineage sorting; however, the relative accuracy of these methods compared to the usual "concatenation" approach is a matter of substantial debate within the research community.

I will present results showing that coalescent-based estimation methods are impacted by gene tree estimation error, so that they can be less accurate than concatenation in many cases. I will also present two new methods (ASTRAL and statistical binning) for estimating species trees in the presence of gene tree conflict due to ILS that are more accurate than current methods. Key to these methods is addressing gene tree estimation error more effectively. Finally, I present preliminary results investigating whether statistically consistent accurate species tree estimation is possible when gene trees have estimation error.

ASTRAL (Mirarab et al., Bioinformatics 2014) was used in the Thousand Plant Transcriptome Project (Wickett et al., PNAS 2014), and Statistical Binning (Mirarab et al., Science 2014) was used in the Avian Phylogenomics Project (Jarvis, Mirarab, et al., Science 2014).

2.2

Species Tree Inference with Polymorphism-Aware Phylogenetic Models

Dominik Schrempf [1,2], Nicola De Maio[3], Carolin Kosiol[1]

[1] Institut für Populationsgenetik, Vetmeduni Wien, Wien, Austria, [2] Vienna Graduate School of Population Genetics, Wien, Austria, [3] Nuffield Department of Medicine, Oxford University, Oxford, UK

The availability of genome-scale inter- and intraspecies data leads to new opportunities in phylogenetics to improve tree accuracy and resolution as well as to take important steps towards understanding the process of speciation.

We present a novel maximum likelihood implementation of a Polymorphism-Aware Phylogenetic Model (PoMo, De Maio et al., MBE 2013) that can do both, parameter estimation and species tree inference for genome-wide data of a moderate number of species while still allowing for many individuals per species. It extends any DNA substitution model and additionally accounts for polymorphisms in the present and in the ancestral population by expanding the state space to include polymorphic states. It is a selection-mutation model which separates the mutation process from the fixation process. PoMo has been shown to efficiently and accurately estimate evolutionary parameters relevant for GC variation along mammalian genomes from exome-wide alignments of four great ape species (De Maio et al., MBE 2013).

Now, we have extended PoMo to infer species trees. Extensive simulations with different demographic scenarios and evaluation against other state-of-the-art methods show that it is fast while being more accurate. Although a single phylogeny — the species tree — is considered, PoMo naturally accounts for incomplete lineage sorting because ancestral populations can be in a polymorphic state. We show results of applying PoMo with species tree inference to data of a recent study on the genetic diversity and population history of the great apes (12 populations in total, Prado-Martinez et al., Nature 2013).

2.3

Statistical Inference of Reticulate Evolutionary Histories

Luay Nakhleh [1], Yun Yu[1], Jianrong Dong[1], James Degnan[2], Kevin Liu[3]
[1] Rice University, Houston, Texas, USA, [2] University of New Mexico, Albuquerque, New Mexico, USA, [3] Michigan State University, East Lansing, Michigan, USA

For various groups of eukaryotic species, phylogenetic reconstruction involves accouting for the possibility of hybridization between species. Further, as hybridization occurs between closely related species, incomplete lineage sorting (ILS) tends to also occur. Thus, phylogenetic reconstruction in this case must account for both processes simultaneosuly. One approach to such reconstruction is to first estimate gene trees for unlinked loci, and then infer an evolutionary history of the species while accounting for both processes. In this case, uncertainty in gene tree estimates must also be accounted for.

In this talk, we present a phylogenetic network model that accounts for both hybridization and ILS. We then present a maximum likelihood formulation for the inference of such a phylogenetic network from a collection of gene trees. To account for uncertainty in the gene tree estimates, we show how the formulation makes use of bootstrap results. For the input gene trees, we will describe the modeling and inference steps when the trees are given only by their topologies and when they are given by both their topologies and branch lengths.

This is joint work with: James Degnan, Jianrong Dong, Kevin Liu, and Yun Yu.

2.4

Approximate Likelihood Estimation of Divergence Time Range in Population Trees Using a Coalescent-based Model

Arindam RoyChoudhury
Columbia University, New York, New York, USA

We present an estimator of divergence time-range in population trees based on a coalescent model. This model sum the probability of coalescent trees, taking into account the effect of incomplete lineage sorting. Maximum likelihood estimator based on this model has been introduced previously; however, a formula for divergence time-range estimator (or confidence interval) has never been presented for this model because the expression of the likelihood makes this estimation difficult to compute. We did not use simulation or bootstrap-based approaches and therefore our method is fast and less computationally intensive. We demonstrate that our method is much faster, but as accurate as a simulation-based approach.

2.5

Partition-Aware Tree Testing

Heiko Schmidt [1,2], Florian Stertz[3,2]

[1] University of Vienna/Medical University of Vienna, Vienna, Austria, [2] CIBIV, Max F. Perutz Laboratories, Vienna, Austria, [3] Medical University of Vienna, Vienna, Austria

Nowadays species-phylogenies are commonly reconstructed from multiple genes, aiming at resolving more splits and studying larger species-sets with higher accuracy. However, the (possibly contradictory) gene-histories introduce more variance. This together with the exponentially growing tree-space often causes one tree reconstruction heuristic to produce different phylogenies than another, or even a repetition of the same heuristic. This emphasizes the importance of testing which trees are significantly worse to assess which of the alternative trees have to be considered. Typically, tests based on resampling estimated site-log-likelihoods (RELL) are used for this purpose like Simodaira-Hasegawa and the approximately unbiased (AU) test.

However, multi-gene datasets often expose varying taxon coverage. In addition, evolutionary models differ from gene to gene when using partition-models to reflect the gene-specific evolutionary constraints. This suggests that the partition structure should be taken into account when testing trees.

We suggest a partition-aware approach to obtain the Null-distribution for the aforementioned tests. The RELL strategy is adopted to take partitions into account to reflect the differences among the genes. We have implemented this ParConsel strategy based on the Consel implementation. We compare the performance of ParConsel (with and without partitions) to Consel using several phylogenomic datasets.

In addition ParConsel allows for testing based on subsets of genes either to reduce the patchiness of data or to study the impact of certain genes on the test results. This is reasonable if so-called full partition-models are used, where the branch length estimates of one partition does not influence those of the other partitions.

2.6

Speed Dating with Least Squares

Thu Hien To[1], Matthieu Jung[1], Samantha Lycett[2], Olivier Gascuel [1]

[1] Institut de Biologie Computationnelle, LIRMM, CNRS & Université de Montpellier, Montpellier, France, [2] Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh, UK

Phylogenies provide a useful way to understand the evolutionary history of genetic samples, and data sets with more than thousand taxa are becoming increasingly common, notably with viruses (*e.g.* HIV). Dating ancestral events is one of the first, essential goals with such data. However, current sophisticated probabilistic approaches struggle to handle data sets of this size. I'll present very fast dating algorithms, based on a normal approximation of the Langley-Fitch molecular-clock model. These algorithms apply to serial data, where the tips of the tree are dated, and could be extended to trees with time calibration points. These algorithms estimate the substitution rate and the dates of all ancestral nodes, and the root position with unrooted trees. They exploit the tree (recursive) structure of the problem at hand, and the close relationships between normal distribution, least squares, and linear algebra. Temporal precedence constraints are accounted for using an active set method. Very large input trees (>10,000 taxa) can easily be processed. Comparisons with standard methods, using both simulated and real data sets, show that these algorithms are surprisingly accurate. They are implemented in the LSD software (Least Squares Dating), which can be downloaded from http://www.atgc-montpellier.fr/LSD/.

2.7

Using a phylogenetic approach to investigate genomic evolution

K Jun Tong, Sebastian Duchene, Charles SP Foster, Nathan Lo, Simon YW Ho
University of Sydney, New South Wales, Australia

The increasing availability of genome-scale data allows the use of filtering approaches that identify the optimal genes for phylogenetic inference. This has great implications for molecular dating, which can be confounded by variation in rates of evolution. We can conveniently compartmentalise this variation into three components: among lineages, genes, and nucleotide sites. These components can add upon as well as interact with one another to produce complex rate variation between and within genomes. This variation is driven by factors ranging from neutral drift to strong selection. Using a phylogenetic approach, we posit that the branch-length patterns of gene trees reflect the relative influence of drift and selection upon different genes. We expect long gene trees to share very similar branch-length patterns because these genes are primarily subject to lineage effects. These genes are ideal candidates for dating analyses. Conversely, short gene trees are expected to have very distinct branch-length patterns from each other because they are primarily affected by gene-lineage interactions. We test our predictions by employing a phylogenetic clustering approach to genomic datasets from four very distinct groups of organisms: holometabolous insects, mammals, angiosperms and aphid bacterial endosymbionts. Our results will contribute to new analytical frameworks that seek to filter genome-scale data in order to identify the optimal genes for molecular clocks.

2.8

LS³: A data subselection algorithm to reduce long branch attraction artifact in multi-gene phylogenies

Carlos J. Rivera-Rivera, Juan I. Montoya-Burgos
University of Geneva, Geneva, Switzerland

Phylogenetic inference artifacts can occur when sequences evolve in ways that violate the assumptions made by the models of sequence evolution employed to analyze them. Current models of sequence evolution assume that evolution is a lineage-independent process, an assumption that is violated when rates of evolution are heterogeneous across taxa. Strong lineage rate heterogeneity can lead to the well-described long branch attraction (LBA) artifact. We define an objective criterion for assessing taxon rate homogeneity in any alignment: the result of a Likelihood Ratio Test (LRT) comparing the likelihood scores of (i) a model in which ingroup lineages all evolve at the same rate (homogeneous model) and (ii) a model in which ingroup lineages are allowed to evolve at their own rates (heterogeneous model). We implement this criterion in an algorithm called LS³, which sequentially removes the fastest-evolving taxon in a gene dataset and tests for lineage rate homogeneity until a taxon sample with homogeneous lineage rates is found for that gene. The resulting dataset will better fulfill the assumption of lineage-independent evolution, reducing the effects of LBA artifacts. This method was tested with multi-gene datasets leading to verified LBA artifacts: simulated nucleotide datasets and datasets from two real, well documented biological cases; one nucleotidic (the position of rodents within mammals) and one composed of amino acids (the position of nematodes within bilaterians). In the three cases, the misled phylogenetic inferences were unambiguously corrected upon processing the data with the LS³ algorithm.

2.9

A whole-genome phylogenetic hypothesis across the three domains of life

Rebecca Dikow
Smithsonian Institution, Washington, DC, USA

Whole genome phylogenetics has often been limited to closely related and small numbers of species due to computational challenges and the difficulty of detecting homology across deeply diverging lineages. The relationships among the three domains of life, Archaea, Bacteria, and Eukaryota, have previously only been hypothesized based on very small numbers of genes and often based on gene content rather than explicit phylogenetic hypotheses. Here, these relationships are hypothesized based on explicit phylogenetic analysis of complete genome data from 2,000 species across the three domains employing whole genome alignments. The scale of this work illuminates the practical and bioinformatics challenges that exist when comparing complete genomes across deep divergences. Phylogenetic results demonstrate that Archaea, as well as most Bacterial and Eukaryotic phyla as currently circumscribed are monophyletic, but Bacteria as a whole appears to be polyphyletic. Eukaryotic taxa are less densely represented due to the small number of finished genomes for many groups. Inclusion of draft eukaryote genomes points to the challenges of genome assembly for large and repetitive genomes, as contamination by symbionts is shown in the presented phylogeny. This whole-genome phylogenetic approach provides the opportunity to resolve branching order of ancient divergences because it is an heuristic approach to homology detection. These homologies can subsequently be compared with functional annotations, allowing us to better understand how homology relates to function across diverse genomes.

2.10

Gene mode of evolution, phylogenetic informativeness and tree structure: A study in land plants

Khidir Hilu
Virgina Tech, Blacksburg, Virginia, USA

Phylogenetic informativeness/noise of genomic regions displaying multiple permutations of rates of substitution and synonymous/non-synonymous ratios is empirically evaluated. Representatives of major land plant lineages were used as a platform to asses the impact of gene mode of evolution on tree robustness at historic depths of ~ 460 MY. Gene/gene-combinations partitioned data sets were constructed for seed plants and land plants from plastid *matK*, *atpB*, *rbcL,* and mitochondrial *matR*. RAxML and maximum parsimony trees were computed to evaluate structure. Fossil-calibrated ultrametric trees were generated, and the output was executed in the PhyDesign program to evaluate phylogenetic informativeness. Noise was measured with the ensemble consistency and retention indices. The rapidly evolving/unconstrained *matK* faired best, while remaining genes varied in degree of informativeness and contribution to tree robustness depending on their rate and mode of substitutions. Third codon position was the most informative compared with the 1^{st} and 2^{nd}. These findings are in clear contrast to the views that rapidly evolving regions and the 3^{rd} codon position has inevitable negative impact on phylogenetic reconstruction at deep historic levels due to accumulation of multiple hits and subsequent elevation in homoplasy and saturation. However, excessive rates of substitutions and substantial historic depths may present constraints on homology assessment of alignments.

2.11

Rapid identification of phylogenetically informative data from next-gen sequencing reads

Rachel Schwartz, Reed Cartwright
Arizona State University, Tempe, AZ, USA

Complete genomic data provides us with the information to determine relationships among organisms; however, extracting phylogenetically useful information from these data has been slow and complicated. Here we present a method called SISRS (Site Identification from Short Read Sequences), which extracts phylogenetically informative data from next-generation sequencing (NGS) reads without a priori knowledge of the genomes of the sampled species. SISRS first produces a de novo assembly using a subset of reads from each species. This "composite" assembly consists primarily of loci that are conserved across species, as sequences unique to one or a few species will be relatively rare in the pooled data. Reads from all species are then aligned to the composite genome to identify the sequence for each species for each locus. For both simulations and empirical genome/transcriptome data, SISRS rapidly identifies conserved, homologous, phylogenetically informative data, resulting in accurate estimates of tree topology and divergence dates, even at high levels of divergence. The datasets we produce using SISRS confirm that AT-biased data support a different phylogeny in mammals than data consistent with mammalian GC content. The SISRS approach is implemented in easy-to-use, open-source software, and can generate well-resolved phylogenies directly from NGS data in a few days. Having rapidly identified phylogenetically informative data, we can use these loci to design probes to do phylogenetics for large numbers of species simply and cheaply. We demonstrate that this "optimal genes" approach works as well or better than a whole genome approach due to the reduction in noise.

2.12

Bayesian information content estimation and application to phylogenomics

Karolina Fucikova, Paul Lewis, Ming-Hui Chen, Lynn Kuo, Louise Lewis
University of Connecticut, Storrs, CT, USA

The accelerated pace of genomics underscores the need for accurate measures of phylogenetic information in sequence data. Faced with prevalent gene tree conflict due to incomplete lineage sorting, horizontal transfer, and non-stationarity of the substitution process, concatenation is under increasing scrutiny and criticism, lending support to methods that can both identify data subsets with strong phylogenetic signal and measure the degree of conflict among subsets.

The Bayesian statistical framework lends itself to the measurement of information. Lindley pioneered using the difference in entropy between the posterior and prior distributions as a natural measure of the amount of information provided by the data (via the likelihood) relative to the information provided by the prior. In phylogenetics, maximum information implies a posterior with minimum entropy (all probability concentrated on a single tree topology), while no information implies entropy equal to the prior, which has maximum entropy if the prior is a discrete uniform distribution over all possible topologies.

We combine Larget's conditional clade framework with Lindley's method to measure Bayesian phylogenetic information content. We show how topological information content is distributed across the phylogeny and partitioned among data subsets, and how the difference between the average information content of individual genes and the information content of the merged posterior distribution measures the degree of gene tree discordance. We conclude with results of simulations and empirical examples that illustrate the utility of our approach.

2.13

Parametric estimation of evolutionary constraint using amino acid sequences

Daniel Jordan, Ivan Adzhubey, Shamil Sunyaev
Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Estimating the level of selective constraint in protein-coding sequences has important applications in designing and interpreting biological experiments and prioritizing variants for medical research, and predicting the phenotypic effects of genetic variation. Several tools exist that use parametric methods to estimate the rate of evolution, including PAML, MrBayes, phyloP, GERP++, and many others. In general, such tools are either focused on estimating the distribution of rates over large sequence regions (PAML, MrBayes), or are limited to nucleotide sequences and fail to account for sequence features of proteins (phyloP, GERP++). We report a new method, BARE (Bayesian Amino-Acid Rate Estimator), that augments a parametric estimator of substitution rate at a specific site with a site-specific estimate of the stationary state distribution of amino acids, given a multiple sequence alignment of amino acid sequences and a precomputed phylogeny. This method outperforms nucleotide-level parametric scores at the task of separating neutral and damaging mutations, which is an important application of such tools. Using this method, we have generated a whole-exome dataset of rate estimates computed using the MultiZ 100-vertebrate multiple sequence alignment. Comparing this dataset with similar datasets produced with nucleotide-level methods such as phyloP and GERP++ reveal systematic differences and similarities between protein-level and nucleotide-level conservation scores. This comparison demonstrates which physicochemical properties of amino acids matter most to protein evolution, as well as highlighting certain features that are equally difficult for protein-level and nucleotide-level methods and may require a completely novel approach.

2.14

Using three-dimensional structure in inferring deep homology

Toni Daly [1,2], Andrew Sutherland-Smith[1], David Penny[1]
[1] Massey University, Palmerston North, New Zealand, [2] NorthTec, Whangarei, New Zealand

Primary DNA and protein sequences are typically used to build phylogenetic trees and infer evolutionary relationships. This approach has at least two limitations. Firstly, at the deepest levels of phylogeny the Markov models used are expected to saturate and lose information. Secondly, we need to understand much more about the nature and principles of living systems. In many cases substitutions, even deletions and insertions, may not significantly affect the protein structure at the tertiary and quaternary level. Structural homology is likely to be retained long after sequence homology has all but evolved away.

We used freely available servers (e.g. I-TASSER) to predict the structure of both extant proteins, and of calculated ancestral sequences. We use inferred ancestral sequences to find (BLAST) more distantly-related sequences than we could find previously and to infer ancestral 3D structures. We analyzed catalytic site geometry to assess conservation of function and used docking algorithms to determine the likelihood of oligomeric assembly. Our targets to date have been the Major Vault Protein (MVP) that spontaneously forms an ornate oligomeric vault particle, and the Argonaute/PIWI proteins whose expansion is in some cases astonishing (and puzzling) and whose functional capabilities are expanding with every publication.

We present a simple and inexpensive method of analyzing sequence and structural data to infer evolutionary relationships and resolve questions not amenable by BLASTing primary sequences alone. From this approach, protein sequences can be screened and ruled in or out before expensive lab work is undertaken.

2.15

How Evolution of Genomes Is Reflected In Exact DNA Sequence Match Statistics

Florian Massip [1,2], Michael Sheinman[2], Sophie Schbath[1], Arndt Peter[2]
[1] INRA, Jouy-en-Josas/Ile de France, France, [2] Max Planck Institute for Molecular Genetics, Berlin, Germany

Genome evolution is shaped by a multitude of mutational processes, including point mutations, insertions, and deletions of DNA sequences, as well as segmental duplications. These mutational processes can leave distinctive qualitative marks in the statistical features of genomic DNA sequences. One such feature is the match length distribution (MLD) of exactly matching sequence segments within an individual genome or between the genomes of related species. These have been observed to exhibit characteristic power law decays in many species. Here, we show that simple dynamical models consisting solely of duplication and mutation processes can already explain the characteristic features of MLDs observed in genomic sequences. Surprisingly, we find that these features are largely insensitive to details of the underlying mutational processes and do not necessarily rely on the action of natural selection. Our results demonstrate how analyzing statistical features of DNA sequences can help us reveal and quantify the different mutational processes that underlie genome evolution.

2.16

Statistical Properties of Pairwise Distances between Leaves on a Random Yule Tree

Michael Sheinman [1], Florian Massip[1,2], Peter Arndt[1]
[1] Max Planck Institute for Molecular Genetics, Berlin, Germany, [2] INRA, UR1077 Unite Mathematique Informatique et Genome, Jouy-en-Josas, France

A Yule tree is the result of a branching process with constant birth and death rates. Such a process serves as an instructive null model of many empirical systems, for instance, the evolution of species leading to a phylogenetic tree. However, often in phylogeny the only available information is the pairwise distances between a small fraction of extant species representing the leaves of the tree. In this article we study statistical properties of the pairwise distances in a Yule tree. Using a method based on a recursion, we derive an exact, analytic and compact formula for the expected number of pairs separated by a certain time distance. This number turns out to follow a increasing exponential function. This property of a Yule tree can serve as a simple test for empirical data to be well described by a Yule process. We further use this recursive method to calculate the expected number of the n-most closely related pairs of leaves and the number of cherries separated by a certain time distance. To make our results more useful for realistic scenarios, we explicitly take into account that the leaves of a tree may be incompletely sampled and derive a criterion for poorly sampled phylogenies. We show that our result can account for empirical data, using two families of birds species.

194A

Phylogenetic relationship of coelacanths, lungfishes, and tetrapods

Naoko Takezaki [1], Hidenori Nishihara[2]

[1] Kagawa University, Kitagun, Kagawa, Japan, [2] Tokyo Technology Institute, Yokohama, Kanagawa, Japan

Lobe-finned fishes are close relatives of terapods and coelacanths and lungfishes are the two extant lineages in this group of fishes. Despite the importance in understanding the origin of tetrapods the phylogenetic relationship among the coelacanths, lungfishes, and tetrapods has been controversial morphologically and in studies with molecular data. Recently, owing to the genome sequencing of coelacanth, two studies [Amemiya et al. 2013 (251 genes of 22 species with 100,583 amino acid sites); Liang et al. 2013 (1,290 genes for 10 species with 690,838 amino acid sites)] reconstructed the sister relationship of lungfishes and tetrapods with high statistical support. In this study we analyzed the data used in these studies and our newly collected data set (835 genes of 26 species with 242,475 amino acid sites). In our result the lungfish-tetrapod sister relationship is strongly supported by most of the phylogeny construction methods with the use of cartilaginous fishes as outgroup, as in the previous studies, whether or not ray-finned fishes are included. However, with the use of ray-finned fishes as outgroup excluding cartilaginous fishes the coelacanth-tetrapod sister relationship is most strongly supported, although the statistical supports tend to become weak and the lungfish-coelacanth and lungfish-tetrapod sister relationships cannot be excluded. This result indicates that the sister relationship of tetrapods and lungfishes is not firmly established and that the phylogenetic relationship among the coelacanths, lungfishes, and tetrapods is still to be resolved.

195B

Evaluating the impact of the tree prior on molecular dating

Andrew Ritchie, Nathan Lo, Simon Y. W. Ho
School of Biological Sciences, University of Sydney, NSW 2006, Australia

Sampling schemes incorporating multiple individuals per species are a common feature of phylogenetic studies. In particular, such mixed datasets are essential for ensuring sufficient coverage of species in groups with unknown or disputed species boundaries. However, these datasets can cause difficulties for molecular dating. Bayesian phylogenetic methods for estimating evolutionary timescales require the specification of a prior probability on the tree and divergence times. Current software packages typically offer time-tree priors that are based on either a speciation or a coalescent process. In the absence of tree priors designed specifically for mixed datasets, researchers must generally choose from among the aforementioned prior classes. It is at present unclear what impact this choice might have on dating results.

Here, we compare the results of Bayesian divergence time estimation on three empirical mixed datasets assuming different time-tree priors. We assess the fit of the most commonly-used tree models to real data and conduct simulations to investigate the effects of different sampling strategies. We make recommendations for choosing appropriate tree priors in molecular clock analyses and discuss implications for the sensitivity of molecular dating methods to the choice of tree prior.

196C

Estimating the evolutionary timescale of flowering plants using complete chloroplast genome sequences

Charles Foster, Simon Ho
University of Sydney, Camperdown, NSW, Australia

The origins and evolution of flowering plants (Angiospermae) have been major topics of research interest in phylogenetics. In particular, the evolutionary timescale of angiosperms has proven to be a source of considerable attention, although most studies have been based only on a small number of loci and/or taxa. However, the application of high-throughput sequencing techniques has produced substantial amounts of genetic data, with many chloroplast genome sequences now being available.

We estimated the evolutionary timescale of angiosperms by analysing sequences of whole chloroplast genomes. Our data set comprises published sequences from GenBank as well as novel data produced by collaborators at the Royal Botanic Garden, Sydney. In total, our data set comprises full chloroplast genomes from 198 taxa. We analysed this data set using a Bayesian phylogenetic relaxed-clock approach in MCMCTREE, which incorporates a fast algorithm designed for molecular dating of genome-scale data sets. Our analysis included a large number of fossil-based age constraints. We also investigated the robustness of our estimates by varying the number of independent molecular clocks, and the gamma priors for overall rates and rate variation across branches. Additionally, we investigated the effects of taxon sampling and gene sampling on date estimates for the angiosperm evolutionary timescale. Our analysis provides the most comprehensive and reliable estimate of the timescale of angiosperm evolution to date.

197D

A multilocus dataset for a recent newt radiation delimits taxa but fails to recover evolutionary relationships

Maciej Pabijan, Piotr Zielinski, Katarzyna Dudek, Michal Stuglik, Wieslaw Babik
Jagiellonian University, Krakow, Poland

Multilocous phylogenetic inference based on NGS datasets may be obscured by high levels of genealogical discordance. Dealing with multiple alleles per locus per species can be particularly problematic for recent radiations. We present an empirical study in subspecies of Lissotriton newts for which 74 markers (500 bp each) located in 3' UTR regions were sequenced for 128 individuals. On average, 12.4% (range 2.8-28%) of sites in each sequence alignment was parsimony informative. While the delimitation of lineages (clustering algorithm in STRUCTURE) was robust and supported by moderate to high exclusive ancestry for each lineage (ensemble genealogical sorting index 0.28-0.69; P<10-7), the relationships among lineages could not be confidently reconstructed. Species trees based on the multilocus coalescent (*BEAST), Bayesian concordance analysis (BUCKy) and a supermatrix approach (MrBayes) gave inconsistent results despite the implementation of subsampling strategies aimed at reducing the complexity of the dataset, and by treating individual alleles as operational taxonomic units. We hypothesize that (i) retention of ancestral alleles due to large effective population sizes and (ii) gene flow between some of the newt lineages, including non-sister lineages, may have overwhelmed the phylogenetic signal in the dataset. We assess the extent to which these demographic/confounding factors have contributed to the inconsistencies in our results using approximate Bayesian computation modeling.

198A

Speeding up tree likelihood computation using state aggregation

Iakov Davydov [1,2], Marc Robinson-Rechavi[1,2], Nicolas Salamin[1,2]
[1] *University of Lausanne, Lausanne, Switzerland,* [2] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*

A wide range of phylogenetic methods model the evolution of discrete traits using a continous-time Markov chain. The likelihood computations for these models is done using Felsenstein's tree pruning algorithm and most of the computation time is spent on matrix exponentiation and partial likelihood computations. The evaluation of the likelihood function has therefore a critical impact on the overall performance of phylogenetic methods.
We purpose here a method to speedup the evaluation of matrix exponentiation and partial likelihood by reducing the number of states in a continous-time Markov chain without loosing the dimensionality of the models. We used state aggregation techniques to selectively combine states of the instantaneous rate matrix. Depending on the particular model used, a number of aggregation strategies may be employed. Maximum reduction is achieved when all the states, unobserved at the tips of the tree, are aggregated into a single state.
We implemented the aggregation optimization in FastCodeML (Valle et al., 2014, Bioinformatics), which uses Branch-Site model (Yang and Nielsen, 2002, Mol Biol Evol) to infer positive selection along positions of a protein-coding gene. We use biological data as well as simulations to show that the proposed approximation does not lead to a bias in the parameter values or positive selection detection while giving a twofold speedup. We also measure the speedup with variable tree sizes and an alignment length and discuss applicability of the optimization for a number of phylogenetic methods.

199B

RiboDB : A dedicated database of prokaryoticribosomal proteins

Frédéric Jauffrit [1,2], Jean-Pierre Flandrois[1], Simon Penel[1], Jean-Philippe Charrier[2], Manolo Gouy[1], Céline Brochier-Armanet[1]

[1] Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France, [2] Technology Research Department, BioMérieux SA, Marcy l'Etoile, France

Phylogenies of prokaryotes relied mainly on the analysis of the RNA component of the small subunit of the ribosome or a small set of housekeeping genes. The resulting phylogenies have provided interesting but partial information on the evolutionary history of these organisms because the corresponding genes do not contain enough phylogenetic signal to resolve all nodes of the Bacteria and Archaea domains. Thus, many relationships, and especially the most ancient and the most recent ones, remained elusive.

The recent burst of complete genome sequencing projects have made a lot of protein markers available as an alternative to SSU rRNA. To assess systematic and taxonomic issues, these markers should be largely conserved across prokaryotic lineages, rarely transferred, and harbor a robust and reliable phylogenetic signal at various taxonomic levels. Among protein markers, ribosomal proteins fulfill most of these criteria. In addition, mass spectrometry (MALDI-TOF) studies showed that ribosomal proteins can be used to discriminate bacterial species. It is worth noting that the phylogenetic signal contained in r-proteins is a good proxy of the phylogenetic signal contained in larger sets of conserved core genes, while allowing applying ML and Bayesian approaches in acceptable computational time.

Here we present RiboDB, a database of prokaryotic ribosomal proteins. RiboDB is built from the automated reannotation of publicly available complete and assembly genomes using an dedicated in-house annotation engine. It's first applications to prokaryotic phylogeny and systematics will be presented and compared to other whole genome methods.

200C

Distinguishing evolution of coding sequences at three layers : nucleotides, codons and amino acids

Fanny Pouyet, Marc Bailly-Bechet, Laurent Guéguen
Laboratoire de Biométrie et Biologie Evolutiive, Villeurbanne, France

Redundance of the genetic code does not lead to a uniform or a random usage of the synonymous codons. How this bias arose and is maintained in a set of genes between species, is not often taken into account to resolve phylogenetic trees. To perform better estimation of trees, we need an evolutionary model that studies codon usage.

We develop such a model, inspired from Yang and Nielsen [Yang and Nielsen (2008)]. It is available in Bio++ suite [Guéguen et al. (2013)]. Our model untangles evolution of coding sequences at three layers : nucleotidic, codons and amino acids (AA). It explicitely describes separately mutational bias and BGC which applies on all nucleotides independently of their position in the codon, the selection between synonymous codons and the preferences among AA. We argue that synonymous substitutions are no longer neutral and our model performs new estimates for AA changes.

We apply our model in an homogeneous and non-stationary context, in a maximum likelihood framework. We study the forces that drive the genomic content in the core genome of twenty pathogens bacterias [Lassalle et al. (2015)]. We compare our model to classical codon models [Yang and Nielsen (1998)]. We observe a global AT enrichment in every genomes which agrees with Hershberg and Petrov [Hershberg and Petrov (2008)]. We confirm the existence of both a universal mutation bias towards AT and selection pressure on codon usage, and are able to measure the relative importance of each one in the evolutionary process.

201D

A Flocking Algorithm for Isolating Congruent Clusters of Genes in Phylogenomic Datasets

Apurva Narechania [1], Richard Baker[1], Rob Desalle[1], Barun Mathema[2], Barry Kreiswirth[2,4], Paul Planet[1,3]

[1] Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA, [2] Public Health Research Institute Center, Rutgers New Jersey Medical School, Newark, New Jersey, USA, [3] Department of Pediatrics, Division of Pediatric Infectious Diseases, Columbia University, College of Physicians and Surgeons, New York, NY, USA, [4] Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

Collective animal behavior such as the flocking of birds or the shoaling of fish has inspired a class of algorithms designed to optimize distance-based clusters in a wide range of applications including document analysis and DNA microarrays. In the flocking model, individual agents only respond to their immediate environment and move according to a few simple rules. After several iterations the agents self organize and clusters emerge without the need for partitional seeds. In addition to its unsupervised nature, flocking offers several computational advantages including the potential to decrease the number of required distance comparisons. Here, we use a flocking approach for the computationally intensive process of finding groups of genes that share a common history. Gene flocks are determined using pairwise calculations of phylogenetic incongruence. We test the approach on a known large-scale hybridization event in strains of Staphylococcus aureus, and then expand our analysis to genomes from the species Mycobacterium tuberculosis and Streptococcus agalactiae, organisms that are predicted to have very different rates of horizontal gene transfer and recombination. We show that an incongruence analysis grounded in the flocking model can recover areas of hybridization and clusters of genes with common phylogenetic signal without the need to estimate the number of groupings or designate a distance cutoff for inclusion. This approach can be used to discover horizontally transferred genes, recombined areas of the chromosome, and the genes that comprise the phylogenetic core.

202A

Scalable alignment-free phylogenomics

Guillaume Bernard, Cheong Xin Chan, Mark A. Ragan
Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

Multiple sequence alignment (MSA) is a key step in phylogenetic analysis. However, its assumption that full-length contiguity is conserved among homologous sequences does not agree well with the highly dynamic molecular mechanisms in genome evolution, including genome rearrangement, insertions/deletions and lateral genetic transfer (LGT). The computational complexity of MSA also limits its scalability to multi-genome data. This calls for the development of scalable techniques for phylogenomic analysis. So-called *alignment-free* (AF) methods based on statistical comparison of short sub-sequences (e.g. *k*-mers) could be an alternative to MSA in phylogenetic inference, but their sensitivity to different biological scenarios, and their scalability to large data, are little known. In this study, we used both simulated and empirical data to systematically assess the sensitivity of available AF approaches to a variety of evolutionary scenarios, and to examine their scalability to large data. Interestingly, we found that AF approaches are more sensitive than the standard MSA-based approaches to sequence divergence and LGT, but more robust to insertions/deletions, compositional biases and genome rearrangement. Our findings demonstrate that these approaches are faster and less computationally intensive than the standard approach. I will present some of these results, including our novel AF approach to infer phylogenetic relationships (including node support via Jackknifing techniques), and how AF approaches can be applied to understand the evolutionary dynamics of microbial genomes.

203B

Can we use structural information in evolution?

Cedrik Magis [1], Maria Chatzou[1], Damien de Vienne[2], Cedric Notredame[1]
[1] Center for Genomic Regulation (CRG), BARCELONA, Spain, [2] CNRS Biometry and Evolutionary Biology (LBBE), LYON, France

The study of biological sequences currently heavily rely on Multiple Sequence Alignment (MSA), a first step towards many other analysis, including phylogenetic tree reconstruction. While many different methods and algorithms were developed, the available sequence information was growing nearly without boundaries, and new challenges arise when trying to apply alignment strategies to large number of sequences ("Big Alignment"). One of the main issue is to be able to evaluate the accuracy of large scale alignment, from which all downstream analysis will be performed. Commonly, aligners are evaluated on benchmarks or reference alignments; however, they were not designed for "Big Alignments" and the accuracy measurement of an MSA containing a wide range of very diverse sequences may be unreliable. On the other hand, structural information, in a lesser extent, was also growing fast, in quantity and diversity; structures are available for a wide range of different species and families. This information should be used in the current framework to assess its efficiency and accuracy. We are investigating the effects of using protein structures at the different steps of the whole procedure, ranging from the alignment step, the alignment evaluation through new generated benchmarks (scalable and reproducible), and the phylogenetic reconstruction taking into account structural information, more resilient than its sequence counterpart. The use of extra information, such as structure, might open new perspectives in order to improve, refine and/or evaluate biological sequence analysis.

204C

Systematic error and gene tree heterogeneity in the quest for the identification of the closest living relative(s) of tetrapods

Iker Irisarri, Axel Meyer
University of Konstanz, Konstanz, Baden-Wuerttemberg, Germany

Finding the closest living relative of tetrapods has been a long-standing open question. Latest studies, including a phylogenomic analysis with the coelacanth genome, favored lungfishes (and not coelacanth) to be closest to tetrapods. However, even using genome-scale data, topology tests cannot reject the alternative where lungfishes and coelacanth are sister taxa, equally close to tetrapods. The internode separating coelacanth, lungfishes and tetrapods has been difficult to resolve due to its age (>400 my) and short length. In such a situation, using genome-scale data, maximizing phylogenetic signal and accounting for gene tree heterogeneity can contribute to resolve this controversy. Given lungfishes' huge (50-130 Gb) and repetitive genomes – that greatly complicate their sequencing – RNA-seq was used to obtain genomic data from all thee living lungfish lineages. Additional lungfish genomic information is key in resolving this question and can reduce possible long-branch attraction artifacts. The new data (representing the Australian, African, and South American lungfishes) are analyzed together with many vertebrate genomes using both concatenation and species-tree methods that account for conflict among gene trees. Gene information content and factors that can bias phylogenomic estimates due to systematic error are studied. Our results demonstrate that model misspecification plays a major role in estimating the correct species relationships. Lungfishes are recovered closest to tetrapods, but only when among-site heterogeneity in the replacement process is accounted for. We further studied amino acid replacements at the base of the sarcopterygian tree. Multispecies coalescent methods found a congruent topology despite pervasive conflict among gene trees.

205D

Moving beyond black box, GTR models in phyogenetic analyses through the use of mechanistic models of sequence evolution.

Michael Gilchrist [1,2], Jeremy Beaulieu[2], Juan Juan Chai[2], Brian O'Meara[1,2]
[1] University of Tennessee, Knoxville, TN, USA, [2] National Institute for Mathematical Biology, Knoxville, TN, USA

Models of sequence evolution trade off realism, such as having different transition rates between different amino acids, with simplicity. As a result, the standard GTR transition matricies lead to biologically absurd interpretations. As an alternative, we introduce a new Markov model that explicitly includes the effects of mutation bias, genetic drift, and natural selection for an optimal amino acid at each site within a protein. Because our model is mechanistic nature, this substantial increase in realism is obtainable using only a small number of additional parameters. Analysis of a multi-locus yeast data set shows that our new model provides a substantially better fit data than the standard empirical models, more realistic behavior over time, and to allows researchers to estimate biologically meaningful parameters such as the sensitivity of a protein's function to an amino acid substitution and the optimal amino acid at a given site.

206A

Towards a gene tree aware reconstruction of early animal relationships

Walker Pett, Nicolas Lartillot
LBBE - Université Claude Bernard, Lyon 1, Lyon, France

A well-supported phylogenetic hypothesis of the earliest animal lineages, including sponges, ctenophores, cnidarians and placozoans, is crucial to an understanding of the last metazoan common ancestor. However, the relationships among these groups are among the most controversial in all of animals. Several recent phylogenomic studies have highlighted the difficulties associated with the use of traditional phylogenetic methods for resolving deep metazoan relationships. We find that super-matrix methods based on EST amino acid data are sensitive to model and outgroup selection, and do not always agree with results based on rare genomic changes, such as changes in gene content and synteny. For this reason, competing hypotheses continue to be debated regarding which animal lineage diverged first. In attempt to uncover the source of this conflict, here we present a novel gene-tree within species-tree phylogenetic model for analyzing gene content based on EST data, and apply it to the problem of reconstructing the genome of the last metazoan common ancestor.

207B

The Open Tree of Life - Curating, synthesizing, and updating phylogenetic information across 1.8 million taxa

Emily Jane McTavish [2,1], The Open Tree of Life Project[2]
[1] Heidelberg Institute for Theoretical Studies, Heidelberg, Germany, [2] University of Kansas, Lawrence, KS, USA

An understanding of species' evolutionary relationships is crucial for making conclusions about biological processes. Since Darwin, considerable research and resources have been devoted to understanding the tree of life. Through this work phylogenies have been generated for taxa across much of the diversity of life. However, there have been several barriers to downstream application of these phylogenetic estimates by other researchers. Many trees are not available in a re-usable format, for example, following publication trees may be available only as figures, or as files with obscure labels. A single tree may not contain all species of interest. In some groups taxonomy still captures our best understanding of species relationships. As well, estimating very large phylogenies is a major computational endeavor, therefore performing full re-estimates of relationships as new taxa are sequenced is often not possible.
The Open Tree of Life project is an effort to reduce these impediments to accessing phylogenetic information. We are a collaboration building infrastructure to synthesize, update and share a comprehensive tree of all 1.8 million named species. I will discuss three major contributions of this project: An open database and community curation of phylogenies; synthesis of a single tree from many phylogenetic estimates using taxonomy as a scaffold; and development of approaches to update large phylogenies with new sequence information. These contributions together create a growing tree of life, which is not static but rather can be continually revised as new data become available.

208C

Fitting evolutionary rate models to complex phylogenies

Anais Brown, Fabia U. Battistuzzi
Oakland University, Rochester, MI, USA

A chronological framework of life is fundamental to the reconstruction of life's major evolutionary steps. While estimations of timetrees are now a common step in phylogenetic studies, questions remain about the accuracy of these timelines in light of the multiple assumptions inherent to molecular clock approaches. Here we investigate one of these assumptions, the model used to estimate rate variation among branches in the presence of complex evolutionary processes that are characteristic of large phylogenies. Focusing on the two most common rate models (autocorrelated, AR, and uncorrelated, UR), we compare the distributions of ancestor-descendant rate changes in simulated and empirical data to determine their fit in phylum-level and class-level phylogenies. As expected, we find variable rates among branches but no significant clustering among groups sharing the same common ancestor, even for closely related lineages. Additionally, we find that patterns of rate changes do not uniformly follow either the AR or UR model, but rather can be better explained by a combination of the two. These results suggest caution when applying these assumptions in divergence time estimations and encourage the use of molecular clock methods that implement fewer assumptions to derive timeline estimations.

209D

Considering MSA uncertainties in phylogeny inference

Haim Ashkenazy, Itamar Sela, Tal Pupko
Tel Aviv University, Tel Aviv, Israel

Multiple sequence alignment (MSA) is often the first step in phylogenetic inference. Specifically, most current analyses rely on a single MSA and it was previously shown that in many cases there are substantial amount of errors in the reconstructed MSA. Such errors might affect and bias phylogenetic inference. It was therefore suggested that considering uncertainty of the MSA might improve phylogenetic inference. While the common practice often involves filtering out unreliable positions from the MSA prior to the phylogenetic inference, we suggest a different approach to account for uncertainties in the MSA.

Here, GUIDANCE2 methodology is used to produce alternative MSAs accounting for (1) uncertainty of the guide tree used in progressive alignment algorithms, (2) uncertainty due to different co-optimal solutions and (3) uncertainty due to different possible gap penalty scores used during the alignment. We demonstrate, using both simulated and real data, that weighting positions of the MSA by using such alternative MSAs can significantly improve maximum likelihood (ML) based inferred phylogeny. Moreover, we suggest that considering such alternative MSAs explicitly, as additional information (e.g. by concatenating them to the default MSA) holds promise to further improve ML phylogenetic inference.

210A

Concatabominations: identifying unstable taxa in phylogenomic analyses

Karen Siu-Ting [1,2], Christopher Creevey[1], Wasiu Akanni[3], Davide Pisani[4], Mark Wilkinson[3]
[1] IBERS, Aberystwyth University, Aberystwyth, UK, [2] School of Biotechnology, Dublin City University, Dublin, Ireland, [3] Department of Life Sciences, Natural History Museum, London, UK, [4] School of Biological Sciences, University of Bristol, Bristol, UK

Rogue taxa can be phylogenetically unstable because of limited and extensive missing data. Their inclusion in phylogenetic reconstructions often leads to multiple trees, unresolved consensus topologies and an increase in run times. Safe Taxonomic Reduction (STR) has been used, with varying success, as an *a priori* method to determine the taxa that are potentially unstable and safe to exclude (in the sense that their exclusion can have no impact upon relationships inferred among the remaining taxa) from an analysis. We recently developed an heuristic extension to STR and illustrated its application to morphological data. Here we will show its potential in genomic studies (using both supermatrix or supertree approaches). This approach differs from, and is more powerful than, current alternatives based only on analysis of taxon presence/absence because it exploits information present in individual gene trees. A pipeline implementing the approach allows visualisation of taxonomic equivalence relations as connections between taxa in a network. We assessed the performance of this method using genomic MRP (Matrix Representation with Parsimony) datasets. The approach can substantially decrease computational burden and increase the resolution of the resulting phylogenies.

211B

Efficient Bayesian Evolutionary Analysis using Hamiltonian Monte Carlo

Arman Bilge [1,2], Timothy Vaughan[1,2], Alexei Drummond[1,2]
[1] Department of Computer Science, University of Auckland, Auckland, New Zealand, [2] Allan Wilson Centre for Molecular Ecology and Evolution, Palmerston North, New Zealand

Bayesian statistical inference via Markov chain Monte Carlo (MCMC) has transformed the field of phylogenetics by enabling the use of complex models for studying evolutionary processes. However, the standard Metropolis-Hastings MCMC algorithm, which utilises a random walk to draw samples from the posterior distribution, performs poorly in the high-dimensional parameter space associated with large datasets. Thus, an improvement on MCMC could dramatically improve our ability to work with large genomic data. The Hamiltonian Monte Carlo (HMC) algorithm has been shown to outperform MCMC on certain problems by avoiding random-walk behaviour. It uses the first-order gradient of the posterior distribution and a vector of momentum variables to simulate Hamiltonian dynamics and make distant proposals that are likely to be accepted, although at greater computational cost than an MCMC proposal. We applied the HMC algorithm to Bayesian phylogenetic inference and compared its performance to that of MCMC for various problem sizes and complexities. The HMC implementation involved developing efficient but accurate methods to compute the gradient for several popular models in phylogenetics. Because HMC is applicable only to differentiable spaces, we focused on inferring times and parameter values for fixed tree topologies. We discuss the cases where the use of HMC yielded a significant advantage over MCMC and the tuning necessary to achieve these advantages. Future work will focus on methods for auto-tuning the HMC algorithm as well as developing a variation on HMC (or other intelligent proposal mechanisms) that can effectively explore treespace.

212C

Phylogenetic placement of ancient genomes

Jeffrey Spence, Jonathan Terhorst, Yun Song
University of California, Berkeley, Berkeley, CA, USA

When DNA is sampled from an individual from an unknown species, understanding how that species is related to known species is a fundamental and important question. When the sample comes from an ancient specimen, the resulting ancient DNA (aDNA) sample poses an interesting challenge for this phylogenetic placement problem. While gigabases of DNA data are generated by state-of-the-art aDNA sequencing techniques, the overwhelming majority of reads come from contaminating modern and ancient bacterial species. Contaminating modern human DNA and the rapid degradation of aDNA also pose problems for existing phylogenetic placement methods. Furthermore, since the sample is from an unknown species, no reference genome exists for the sampled individual.

We explore the impact of the difficulties associated with aDNA phylogenetic placement and develop several techniques for overcoming them. We combine these techniques in a coalescent-based method to address phylogenetic placement specifically for aDNA samples. Our algorithm operates directly on reads and so obviates the need for a reference genome for the unknown species. We test the method on simulated and real data and show that it accurately infers the topology and the join on time of the unknown species using reads from a single individual.

213D

Phylogenomic insights into genome evolution and speciation in mice

Brice Sarver, Jeffrey Good
University of Montana, Missoula, Montana, USA

Comparative genomic data can offer powerful insights into molecular evolution, adaptation, and speciation. While most comparative genomic studies have relied upon contrasts within one or among a few species, the increasing ease of generating genome-scale datasets now enables comparisons across many. These increasingly complex datasets necessitate a strong phylogenetic framework. However, phylogeny estimation is complicated by the diverse forces (selection, gene flow, and incomplete lineage sorting) that generate phylogenetic discordance across the genome. Here, we used targeted capture to sequence the entire exome (55Mb) of 11 species of mice spanning the *Mus* radiation. We employ an iterative mapping-consensus approach that allows us to mitigate reference bias while maintaining the detailed annotation provided by the finished house mouse reference genome. We then resolve the evolutionary relationships among these species. The validity of our estimate is assessed using species tree approaches with and without random binning of loci. We evaluate the relative contribution of incomplete lineage sorting and introgression to patterns of phylogenetic discordance across the genome. Finally, we assess the influence that phylogenetic history has on patterns of molecular evolution and the inference of positive selection across the genome. Our work highlights both the computational challenges and powerful insights into genome evolution and phylogenetic history afforded by phylogenomic data.

214A

Evolution as a Communication Channel

Ercan E. Kuruoglu [1,2], Peter F. Arndt[2]
[1] *ISTI-CNR, Pisa, Italy,* [2] *Max Planck Institute for Molecular Genetics, Berlin, Germany*

We present a mathematical framework based on the Information Theory of Claude Shannon for modelling the evolution as an information transmission process. The evolution is visualised as a communication system in which the message being sent is the genetic code. Due to the mutations, errors are introduced on the message sent. These errors define the noise in the communication system analogy. Using various substitution matrix models in the literature we calculate the capacity of the evolution channel for both non-coding RNA and the protein coding DNA. We calculate the capacities also based on experimental data for both types of channels. The capacities calculated give us the theoretical limits of information transferable via genetic code without error and hence the number of generations (or unit time) required for the alteration of the sent message. The difference between protein coding and non-coding channels gives us the measure of the robustness to mutations provided by the protein code.

215B

Molecular phylogenetics and the biogeographic dilemma of Marsupial Mammals

Lucas Marques, Alexandre Pedro Selvatti, Claudia Russo
Universidade Federal do Rio de Janeiro, Instituto de Biologia, Departamento de Genética, Rio de Janeiro, RJ, Brazil

Marsupial species are generally classified on two major cohorts: Ameridelphia and Australidelphia. Several molecular analyzes have confirmed this hypothesis, but relationships between orders and families remain still controversial in the group. Of particular interest is the position for the South American order Microbiotheria that clusters either as sister group or within the Australian clade, which is critical to understand the early radiation of the marsupials. In this study, we are exploring key events on the diversification of the group by assuming genera monophyly in order to increase matrix-filling levels. The final alignment had 18% gap proportion assembled with sequences from 35 mitochondrial and four nuclear *loci*, representing 91 recent marsupial genera. For the phylogenetic analysis, we have used the maximum likelihood method implemented on the program RAxML and bootstrap test was used to verify reproducibility levels. Our topology recovered the monophyletism of all major marsupial orders usually with a moderate or high support and confirmed the paraphyletic status for Ameridelphia. Thus, we now provide more evidence to support a South American origin for the Australian lineage. Also, the South American Microbiotheria order was recovered within the Australidelphia cohort. In this scenario two hypotheses are likely to explain the early radiation in Australasia. The first includes a single invasion event followed by a secondary dispersion of Microbiotheria back to South America whereas the second considers multiple invasion events in Oceania with Microbiotheria remained in South America.

216C

Failure of phylogenetic tree reconstruction by heterogeneous substitution pattern among lineages and the method for its recovery.

Eisuke Iwamoto [1], Koichiro Tamura[1,2]

[1] Department of Biological Sciences, Tokyo Metropolitan University, Tokyo, Japan, [2] Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Tokyo, Japan

In molecular phylogenetic analyses, phylogenetic trees are inferred from molecular sequence data such as nucleotide sequences and amino acid sequences on the basis of mathematical model of molecular evolution. Recently, with the rapid development of sequencing technology, data sets consisting of many genomic regions are available. In such cases, the substitution pattern is usually assumed to be constant through evolutionary processes within the individual genomic regions. When this assumption is violated, the accuracy of phylogenetic tree reconstruction may deteriorate. In this study, we examined the influence of substitution pattern variations among lineages on phylogenetic tree reconstruction using computer simulations and then developed a method to improve the accuracy of phylogenetic tree reconstruction by identifying and removing the regions where substitution pattern has altered in some lineages. In this method, we (1) reconstruct an initial tree from a distance matrix computed from all the sequences, (2) compute the patristic distance matrix from the initial tree and (3) identify and remove the genomic regions showing high discrepancy between the original and patristic distance matrices. Our computer simulations demonstrated that removal of such genomic regions improves the accuracy of phylogenetic tree reconstruction when the total number of sites was 10,000 or more. However, when the total number of sites was 2,000 or less, the improvement by this approach was not significant. Furthermore, we applied this method to real sequence data.

217D

Choosing subsamples for sequencing studies by minimizing the average distance to the closest leaf

Jonathan Kang [1], Peng Zhang[2], Sebastian Zöllner[3], Noah Rosenberg[1]
[1] Stanford University, Stanford, CA, USA, [2] Johns Hopkins University, Baltimore, MD, USA, [3] University of Michigan, Ann Arbor, MI, USA

Geneticists routinely make choices about which individuals, strains, or populations merit prioritization for genotyping or DNA sequencing, and techniques based on phylogenetic principles can assist in these experimental design decisions. Here, we consider two phylogenetically derived algorithms for prioritizing samples for use in genotype imputation studies in human genetics.

Dense imputation of genotypes in a sample of individuals genotyped at a relatively small number of markers can proceed from use of the linkage disequilibrium pattern in a set of sequenced samples. But which individuals should be sequenced? Beginning from tree diagrams relating a set of DNA sequences, we introduce a new method for selecting samples for sequencing—minimizing the average distance to the closest leaf (ADCL)—and compare its relative performance to an earlier algorithm, namely, maximizing phylogenetic diversity (PD). Employing both simulated data and sequences from the 1000 Genomes Project, we show that ADCL provides a significant improvement to imputation accuracy, especially for the imputation of sites with low-frequency alleles. We find that this improvement in imputation accuracy is robust to changes in reference panel size, marker density, and target imputation length.

218A

CVG: core vertebrate genes for genome and transcriptome assembly completeness assessment

Yuichiro Hara, Shigehiro Kuraku
CLST, RIKEN, Kobe, Japan

Assessment of *de novo* assemblies is cumbersome in that suitable metrics have to be chosen from numerous benchmark tests with varied objectives. Among these benchmarks, coverage of protein-coding genes has been employed as an essential one. Identification of orthologs to highly conserved genes across wide-ranged taxa is a standard of the coverage evaluation, which is accomplished by the CEGMA gene prediction pipeline referring to the Core Eukaryotic Genes (CEGs). This approach provides a common measurement of 'completeness' among assemblies from different species. On the other hand, this approach potentially misidentifies paralogs as orthologs, leading to overestimation of gene coverage. To avoid this problem, a subset of the CEGs composed of no or minimal duplicates (248 CEGs) has been selected and used in conventional completeness assessment, but our examination revealed quite a few cases leading to overestimation. In this study, we introduce Core Vertebrate Genes (CVGs), a new reference gene set composed of 233 one-to-one ortholog groups strictly confirmed with 29 vertebrate genomes. The gene set includes a cyclostome and a chondrichthyan belonging to the deep-branching taxa. Using the 233 CVGs, we evaluated gene coverage of embryonic transcriptome assemblies of Madagascar ground gecko (*Paroedura picta*) built with multiple strategies. The result demonstrated that the evaluation referring to the 233 CVGs achieved higher accuracy and resolution than that with the 248 CEGs. The assessment with CVGs is applicable to *de novo* assemblies of any vertebrates, and this approach is also useful in assembly assessments in other lineages.

219B

Accounting for Sequence Error in Maximum Likelihood Phylogenetic Inference

Alexey Kozlov [1], Alexandros Stamatakis[1,2]

[1] Heidelberg Institute for Theoretical Studies (HITS gGmbH), Heidelberg, Germany, [2] Karlsruhe Institute of Technology, Karlsruhe, Germany

In the Maximum Likelihood (ML) framework, the phylogenetic tree is typically estimated from a multiple sequence alignment (MSA) of the taxa under study. MSAs are usually deterministic, that is, character states at each MSA site are assumed to be known with certainty. This is an over-simplification, because character states can exhibit uncertainty due to (i) sequencing error, (ii) the read assembly process and (iii) the MSA procedure. In practice, one usually resorts to ad hoc filtering methods to eliminate such errors (e.g., skipping or trimming low-quality reads, using "undetermined" characters for positions with poor consensi etc.). This standard approach is sensitive to threshold selection and can lead to loss of potentially useful information.

To this end, we assess if explicitly modeling such errors improves the accuracy of phylogenetic inferences. Therefore, we modified RAxML to support tip sequence uncertainty. We experimented with both global (per-alignment) and per-site error rates. These error rates can either be specified by the user or explicitely estimated from data. On simulated alignments, we show that, incorporating sequence uncertainty yields more accurate branch length estimates. This effect is most pronounced when the phylogenetic signal is weak and the error rate is high.

Currently, we are evaluating our approach for several real-world scenarios, where the expected error is relatively high: evolutionary placements of short reads (EPA), RADSeq marker analyses, and amplicon-based metagenetic studies.

220C

Estimating genomic introgressions across mouse species

Emre Karakoc, Diethard Tautz
Max Planck Institute for Evolutionary Biology, Ploen,Schleswig - Holstein, Germany

Allelic introgression from other subspecies or closely related species, is an important source of adaptive evolution and plays a crucial role for understanding the genetic composition of natural populations. While there are examples of hybridization in animals, there is still a few examples of introgression of chromosomal regions due to the difficulty of differentiation introgression from parallel evolution, incomplete lineage sorting and balancing selection with ancestral polymorphisms. Although the introgression can be neutral and does not affect the phenotype, it can also be adaptive and affect the phenotypes in the form of adaptive evolution. It is great interest to us to detect regions of introgression in animal genomes given the great consequences of it in evolutionary biology and speciation. We hypothesize that the introgression plays a major role in the adaptive genomic landscape of natural populations and we aim to develop computational methods to systematically analysis the recently abundant next generation sequencing data to detect introgression regions at the whole genome scale.

221D

Recombination rate governs patterns of molecular evolution in the collared flycatcher by an interplay between selection and GC-biased gene conversion.

Paulina Bolívar, Carina F. Mugal, Alexander Nater, Hans Ellegren
Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University., Uppsala, Sweden

The ratio of synonymous (d_S) to non-synonymous (d_N) substitutions provides information on the evolutionary processes driving sequence evolution of protein coding genes and is commonly applied to infer natural selection. However, sites within genes are not independent units but are linked to each other. This linkage between targets of selection leads to interference, referred to as Hill-Robertson Interference (HRI), which reduces efficacy of selection. Recombination breaks down the linkage between sites and thereby increases the efficacy of selection. However, recombination may also affect substitution patterns by means of GC-biased gene conversion (gBGC), a process that leads to the preferential fixation of "*strong*" (G or C) over "*weak*" (A or T) nucleotides in the proximity of a recombination event. Here, we explore the impact of recombination on patterns of molecular evolution in the collared flycatcher (*Ficedula albicollis*) genome. Capitalizing on a high-density genetic linkage map and whole-genome re-sequencing data from population samples, we are able to disentangle the role of HRI and gBGC in shaping patterns of divergence and diversity in protein coding genes. Our results indicate that signatures of gBGC are pervasive in the collared flycatcher genome, probably as a result of high recombination rates, a stable recombination landscape and a large effective population size. Therefore, we advise caution when making inferences of natural selection in avian species based on the d_N/d_S ratio as this test statistic may be strongly affected by gBGC.

222A

R-opsin evolution: one or two families?

James Fleming [1], Davide Pisani[1], Nicholas Roberts[1], Ronald Jenner[2]
[1] *University of Bristol, Bristol, UK,* [2] *Natural History Museum, London, UK*

Opsins mediate light perception in animals. In invertebrates, opsins belonging to the rhabdomeric group (r-opsin) subfamily are expressed in the eye and used for image-forming vision. Duplications of r-opsin genes can be tracked to uncover the evolutionary history of image forming vision in protostomes and colour vision in arthropods such as spiders, insects and crustaceans. Until recently it was assumed that there was only one group of r–opsins. However, the existence of a second subfamily of "r–opsin-like" sequences has recently been proposed, following the identification of sequences from the molluscs and most importantly from the lancelet (Branchiostoma). These seem to cluster within the arthropsins - a group of opsins that were previously assumed to be arthropod-specific r–opsins. Accordingly, it has been suggested that the arthropsins might represent the sister group of the entire r-opsin subfamily. If this is true, then the arthropsins would have emerged from the duplication of a single gene in the common ancestor of the r-opsins and arthropsins in a stem bilaterian lineage. However, whether the arthropsins represent a real family or a long branch attraction artifact remains unclear. Here we shall test these hypotheses to clarify r–opsin evolution and the origin of vision in protostomes more broadly.

223B

Conserved Ortholog Set Genes across Gymnosperms as Effective Markers for Phylogenomics in the Seed Plants

Zhen Li [1,2], Lieven Sterck[1,2], Yves Van de Peer[1,2]
[1] Department of Plant Systems Biology, VIB, Ghent, Belgium, [2] Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, [3] Department of Genetics, Genomics Research Institute, University of Pretoria, Pretoria, South Africa

Over the past few years, numerous genomic and transcriptomic resources have been constructed for gymnosperms, the genomes of which are difficult to assemble because of their tremendously large genome sizes compared with those of other eukaryotic species. Although most of the current sequences are from transcriptomes, comparative analyses based on transcriptomic data can provide important evolutionary insights in the absence of complete gymnosperm genomes. Comparative transcriptomics can also accelerate comparative genetic mapping between different gymnosperm species, especially between conifers, which are of primary interest for quantitative genetics because of their economical importance. To overcome the limits that, hitherto, only a few genetic markers were available in gymnosperms, we constructed a set of 3,072 conserved ortholog set (COS) markers in 31 gymnosperms out of 11,152 low-copy number gene families from six Pinaceae species, which either have genomes (*Picea abies* and *Pinus taeda*), high quality transcriptomes (*Picea glauca* and *Picea sitchensis*), or deep sequenced transcriptomes (generated by ProCoGen project for *Pinus pinaster* and *Pinus sylvestris*). Using a similar protocol as was applied for the gymnosperm analysis we were also able to identify 2,539 COS markers in the angiosperms (PLAZA 3.0), of which 1,468 are present in all seed plants. Moreover, 42 high quality phylogenetic markers of nuclear genes for the seed plants were selected to help resolve the phylogenetic position of gnetophytes and to elucidate the relationships within the order Pinales.

224C

Embracing Phylogenetic Uncertainty - Testing the Congruence of Sets of Trees.

Giddy Landan [1], David Bryant[2], Arnold Janssen[3], William F. Martin[3], Tal Dagan[1]
[1] Christian-Albrechts-Universität, Kiel, Germany, [2] University of Otago, Dunedin, New Zealand, [3] Heinrich-Heine University, Düsseldorf, Germany

The emergence of phylogenomics presents us with challenging new data: large sets of phylogenetic trees. In general, the trees will conflict with each other to some extent, and the question arises whether these incongruences are due to phylogenetic inference errors or to lateral gene transfer. Further complications very often arise: the trees have non-identical taxa sets, and there is no reliable reference tree to draw on. We present a similarity measure that quantifies the congruence of a given tree with an entire collection of trees, and which is sensitive to LGT events. We use this measure to develop a testing procedure for the congruence of two sets of trees. The test procedure entails the use of a new statistical methodology for unbiased pooling of layered data. We demonstrate our testing procedure for the case of two sets of trees over 10 Halobacteriales species. One set of trees represents the vertical inheritance pattern of the 10 species. This set serves two purposes: first, it takes over the role of a classical reference species tree, thus avoiding the need for a 'gold-standard' reference tree; and second, it captures the level of phylogenetic noise encountered in typical trees over this set of taxa. The second set consists of trees that are suspected of harboring rampant LGT events. We show that our testing protocol can tolerate substantial phylogenetic errors while being powerful enough to detect LGT.

225D

Inferences from 450 high coverage complete human genome sequences from 156 populations of predominantly Eurasian origin

Mait Metspalu
on behalf of all authors, Tartu, Estonia

Complete high coverage individual genome sequences carry the maximum amount of information for reconstructing the evolutionary past of a species in the interplay between random genetic drift and natural selection. Here we present a novel dataset of 450 human genomes sequenced at 40X coverage on the same platform (Complete Genomics) and uniform bioinformatic pipelines. Based on SNP-chip data we generally chose three samples to represent each population of interest. We cover a wide range of mostly Eurasian populations with additional populations from Oceania, South America and Africa.

Here we use Treemix, f3 and D statistics to explore the past demography of human populations by reconstructing the splitting and admixture patterns. We show that several analyses support a scenario of more than one Out of Africa events for anatomically modern humans. The genetic vestiges of the first event are detectable today mostly in Australian, Papua New Guinea and negrito populations from the Philippines. Importantly the evidence for the special genetic history of these populations is retained when we control for admixture with Denisova and Neanderthal man.

However, it should be also noted that while different analyses yield results that are in agreement in indicating a special relationship between the aforementioned populations and Africans, the nature of the specialness seems to vary. This suggest that early migrations of AMH into Eurasia and beyond may be more complex than a simple pattern of two migrations followed by admixture of the two in areas where thy met.

226A

Is convergence less common in protein sequences than morphological traits?

Zhengting Zou, Jianzhi Zhang
University of Michigan, Ann Arbor, Michigan, USA

Molecular phylogenies are often considered more reliable than morphology-based trees, largely due to the belief that convergent evolution, which confuses phylogenetic reconstruction, is much rarer at the level of DNA or protein sequences than at the level of morphology. Recent genomic studies, however, revealed abundant protein sequence convergence, casting doubt on a commonly perceived benefit of molecular systematics. Using 3,433 parsimony informative morphological traits and 5,843 parsimony informative amino acid sites of 46 mammals previously compiled for phylogenetic inference, we compare the frequency of convergence for the two types of traits, under both the morphological tree and molecular tree. We find the mean frequency of convergence for morphological traits to be over twice that for molecular traits, regardless of the tree topology assumed. Consistency index (CI) and that corrected by retention index (RI) are both significantly lower for morphological traits than for molecular traits. These results confirm that convergence is generally less common in molecular sequences than morphological traits, even though only those morphologies that are considered by systematicists to be suitable for phylogenetics are included here. Nevertheless, a small fraction of branch pairs exhibit more molecular convergence than morphological convergence, an intriguing phenomenon whose underlying cause is under active investigation.

3 The biological impact of transposable elements

3.1

SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age

Michael van Meter, Mehr Kashyap, Rezazadeh Sarallah, Anthony Geneva, Timothy Morello, Andrei Seluanov, Vera Gorbunova
University of Rochester, Rochester, NY, USA

L1 retrotransposons are an abundant class of transposable elements which pose a threat to genome stability and may play a role in age-related pathologies such as cancer. Recent evidence indicates that L1s become more active in somatic tissues during the course of aging; the mechanisms underlying this phenomenon remain unknown, however. We found that the longevity regulating protein, SIRT6, is a powerful repressor of L1 activity. Specifically, SIRT6 binds to the 5'UTR of L1 loci, where it mono-ADP ribosylates the nuclear corepressor protein, KAP1, and facilitates KAP1 interaction with the heterochromatin factor, HP1α, thereby contributing to the packaging of L1 elements into transcriptionally repressive heterochromatin. Interestingly, upon DNA damage SIRT6 leaves L1 promoters and relocalizes to the sites of DNA breaks. We also observe activation of L1 elements during aging and cellular senescence. This suggests a paradigm where during the course of aging chronic DNA damage and short telomeres accumulate, SIRT6 is redeployed to the sites of DNA damage, and the sleeping retrotransposons are left unguarded leading to transposon activation and a vicious cycle of increased DNA damage.

3.2

Transposable elements contribute to rapid adaptation in interspecific hybrids

Caiti Smukowski Heil, Maitreya Dunham
University of Washington, Seattle, WA, USA

Understanding the genetic changes by which organisms adapt to novel environments is a central goal of evolutionary biology. One possible mechanism is hybridization: hybridization may transport genetic variation from a related species that allows organisms to adapt to a new environment, or alternatively, hybridization may stimulate new variation to form within genomes, such as the mobilization of transposable elements (TEs). TE insertions can mediate a cascade of large scale changes in chromosome structure, gene expression, and gene content, thereby creating a rapid path to adaptation. To test this hypothesis, we created interspecific hybrids using the extremely tractable model system *Saccharomyces cerevisiae* and its relative *Saccharomyces uvarum*, utilizing a genetic background carrying ten times the wild type level of TEs for half of the populations. We then subjected these hybrids to hundreds of generations of selection to nutrient limited conditions and characterized insertion sites of TEs. After characterizing fitness and determining other mutations present, we used allele replacement techniques and competition assays to quantify the proportion of increased fitness attributable to TEs. We report on TE insertions mediating adaptive genomic changes such as copy number variants and loss of heterozygosity, transposition rate in hybrids versus nonhybrid populations, and finally, we determine whether hybridization provides a means for TEs to invade naïve genomes.

3.3

Frequency and impact of horizontal transfer of transposable elements from moth to virus

Clément Gilbert [1], Jean Peccoud[1], Aurélien Chateigner[2], Elisabeth Herniou[2], Richard Cordaux[1]
[1] UMR CNRS 7267 EBI, Université de Poitiers, Poitiers, France, [2] UMR CNRS 7261 IRBI, Université de Tours, Tours, France

Horizontal transfer (HT) of transposable elements is an important factor shaping animal genome evolution. However, the vectors and mechanisms underlying these transfers remain poorly understood. Using a population genomics approach, we show that when a caterpillar is infected by the baculovirus AcMNPV, a large diversity of caterpillar sequences can become integrated in viral genomes. We uncover 61 caterpillar sequences integrated in AcMNPV genomes, comprising at least 30 different TEs which belong to 6 superfamilies, as well as 14 non-TE sequences. Several junctions between caterpillar and viral genomes correspond to target site duplications, a hallmark of DNA transposition. The other junctions are characterized by the presence of microhomologies (1 to 23 bp) between caterpillar and viral sequences, suggesting that caterpillar sequences can be integrated in the viral genome by a form of end-joining repair mechanism. We also show that 12 of the 61 sequences integrated in viral genomes have undergone one or more HT events during the evolution of Lepidoptera, and that the frequency of caterpillar sequences in baculovirus populations can reach more than 4%. Together, these results support viruses as vectors of HT between animals and suggest integration of host TEs into viral genomes may have a measurable impact on viral replication.

3.4

Activities and Biological Impact of Mobile Elements in Canines

Julia Wildschutte [1], Aurelien Doucet[1], Robert Wayne[2], Carles Vila[3], Tomas Marquès-Bonet[4], Adam Boyko[5], John Moran[1,6], Jeffrey Kidd[1]

[1] University of Michigan, Ann Arbor, MI, USA, [2] University of California Los Angeles, Los Angeles, CA, USA, [3] Estación Biológica de Doñana, Sevilla, Spain, [4] Institut de Biologia Evolutiva, Barcelona, Spain, [5] Cornell University, Ithaca, NY, USA, [6] Howard Hughes Medical Institute, Ann Arbor, MI, USA

Mobile element insertions (MEIs) cause sporadic disease and phenotypic variation, and provide a repetitive substrate that seeds structural variation. The dog has emerged as an important model system for mapping traits relevant to human disease. Studies have linked canine MEIs to disease-associated variants in human orthologs, for example in narcolepsy and centronuclear myopathy. Akin to *Alu* and L1Hs in humans, the canine SINEC_Cf/L1_Cf pair has undergone recent expansion leading to thousands of dimorphic copies, indicating these elements remain active and continue to influence canine genome structure and evolution. To investigate the genomic impact of canine MEIs, we searched whole genome sequence data of ~90 canids, identifying >32,000 non-reference SINEC_Cf and >1,400 L1_Cf copies. *De novo* assembly of MEI-supporting reads per site reconstituted >19,000 SINEC_Cf, and permitted inference of ~140 full-length L1_Cf copies. We find ~7,184 intronic and 33 exonic SINECs and ~371 and ~9 intronic and exonic L1_Cf insertions, implicating potential biological impact. To assess activity, we have subcloned ~10 L1_Cf copies utilizing a fosmid library we constructed from a single breed dog. In standard retrotransposition assays, we demonstrate L1 is competent for retrotransposition in human cells. We also demonstrate retrotransposition of the canine L1 and its ability to mobilize SINEC_C, including disease-causing variants. We anticipate these data will provide a MEI variant resource for incorporation to future canine GWAS analysis and will be critical for understanding the biology of canine mobile element expansion, contribution to canine structural variation, and impact on canine phenotypes and evolution.

3.5

Exploring the phenotypic space and the evolutionary history of a natural mutation in *Drosophila melanogaster*

Anna Ullastres, Natalia Petit, Josefa González
Institute of Evolutionary Biology (Universitat Pompeu Fabra - CSIC), Barcelona, Spain

How mutations allow organisms to adapt to different environments is still an open question in Evolutionary Biology. Recently, transposable elements (TEs) have re-gained attention as an important source of mutations playing a crucial role in genome structure and regulation. In this work, we explored the evolutionary history, the phenotypic space, and the molecular mechanism of a previously identified adaptive TE insertion: the invader4 element FBti0019386. We first used several tests that capture different signatures of selection to show that there is evidence of positive selection in the regions flanking FBti0019386 insertion. We then explored several phenotypes previously associated with FBti0019386's nearby genes, and having plausible connections to fitness variation in nature. We found that flies with FBti0019386 insertion had a shorter developmental time and were more sensitive to stress, which are likely to be the adaptive effect and the cost of selection of this mutation, respectively. Interestingly, these phenotypic effects are not consistent with a role of FBti0019386 in temperate adaptation as was previously suggested. Indeed, a global analysis of the population frequency of FBti0019386 showed that clinal frequency patterns are found in North America and Australia but not in Europe. Finally, althoug FBti0019386 insertion could be inducing the formation of heterochromatin by recruiting HP1a (Heterochromatin Protein 1a) protein, the insertion is associated with upregulation of sra.
Overall, our integrative approach allowed us to functionally characterize the biological impact of FBti0019386. Our work highlights the complexity of mapping genotype to phenotype, and exemplifies how TE-induced mutations influence genome evolution.

3.6

Transposon landscape dynamics and the evolution of Piwi-interacting RNA (piRNA) Clusters.

Reazur Rahman[1], Gung-wei Chirn[1], Yuliya Syntnikova[1], Jessica Matts[1], Mei Zeng[1], Daniel Gerlach[2], Michael Yu[3], Bonnie Berger[3], Benjamin Kile[4], Nelson Lau [1]

[1] Brandeis University Dept of Biology, Waltham, MA, USA, [2] Vienna Biocenter, Research Institute of Molecular Pathology, Vienna, Austria, [3] MIT, CS and AI Lab, Cambridge, MA, USA, [4] Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia

The Piwi pathway ensures animal fertility by serving as an effective repressor of transposon mobilization in the germ line. Nevertheless, transposons still remain a sizeable proportion of most model animal genomes. To molecularly explain this equilibrium, we show in Drosophila that transposon landscapes are dynamic and actively fluctuating despite active PIWI silencing. By examining the PIWI/piRNA complex targeting capacity on transposon targets, we propose that transposon-inherent promoters of transcription can stimulate the expression of new long-noncoding RNAs that perhaps antagonize PIWI silencing. Furthermore, we will show the pervasiveness of transposon landscapes divergent from the Drosophila reference genome in fly strains and cell cultures. We describe a tool called TIDAL that simplifies the analysis of individual transposon landscapes. Although the function of piRNAs in guiding Piwi proteins to transposon targets is deeply conserved, very few piRNAs are conserved at the sequence level, even amongst closely related animal species. To examine the evolution of piRNAs at the higher level of piRNA clusters, we discovered and compared piRNA clusters across various Drosophilids and mammals. We discovered that the evolution of most piRNA clusters is remarkably rapid, perhaps even faster than transposons that they are suppressing. Since the Piwi pathway has been compared as an adaptive-like immune system for the genome against invading transposons, the piRNA clusters loci can also be viewed as hyper-mutable regions like immunoglobulin loci. Our analysis suggests both adaptive and non-adaptive forces are acting on the evolution of piRNA clusters.

3.7

The evolutionary consequences of *piRNA*- mediated epigenetic silencing of transposable elements in *Drosophila melanogaster*

Grace Yuh Chwen Lee
Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

Several widely studied deleterious impacts of transposable elements (TEs), such as TE insertion into functional elements and ectopic recombination between nonhomologous TEs, are results of TE-induced physical disruption of DNAs. The potential functional and evolutionary consequences of TEs' epigenetic impacts have been largely unexplored. The *piwi*-interacting RNAs (*piRNAs*) are small RNAs that target selfish TEs in many animals. Euchromatic TEs can be epigenetically silenced via *piRNA*-dependent heterochromatin formation, which can further "spread" into nearby genes. However, the influence of TEs on the chromatin states and functions of nearby genes has not been explored on a genomic scale and from an evolutionary perspective. We hypothesized that the spread of *piRNA*-mediated heterochromatin of TEs to adjacent genes has deleterious functional impacts and leads to selection against individual TEs. Our genomic analyses found that one of the repressive chromatin marks, H3K9me3, is elevated in sequences and genes adjacent to euchromatic TEs in Drosophila. This association is likely *piRNA*-dependent and was observed to influence the expression levels of host genes. Importantly, we found stronger selection against TEs that lead to higher H3K9me3 enrichment of adjacent genes, demonstrating the evolutionary consequences of TE-induced epigenetic silencing. Our simulations further showed that, without other mechanisms removing TEs, selection against *piRNA*-dependent epigenetic impact of TEs could lead to stably contained TE copy number in host genomes. Our results suggest that the *piRNA*-mediated silencing of TEs not only leads to pervasive indirect deleterious impacts on hosts, but also plays a critical role in the evolutionary dynamics of TEs.

3.8

Transposable elements continuously remodel the regulatory landscape of mammalian endometrial stromal cells

Vincent Lynch
The University of Chicago, Chicago, USA

A major challenge in biology is explaining how novel characters originate and maintain their identities despite continuous turnover in the individual regulatory elements that give establish to gene regulatory networks. Here we use RNA-Seq to identify genes expressed in the endometrium during pregnancy in diverse placental mammals and ChIP-Seq to identify active cis-regulatory elements (enhancers, promoters, and transcription factor biding sites) in endometrial stromal cells, we then reconstruct how the gene regulatory network has diverged in different lineages of placental mammals, particularly primates. We show that: 1) transposable elements, which we have previously implicated in the origin of the endometrial gene regulatory network, have continuously remodeled the cis-regulatory landscape in the endometrium and played a major role in the turnover of regulatory elements and gene expression patterns within the mammalian endometrium; 2) Genes associated with younger transposable element derived regulatory elements are more strongly differentially regulated by the hormone progesterone than genes with more ancient transposable element derived regulatory elements, suggesting that the potentially disruptive effects of transposable elements on gene expression patterns are moderated over time; and 3) Genes associated with transposable element derived regulatory elements are located at the periphery of the regulatory network rather than its core. Thus our results suggest that while transposable elements continuously remodel the regulatory network in the endometrium leading to divergence in gene expression patterns, the core regulatory architecture of the network has remained relatively stable.

3.9

Does the presence of transposable elements near genes impact epigenetic modifications and gene expression in human?

Annabelle Haudry, Laura Gregoire, Emmanuelle Lerat
Lab. Biometrie et Biologie Evolutive - Univ Lyon 1, Villeurbanne, France

The effects of epigenetic modifications consist mainly in the modulation of gene expression such as differential expression between tissues. In this way, such modifications can potentially cause diseases under certain circumstances. In the human genome only 2% of the genome code for proteins, while the remaining 98% consists of non-coding regions and repeated elements. Among those, transposable elements (TEs) are present in millions of copies, which represent at least 45% of the human genome. Because of their abundance, TEs have a significant impact on genome evolution, gene expression or regulation. To counteract their deleterious effects, TEs are regulated by the host genome via epigenetic mechanisms. Previous studies have shown that TEs are not randomly distributed in the human genome and this appears to be linked to the function of genes. Moreover, the presence of TEs can change the expression of neighboring genes in cancer cells. The fact that the TE silencing can be removed could explain the change in gene expression in cancer. Indeed, a change in epigenetic modifications associated locally with TE sequences is expected to impact neighboring genes since these modifications occurring at TE sequences can spread to neighboring sequences. In this study, we ask whether the presence of TEs near genes may influence a change in the gene epigenetic modifications and in the gene expression when the cellular environment changes. For that, we have analyzed the associations between the epigenetic modifications, expression divergence and TE environment of genes between normal and tumor conditions in human.

3.10

Co-option of endogenous retroviral envelope proteins in the convergent evolution of the fish placenta

Nathaniel Jue, Rachel O'Neill, Michael O'Neill
University of Connecticut, Storrs, CT, USA

The mechanisms underlying how complexity evolves are of principle interest in evolutionary biology. In the example of the mammalian placenta, the co-option of retroelements in its development and function has been established as having major contributions to this organ's complexity, but the lack of intermediate phenotypes leaves their specific role in the evolutionary process undetermined. Placentas are not exclusive to mammals and post-fertilization maternal provisioning has evolved multiple times in the fish family Poeciliidae. Here, we present evidence from an RNA-SEQ study on 6 species of matrotrophic and lecithotrophic Poeciliids that supports the hypothesis that co-option of retroelements is essential to the evolution of placentation. In these species, endogenous retroviral envelope protein gene expression is correlated with the level of maternal investment across species and visualization of tissues expressing these genes demonstrate them to be associated with regions important to placenta/embryo interactions. Different lineages of retroviral envelope proteins are expressed in different species indicating this phenomenon to be phylogenetical independent. Thus, co-option of endogenous retroviral envelope proteins is posited as a general mechanism necessary for the evolution of the placenta and, thus, a principle contributor to the evolution of organismal complexity.

3.11

ALU exaptation enriches protein repertoire by introducing poly(A) signals

Eitan Lavi, Liran Carmel
The Hebrew University of Jerusalem, Jerusalem, Israel

It is well established that transposable elements introduce extended genomic variation, and that occasionally may be co-opted to play a functional role. It was already shown that intronic ALU elements in antisense orientation may, in some circumstances, present splicing signals that trigger the formation of new alternative exons. Here, we show that ALU elements in sense orientation may present polyadenylation signals that may, at some circumstances, compete with the original signals and lead to new gene ends. Once in the 3'UTR, such ALU elements may lead to an alternative 3'UTR, whereas once in an intron, such ALU elements may to a truncated version of the protein. We show that native ALU elements contain weak polyadenylation signals, which are not active under normal conditions. We hypothesized that specific mutations occasionally enhance these signals, leading to a novel cleavage and polyadenylation site. To test this, we scanned the human genome for ALU elements that overlap gene ends, and compare them to control intronic ALU elements. Using motif-finding algorithms, we have identified a number of mutations that may turn the native weak polyadenylation signal into an active one. We then showed that these mutations can be used as features in a classification algorithm designed to identify polyadenylation-triggering ALU elements. Using 10-fold cross validation, we have achieved over 90% accuracy in the classification.

3.12

The impact of social lifestyle on the repetitive elements landscapes in ten bee genomes - a case of social immunity?

Eckart Stolle [1,2], F Bernhard Kraus[2,3], Sophie Helbing[2], Robin FA Moritz[2,4]
[1] Queen Mary University, London, UK, [2] Martin-Luther-University, Halle-Wittenberg, Germany, [3] Department of Laboratory Medicine, University Hospital Halle, Halle, Germany, [4] German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany

Transposable elements can play major roles in driving evolution. An important aspect is horizontal transfer between even distant species. The transposable elements landscape in eukaryotic genomes can be most diverse, but particular genomic constraints or life history factors can shape it in similar ways. Trophic interactions might play a crucial role to determine pathways and likelihood for horizontal transfers of transposable elements and this way shape genomes of different species. Many recent examples showed that horizontal transfers are often facilitated by pathogens and their vectors. Particularly, viruses can shuttle transposable elements and can then again be transmitted across different host species by parasitic vectors such as mites and other blood feeding arthropods or endoparasites.
Our study focuses on the repetitive elements landscape in the genomes of ten bee species, exhibiting different levels of social organisation from solitary to advanced eusociality. It revealed interesting patterns of diversity and abundance of transposable elements in correlation with social lifestyle. In addition, we found traces of potential viral vectors. Most strikingly, genomes of advanced eusocial bee species appear to contain very low amounts of retro- and DNA elements indicating either low levels of infections or low persistence of transposable elements. We discuss our results in light of some aspects of the genome biology of social bees as well as point out the potential role of eusocial organisation as a superorganism. We show that the transposable elements landscape in highly eusocial bees could be consequence of social immunity in these societies.

3.13

Evolution of structural variation in a hybrid species of *Cottus*

Stefan Dennenmoser, Arne Nolte
Max-Planck Institute for Evolutionary Biology, Ploen, Germany

Genetic admixture is known to raise the rate at which gene copy number changes or transposition events occur. While such mutations may often be detrimental, they have been suspected to create novel variants that increase fitness. This could affect evolution in the wake of hybridization events, yet only few studies have investigated genomic structural variation in natural populations of hybrid origin. Invasive *Cottus* have recently evolved through hybridization between *Cottus rhenanus* and *Cottus perifretum*. We tested whether copy number variation (CNV) of protein coding genes and transposable elements evolved *de novo* in Invasive *Cottus*. Using comparative genomic hybridization arrays, we screened ~11.000 annotated genes for CNVs. Copy number variation of transposable elements was estimated by counting whole-genome sequencing reads that mapped against consensus transposon contigs. Our results suggest that natural hybrids show increased copy numbers at a low number of loci. We will discuss possible adaptive implications of CNV in hybrids in the light of potential candidate genes.

3.14

Massive bursts of transposable element activity in *Drosophila*

Robert Kofler, Viola Nolte, Christian Schlötterer
VetMedUni Wien, Wien, Austria

Transposable elements (TE) are stretches of DNA that propagate autonomously within genomes, but it is not clear whether TEs are moving at a constant rate or if they experience periods of dramatically increased activity (bursts). Determining the genome-wide TE content of *Drosophila melanogaster* and *Drosophila simulans*, we show that bursts of TE activity are a predominant process in *Drosophila*.Since TE insertions are frequently associated with a selective advantage, we suggest that such bursts of activity may have served a central role facilitating the adaptation of the two species to their novel environments after the recent out of Africa habitat expansion. To investigate this hypothesis we test whether insertions of TE families showing evidence for a recent burst of activity more frequently have signatures of positive selection (high population frequency and low Tajima's D) than insertions of other TE families.

3.15

dnaPipeTE, a new bioinformatic pipeline to assemble, annotate estimate abundance and dynamics of repetitive DNA in low coverage sequencing: application to *Aedes* mosquitoes

Clément Goubert [1,3], Laurent Modolo[1,3], Cistina Vieira[1,3], Claire Valiente-Moro[2,3], Patrick Mavingui[2,3], Matthieu Boulesteix[1,3]

[1] CNRS UMR 5558 Laboratoire de Biométrie et de Biologie Evolutive, Villeurbanne, France, [2] CNRS UMR 5557 Ecologie Microbienne, Villeurbanne, France, [3] Université de Lyon 1, Lyon, France

Repetitive DNA, including transposable elements (TEs), is found throughout eukaryotic genomes. Annotating and assembling this "repeatome" during genome-wide analysis often poses a challenge. To address this problem, we present dnaPipeTE - a new bioinformatics pipeline that uses a small amount of raw genomic reads. It produces precise estimates of repeated DNA content and TE consensus sequences, as well as an overview of the relative ages of TE families. dnaPipeTE can annotate and quantify repeats in any genome, using very low coverage sequencing as input (< 0.5X) and works without reference genome assembly. We applied this pipeline to the genome of the Asian tiger mosquito, *Aedes albopictus*, an invasive species of human health interest, for which the genome size is estimated to be over 1 Gbp but whose sequence has not been released yet in spite of multiple ongoing projects. Using dnaPipeTE, we showed that this species harbours a large (50% of the genome) and potentially active repeatome with an overall TE class composition similar to that of *Aedes aegypti*, the yellow fever mosquito. However, intra-class dynamics shows clear distinctions between the two species, with differences at the TE family level that are compatible with the theory of genome ecology. Our pipeline's ability to manage the repeatome annotation problem will make it helpful for new or ongoing assembly projects, and our results will benefit future genomic studies of *Ae. albopictus*.

Goubert et al. 2015. *in minor revision* for Genome Biology and Evolution

3.16

The epigenetic interplay between transposable elements and their plant hosts

Alexandros Bousios [1,2], Nikos Darzentas[4], Concepcion M. Diez[3,5], Shohei Takuno[3], Brandon Gaut[3]

[1] University of Sussex, Brighton, UK, [2] Institute of Applied Biosciences, Thessaloniki, Greece, [3] University of California, Irvine, Irvine, USA, [4] CEITEC, Brno, Czech Republic, [5] University of Cordoba, Cordoba, Spain

Transposable elements (TEs) occupy the largest proportion of plant genomes, but are typically silenced by host epigenetic mechanisms. The main component of these mechanisms are small interfering RNAs (siRNAs) that target TE sequences and guide their methylation. This process is known as RNA-directed DNA methylation (RdDM) and has been extensively studied. However, one aspect that remains poorly understood is the patterns of siRNA targeting and methylation along the length of TEs, and how these vary with TE age. Therefore, we investigated these epigenetic features in 6,456 carefully annotated Sirevirus LTR retrotransposons in maize. Sireviruses are a major component (~20%) of the maize genome, and, as expected, were found highly methylated. Yet, we identified several novel features of the siRNA response. First, although presumably silenced, maize allocates a significant proportion of 21-22 nucleotide siRNAs to Sireviruses, suggesting that few elements may routinely escape silencing *in vivo*, triggering 'epigenetic emergencies'. Second, targeting varies along the sequence. More specifically, a complex region of multiple palindromic motifs within the LTRs acts as a hotspot for both siRNA targeting and sequence evolution. Third, due to its capacity to form stem-loop structures, this region likely produces a class of miRNA-like, hairpin-derived small RNAs. This implies a new mechanism for TE silencing besides the canonical siRNA-based RdDM. Finally, siRNA targeting is a function of age, with older elements targeted at unexpectedly high levels, partially due to specific conservation of the palindrome-rich LTR region. Altogether, these findings provide new insights into the epigenetic interplay between plants and TEs.

476A

Hybrid dysgenesis in *Drosophila simulans* caused by a rapid global invasion of the P-element

Tom Hill, Christian Schlötterer, Andrea Betancourt
Institute of Population Genetics, Vetmeduni, Vienna, Austria

In a classic example of the invasion of a species by a selfish genetic element, the P-element was horizontally transferred from a distantly related species into *Drosophila melanogaster*, where it spread globally in the course of a few decades. *D. melanogaster*'s sister species, including *D. simulans*, remained P-element free, until recently. Here, we survey *D. simulans* strains collected worldwide over the past 25 years for the presence of the P-element. We find that P-element has rapidly spread in *D. simulans*, possibly even more rapidly than in *D. melanogaster.* We also find that, as in *D. melanogaster*, hybrid dysgenesis is associated with the spread of the P-element; crosses between strains collected before and after the invasion suffer from a high proportion of malformed ovaries.

477B

Comparison of recent activities of genomic repeats for multiple species using sequence reads

Chong Chu[1], Tobias Mourier[2], Rasmus Nielsen[3], Yufeng Wu [1]
[1] University of Connecticut, Storrs,CT, USA, [2] University of Copenhagen, Copenhagen, Denmark, [3] University of California, Berkeley, Berkeley,CA, USA

Highly repetitive genomic elements, such as transposable elements (TEs), are of great evolutionary importance phenotypically and genomically. Sequencing data from many species are now available, providing opportunities for finding and comparing genomic repeat activity among species. However, the currently available reference sequence data are typically not suitable for comparative studies of recently expanded repeat families. The construction of reference genomes from repeats regions is computationally challenging and such regions are often omitted from published reference assemblies. The degree to which they have been assembled (correctly) may vary from species to species, greatly hampering efforts to compare recently expanded repeat families among species.

In this presentation, we will present a new computational method for directly analyzing, recently expanded, highly repetitive genomic regions from sequence reads without relying on an existing reference assembly. The method performs de novo assembly of repeat motifs from raw sequence reads, and provides a metric for directly comparing repeat content among multiple species. We show that the method can reconstruct references of repeat motifs from recently expanded families, and analyze raw sequencing data from more than twenty organisms, including Adelie Penguin, Anna Hummingbird, Human, Medium Ground Finch, etc. In each genome, we identify recently expanded repeat families, and show that there are marked differences between species in composition of repeat elements that have recently expanded in the germline.

478C

The epigenetic response to transposable element activity is quickly evolving in natural populations

Bastien Saint Leandre, Isabelle Clavereau, Emilie Robillard, Aurélie Hua-Van, Pierre Capy
EGCE-CNRS-Université Paris SACLAY, Gif-sur-Yvette, France

Transposable elements (TE) are genome parasites present in all eukaryotic genomes, usually silenced in the germile by the conserved PIWI RNA pathway.

Based on the study of one model TE, mariner, in Drosophila, we established the relationship that could exist between this regulatory machinery and TE dynamics into colonizing populations and laboratory lines.

D. simulans, a sister species of D. melanogaster naturally contains mariner elements, whose activity is highly variable in natural populations, especially in invasive ones. Interestingly, we found that these variations are negatively correlated to expression of major piRNA genes and protein expression.

When artificially introduced into D. melanogaster, the element is able to amplify till about 20 copies. Expression analysis and deep sequencing revealed that mariner is presently repressed in the germline by the piRNA pathway. Indeed, de novo mariner insertions into repeat rich regions are able to produce the required piRNA for its own silencing.

Our results suggest that the genome response to TE invasion is a rapidly evolving trait, determinant in shaping the "transposon load".

479D

The Role of *FBti0019985* Transposable Element in *Drosophila melanogaster* Adaptation: molecular mechanism and functional consequences

Miriam Merenciano, Anna Ullastres, Josefa González
Intitute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain

A way to study environmental adaptation is by elucidating the relationship between mutations and their phenotypic effect. Transposable elements are mobile DNA sequences that can induce genetic variations and have been previously shown to be a considerable source of adaptive mutations in *Drosophila melanogaster*. *FBti0019985* is a putatively adaptive transposable element insertion that is present at higher frequencies in out-of-Africa populations compare to Africa populations, where *D. melanogaster* was originated. *FBti0019985* overlaps 54 nucleotides with the 5'UTR region of *CG18446* gene, a gene of unknown function. We found that flies with *FBti0019985* insertion are more resistant to cold stress than flies without the insertion. In addition, the presence of *FBti0019985* increases the expression levels of *CG18446* in normal conditions. Recently, we have found that *FBti0019985* has recurrent insertion sites upstream of *CG18446*. These other insertions are present at lower frequency in natural populations compared to the one previously studied. We found that flies containing these insertions show different patterns of cold resistance compared to flies without the insertion. We are currently performing CRISPR/Cas9 to generate *FBti0019985* knockout fly strains to further demonstrate the role of this transposable element in cold stress resistance. Overall, our results indicate that the effect of *FBti0019985* depends on the insertion site of this element and provide another example of a transposable element insertion involved in environmental adaptation.

480A

A cost-effective whole-genome screen for identifying novel transposable element (TE) insertions

Erika Kvikstad, Gerton Lunter
University of Oxford, Oxford, UK

As mobile genetic elements that randomly propagate within their host's genome, TEs are major drivers of genome evolution. By donating transcription and splice signals, even intronic insertions can potentially affect transcription and cause disease. However, existing screens for mutations underlying genetic disorders typically ignore TEs for technical reasons despite their extensive mutational target.

We hypothesize that the role of TEs in disease has thus far been underestimated. Hence, we developed a genome-wide screen for targeted TE discovery from next-generation sequencing data. Briefly, the strategy consists of standard library preparation, followed by selective amplification of genomic fragments containing TEs and their unique flanking sequence. Unlike existing approaches, we simultaneously target the three most active TE subfamilies (AluYb8/b9, AluYa5/a8, and L1HS) that together comprise the majority of polymorphic insertions.

Our approach yields substantial enrichment (> 4 orders of magnitude) for targeted fragments spanning on average 62 nucleotides of unique sequence flanking the poly-A signature, aiding accurate mapping and detection of polymorphic and novel insertion sites. Indeed, in a single individual we detect 99% of fixed L1HS targets and ~2/3 known polymorphic L1HS (e.g., dbRIP).

High rates of sensitivity and specificity, coupled with a substantial proportion of usable data (~75% total reads) allows for deep multiplexing of hundreds of individuals in a single sequencing experiment. Application of this technique alongside exome sequencing of patients with sporadic disease has the potential to reduce the number of unresolved cases in such screens, and improve estimates of the contribution of TEs to human genetic disease.

481B

Invasion of lizard *Hox* clusters by transposable elements during an adaptive radiation

Nathalie Feiner
University of Oxford, Oxford, UK

Hox genes are pivotal for development, including the patterning of the vertebrate limb. *Hox* gene regulation is tightly coordinated in compact gene clusters. In vertebrates, purifying selection keeps these clusters tightly packed and devoid of transposable elements (TEs). This constraint appears to be relaxed in squamates where *Hox* gene clusters harbor numerous TEs, with the green anole lizard showing exceptionally high density of TEs. The genus *Anolis* is a famous example of an adaptive radiation: upon colonization of Caribbean islands, these lizards diversified into several ecomorphs with different limb morphology. Here we map the invasion of individual TEs onto the well-resolved phylogeny of these lizards and test for association between TE presence and limb morphology.

We sequenced 'posterior' regions of the HoxA and –D cluster of more than 20 *Anolis* species plus closely related lizards and annotated their TE content. We found that (1) all *Anolis* lizards possess significantly more TEs than any other vertebrate in their *Hox* gene clusters and (2) there are marked differences between groups of species within the genus *Anolis*. This shows that the phenotypic evolution of these lizards was paralleled by high levels of TE activity that altered their *Hox* cluster structure. We discuss to what extent TE activity may have contributed to diversification of *Anolis* lizards through changes in *Hox* gene regulation.

482C

Transposable element population dynamics in *Drosophila melanogaster* using next generation sequencing data.

Maite G Barrón [1], Isabelle Luyten[2], Hadi Quesneville[2], Dmitri A Petrov[3], Anna Sophie Fiston-Lavier[4], Josefa Gonzalez[1]

[1] Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain, [2] INRA 'Unité de Recherche en Génomique et Informatique', Versailles, France, [3] Stanford University, Stanford, California, USA, [4] Institut des Sciences de l'Evolution (ISEM), Montpellier, France

Transposable elements are abundant, diverse and active genomic components that generate a great variety of mutations. Hence, understanding TEs population dynamics is crucial if we want to understand the complex organization, function and evolution of genomes. There are two different models to explain TE dynamics; the transposition selection balance model and the burst of transposition model. Several attempts have been made with partial datasets to explain TE dynamics under one or the other model. Here, we measured both the population frequency and age of 1636 well-known TE insertions in four worldwide *Drosophila melanogaster* populations. We also estimate a range of variables specific to each TE insertion such as length, recombination rate, distance to closest gene or family copy number. These comprehensive dataset allow us to test whether TE dynamics can mainly be explained neutrally by the TE age or whether models of negative selection under transposition selection balance need to be invoke to explains the bulk of the TE dynamics. Overall, we found that knowing the age of the TE insertion improves our understanding of TE dynamics. However, it is not the main variable and explains 18% of the variation. Our results corroborate and extend previous observations indicating that ectopic recombination is an important force eliminating TEs and the modest effect that distance to the nearest gene exerts on TE dynamics. Combining all these variables we explain 47% to 70% of the variance in the TE frequencies observed in natural populations.

483D

A likelihood method for detecting cryptic copy number variation from low-coverage, Next-Generation Sequencing data applicable to model and non-model systems

Tyler Linderoth [1], Rasmus Nielsen[1]
[1] University of California Berkeley, Berkeley, California, USA, [2] University of Copenhagen, Copenhagen, Denmark

Determining how various types of genomic duplications are spatially distributed throughout the genome, how they influence genome-size, as well as host fitness consequences first boils down to being able to identify areas of the genome in which duplication has occurred in many cases. To address this issue we have developed a likelihood method for identifying paralogy using Next-Generation Sequencing data from multiple individuals. We used simulations to demonstrate that this method is powerful at detecting paralogy for even small sample sizes with low sequencing coverage (e.g. 10 individuals at 2X average coverage is usually more than sufficient). This method has been implemented in the software ngsParalog which requires nothing other than BAM files and operates on a per-site basis such that it can be used to scan for paralogy in systems with entire, assembled genomes or in cases for which there is only a collection of contigs and no reference genome. We applied ngsParalog to the 1000 Genomes low-coverage human data and that of non-model organisms with only de novo assemblies to detect evidence of duplication. The ability to detect paralogy in this manner has broad application in any situation where genomic duplication is involved (e.g. transposable elements, whole genome-duplication, etc.) ranging from the biological significance of duplication to the important inference problem of inflated heterozygosity when sequencing reads from paralogous loci map to the same location.

484A

Transposable elements influence stress response regulatory networksin the Drosophila melanogaster and the human genomes.

Quirze Rovira, Maite G. Barrón, Josefa González
Institute of Evolutionary Biology (Universitat Pompeu Fabra - CSIC), Barcelona, Spain

Stress response processes are highly conserved across organisms and provide a great opportunity to study how adaptation occurs. Transposable elements (TEs) are active components of genomes that have been shown to rewire and fine-tune regulatory networks. However, the specific role of TEs in stress response regulatory networks has not been studied in detail. Our goal is to identify a catalogue of putatively adaptive TEs that add stress response elements (SREs) in the upstream regions of fruitfly and human's genes. For this study, we selected SREs that are highly conserved between Drosophila and humans. We combine bioinformatics, population genetics, and experimental approaches to identify and validate the TEs involved in stress response. We have identified 159 Drosophila melanogaster TEs that contain at least one of the studied SREs. Several of these TEs add cisregulatory modules containing two or more different SREs. Interestingly, some of the identified TEs are located upstream of genes that were not previously annotated as being involved in stress-response. Preliminary analyses of the human genome showed that SREs are also present in human-specific TEs. These results strongly suggest that TEs play an important role in the modulation of stress response regulatory networks and provide a promising candidate list that will be functionally analyze.

485B

Copy number variation in ribosomal DNA and the transposon, *Pokey* in *Daphnia* .

Kaitlynn LeRiche, Shannon Eagle, Teresa Crease
University of Guelph, Guelph, Ontario, Canada

Pokey is a class II DNA transposon that inserts into 28S ribosomal RNA (rRNA) genes and other genomic regions of species in the subgenus, *Daphnia*. Two divergent lineages, *Pokey*A and *Pokey*B have been identified. Recombination between misaligned rRNA genes changes their number and the number of *Pokey* elements. We used qPCR to estimate rRNA gene and *Pokey* number in isolates from natural populations of *Daphnia obtusa*, and in clonally propagated mutation accumulation lines (MAL) initiated from a single *D. obtusa* female. The change in direction and magnitude of *Pokey* and rRNA gene number did not show a consistent pattern across 87 generations in the MAL; however, *Pokey* and rRNA gene number changed in concert. *Pokey*A and 28S gene number were positively correlated in the isolates from both natural populations and the MAL. *Pokey*B number was much lower than *Pokey*A in both MAL and natural population isolates, and showed no correlation with 28S gene number. Preliminary analysis did not detect *Pokey*B outside rDNA in any isolates and detected no more than 4 copies of *Pokey*A outside rDNA indicating that *Pokey* is primarily an rDNA element in *D. obtusa*. The recombination rate in this species is high and the average size of the rDNA locus is about twice as large as that in other *Daphnia* species such as *D. pulicaria* and *D. pulex*, which may have facilitated expansion of *Pokey*A to much higher numbers in *D. obtusa* rDNA than these other species.

486C

An Ancient Trans-Kingdom Horizontal Transfer of Penelope-like Retroelements from Insects to Conifers

Xuan Lin[1], Nurul Faridi[1,2], Claudio Casola [1]
[1] Texas A&M University, College Station, TX, USA, [2] USDA Forest Service Southern Research Station, Saucier, MS, USA

Penelope-like elements, or PLEs, represent a class of retroelements represented by two main types: EN(+)PLEs, which encode the unique combination of a reverse transcriptase (RT) domain and a GIY-YIG endonuclease (EN) domain; and EN(-)PLEs, encoding only a RT domain. While members of the second type occur in a variety of eukaryotes, EN(+)PLEs have thus far been detected only in animal genomes, with the noteworthy exception of the recently sequenced loblolly pine (Pinus taeda L.) genome. However, the origin of EN(+)PLEs in pine trees and other gymnosperms remains unknown. In this work, we have investigated the evolutionary history of loblolly pine and other conifers EN(+)PLEs, which we have named Dryads. In phylogenies of PLE sequences, Dryads form a monophyletic group placed within a major animal EN(+)PLE lineage known as Poseidon. Furthermore, Dryads appear to be closely related to a clade of EN(+)PLEs primarily found in insects. Bioinformatics surveys revealed no EN(+)PLEs in 625 fully sequenced non-metazoan and non-conifer genomes from twelve major eukaryotic lineages. Additionally, PCR assays indicate that while Dryads occur in non-Pinaceae conifers, they are absent in non-conifer gymnosperms, including Ginkgo biloba and several cycads and gnetales. Taken together, these findings indicate that Dryads emerged following an ancient horizontal transfer of Penelope-like elements from an insect group to a common ancestor of conifers in the late Paleozoic. Our findings suggest that retroelements invasions from non-conifer groups might have played an important role in the expansion and evolution of the very large conifers genomes.

487D

In situ hybridization with P element to *Drosophila simulans* polytene chromosomes

Volha Paulouskaya [1], Tom Hill[1], Wolfgang Miller[2], Andrea Betancourt[1]
[1] *Institut für Populationsgenetik, Veterinärmedizinische Universität Wien, Vienna, Austria,* [2] *Medical University of Vienna, Vienna, Austria*

Transposable elements are selfish genetic elements that insert themselves into new locations in the genome, increasing their copy number. Transposable elements not only relocate within genomes, but are also capable of invading new species. P element, a DNA transposon causing hybrid dysgenesis, invaded natural populations of *Drosophila melanogaster* several decades ago after horizontal transfer from a distant relative, *D. willistoni* (Engels 1992). Recent research has shown that P element, though previously absent *D. simulans*, has now invaded and spread rapidly worldwide in *D. simulans* populations. Analysis of P element insertions using in situ hybridization showed accumulation of the element at the tip of the X chromosome (Ajioka and Eanes 1989), *D. melanogaster*. This region, known as X-TAS, represses P element activity by producing piRNA. Here, we use in situ hybridization of *D. simulans* polytene chromosomes in order to investigate whether P element insertions also accumulate at particular sites in the genome in this species. Data from different populations will be shown, including from recently collected strains from populations with full P element insertions and from old collections from populations where P element is absent.

488A

Retrocopies - treasures among the junk DNA

Joanna Ciomborowska, Magdalena Kubiak, Michal Kabza, Izabela Makalowska
Department of Bioinformatics, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznan, Poland

Retrocopies are RNA-based duplicates originated from reverse transcription of mRNA and incorporation of cDNA into a genomic sequence, so called retroposition. Majority of retrocopies are usually inactive and therefore are commonly called retropseudogenes or just pseudogenes. Since the discovery of the first functional retrogene in 1985 the interest in retrogenes (expressed retrocopies) we have more and more examples showing retrocopies as a driving force in evolution of animals and playing an important role in shaping interspecies differences. Also our studies reveled that we should reconsider retropseudogenes and look at them from a new perspective as we still underestimate the real number of animal retrogenes.

We decided to check whether previously annotated as pseudogenes retrocopies can be transcribed and be in fact functional. Candidates for our analyses come from developed by us database of animal retrogenes - RetrogeneDB (retrogenedb.amu.edu.pl). We were able to confirm expression of 41 human and 11 mouse retrogenes utilizing PCR and pooled human and mouse cDNA libraries as templates. For those retrogenes we performed a set of bioinformatics analyses in order to pinpoint their putative functions. Among them we identified 29 human and 9 mouse genes with disrupted open reading frames, which together with confirmed expression, strongly suggests neofunctionalization. Moreover significant part of examined retrogenes is specific for one organism. 32% of expressed human and 90% of expressed mouse retrogenes has no orthologs in other organisms. This observation shows how big could be the impact of retroposition on variation among species.

489B

A survey of variability in LTR-retrotransposon abundance and proximity to genes between wild and cultivated sunflower genotypes.

Flavia Mascagni, Elena Barghini, Tommaso Giordani, Andrea Cavallini, Lucia Natali
Department of Agriculture, Food and Environment. University of Pisa, Pisa, Italy

Sunflower (*Helianthus annuus*) is an important crop species of the Asteraceae family. Recent characterization of sunflower repetitive fraction has shown that the genome of this species contains a very large proportion of transposable elements, especially long-terminal-repeat retrotransposons. However, knowledge on the retrotransposon-related variability within this species is still limited. We used next generation sequencing technologies to perform a quantitative and qualitative survey of intraspecific variation of the retrotransposon fraction of the genome across different genotypes of *H. annuus*. First, we characterized the repetitive component of a sunflower homozygous experimental line, using 454 reads, and prepared a library of retrotransposon-related sequences. Then, we analysed the retrotransposon fraction of 7 wild accessions and 8 cultivars of *H. annuus* by mapping Illumina reads of the 15 genotypes onto the library. We observed large variations in redundancy among genotypes, at both superfamily and family levels. In another analysis, we mapped Illumina paired reads of the 15 genotypes onto two sets of sequences, i.e. retrotransposons and protein-encoding sequences, and evaluated the extent of retrotransposon proximity to genes in the 15 genomes by counting the number of paired reads of which one mapped onto a retrotransposon and the other onto a gene. Large variability among genotypes was ascertained also for retrotransposon proximity to genes. Both retrotransposon redundancy and proximity to genes showed different behaviour among retrotransposon families and also between cultivated and wild genotypes, indicating a possible involvement in sunflower domestication.

490C

High richness in the dying mobilome of Coffee Berry Borer's genome

Erick Hernandez[1], Lucio Escalante[2], Pablo Machado[2], Claudia Carareto [1]
[1] Sao Paulo State University, Sao Jose do Rio Preto/SP, Brazil, [2] Cenicafe, Manizales, Colombia

Systematic searching strategies using de novo and homology-based methods were implemented in order to build a transposable elements (TE) reference library from a draft genome of the coffee berry borer (Hypothenemus hampei), the most important pest in coffee production worldwide. The library consists of 880 sequences (LTRs: 46%, Non-LTRs: 20%, DNA transposons: 8%, Helitrons: 16% and MITEs: 10%) including families of the three main LTR superfamilies (Gypsy, Bel/Pao and Copia), Non-LTR clades (CR1, Daphne, I, Jockey, Kiri, Nimb, R1, R2 and R4) and DNA transposons (Tc1-Mariner, hAT, Merlin, P, PIF-Harbinger, PiggyBac and Helitrons). RepeatMasker identification suggests that 8.2% of the CBB assembled genome consist of TEs but most of the sequences are degenerate. This content agrees with that of two coleopteran species, Tribolium castaneum (5.9%) and Dendroctonus ponderosae (7%). MITEs reach an unprecedented fraction of 49% of the total content. Broadly, annotation showed that the copies are highly fragmented; being only few copies of LTRs and DNA transposons possibly active. On the contrary, several representatives of different NLTR clades were found in more than one complete and putatively active copy. This is a first step for a better knowledge of the genome structure of CBB into the ongoing genome project for this species, but further studies are necessary in order to elucidate functional relationships of TEs with the genome evolution and if there is insertional variation of TEs at inter-population level and thus, take advantage of these sequences as new genetic markers in innovative pest control strategies.
Financial support: FAPESP

491D

Elucidating the Gene Regulatory Landscape of Human Pregnancy

Mihaela Pavlicev[2], Corinne Simonti[1], Kayleigh Swaggart[2], John Capra [1]
[1] Vanderbilt University, Nashville, TN, USA, [2] Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

We report preliminary results of localizing distal gene regulatory regions in tissues relevant to human pregnancy and characterizing their evolutionary history. The physiology of pregnancy and parturition manifests significant variation within and across species. Genetic variation in regulatory regions, including distal enhancers, likely contributes to much of this phenotypic variation.

Building on the recent identification of gene regulatory enhancer regions in several human reproductive tissues including the placenta, uterus, umbilical cord, and vagina, we used the EnhancerFinder pipeline to create predictive machine-learning models of known enhancers in these tissues. Applying these models to the human genome, we produced genome-wide maps of enhancers likely relevant to gestation. We will experimentally validate several of these and predict potential target genes based on gene expression, DNaseI hypersensitivity, and transcription factor binding patterns.

By characterizing the evolutionary history of predicted enhancers in relevant tissues and relating their evolution to differences in pregancy between species, we will suggest potential gestational functions regulated by these regions. In particular, we will describe the recruitment of transposable elements (TEs) as lineage and tissue specific enhancers.

Our genome-wide reference maps of regulatory enhancers and their evolutionary origins will enable powerful tests for the association of non-coding genetic variants with pregnancy phenotypes. These integrated maps will also facilitate the generation of molecular hypotheses about the basis of such associations.

492A

Evolutionary analysis of deletion polymorphism of the entire GSTM (glutathione S-transferase mu) 1 gene generated by a homologous chromosomal recombination suggests multiple deletion events

Marie Saitou, Takafumi Ishida
Graduate School of Science, University of Tokyo, Tokyo, Japan

The glutathione S-transferases (GSTs) are important phase II enzymes. Deletion types of genetic variants of the GSTM (mu) 1 have been ubiquitously found in human populations. The GSTM1 gene is flanked by two highly similar regions and the GSTM1 null allele is thought to be generated by a homologous recombination. In this study, first, we have genotyped the GSTM1 of 33 Dani individuals by using both short-range PCR and long-range PCR. Next, we have sequenced the entire 13kb recombinant region of a heterozygous individual by primer-walking method. Comparing the recombinant sequence with the two highly similar sequences of the GRCh38 Primary Assembly, we found STRs and SNPs in these sequences and determined the putative breakpoint. The breakpoint was different from previous report of a European individual suggesting that multiple independent deletion events occurred.

493B

Comparison of transposable elements dynamics among taxa with different speciation rates

Valentina Peona, Marco Ricci
University of Bologna, Bologna, Italy

According to many authors bursts of activity of transposable elements may be related with speciation events in animals and plants by facilitating reproductive isolation and increasing speciation rate (Senerchia 2015, Jurka 2011, Rebollo and Vieirà 2010, Oliver and Greene 2010, Zhe et al 2009).

In the present work we tested this hypothesis in Mammal and Bird groups with different speciation rates (25 species).

In particular, we performed large scale analyses aimed to evaluate the total number of recently inserted Transposable elements (TE). In order to estimate insertion density, observed values were normalized to genome size (insertion density). Within Mammalia, we compared pairs of species belonging to different families of the same order.

Within the class Mammalia we compared pairs of species belonging to different families of the same order.

In most cases (16 out of 19), we observed a higher density of insertion in taxa with higher speciation rate. Accordingly, we confirm that insertion density in Mammals is positively related to the speciation rate (Wilcoxon test, $p<.01$).

Within Aves we compared three species from Galliformes with tree species from Passeriformes, respectively the orders with the lowest and highest speciation rate.

Results showed that the TE density in these two taxa does not show significant differences, suggesting that TE activity did not affect speciation in Aves as much as in Mammalia. Hence, the tested hypothesis is not confirmed in this case. Alternative explanations and implications are discussed.

494C

Long-term preservation of the Au SINE retrotransposon family in plant genomes

Jeffrey Fawcett, Hideki Innan
SOKENDAI, Hayama, Kanagawa, Japan

Several TE (Transposable Element) families show surprisingly high levels of similarity between distantly related species. This high similarity and 'patchy' distribution has often been attributed to frequent horizontal transfers of TEs between species, even though the mechanistic basis is usually highly speculative. Here, we studied the evolution of the Au SINE family, in which high similarity between many distant plant species has been reported. Indeed, we found that the Au SINE family is present in >30 different families, exhibiting high nucleotide similarity (>80%) between various distant species such as tomato and Brachypodium. Despite the high similarity, we find no evidence suggesting recent horizontal transfers. Instead, first, we find that the Au SINE family was at least present in the common ancestor of multiple Fabaceae species, multiple Solanacae species, and multiple Rosaceae species. Next, the Au SINE family is inactive and almost undetectable in many species, suggesting that the 'loss' of this SINE family has been common in the evolution of plants. Finally, the 3'-end of this SINE family shares high similarity with the 3'-end of a LINE family, suggesting that the activity of this SINE family is likely constrained at the sequence level. Our results suggest that the high similarity and patchy distribution of the Au SINE family in plants can be explained without having to assume hypothetical horizontal transfers or exaptations.

4 Evolution and Ecology of microbial communities

4.1

Towards understanding microbial communities: From gut to ocean

Peer Bork
EMBL, Heidelberg, Germany

The human microbiome, that is all the microbes living in and around us, has recently become accessible by environmental shotgun sequencing (metagenomics), whereby the most prominent human habitat of our invisible microbial companions is the gut, harbouring in the order of 1000 species. We recently identified three microbial community types in the population, which we dubbed enterotypes (Arumugam et al, Nature, 2011). I will use this stratification of the human population to discuss methodological challenges in defining ecological states, but also describe resilience of this ecosystem in longitudinal studies as well as changes in disease state or after fecal microbiota transplantation, both at species, but also at strain level, the latter we found to serve as a fingerprint of an individual (Schloissnig et al., Nature, 2013). Although we see first geographic differences within gut species in different countries from different continent, a true planetary scale of microbiome analysis requires more systematic sampling, which was done using the TARA Oceans project. I will give a brief introduction of the factors that drive microbiome diversity in the world's oceans using the first results of this expedition that captured 35000 ocean samples within a 4 year journey (Sunagawa et al., Science, in press, Lima-Mendez et al., Science in press).

4.2

Host genetic variation at *B4galnt2* influences intestinal microbial ecology and susceptibility to enteric pathogens in house mice

John Baines [1,2], Philipp Rausch[1,2], Marie Vallier[1,2], Abdulhadi Suwandi[2], Jill Johnsen[3], Guntram Grassl[2]

[1] Max-Planck-Institute for Evolutionary Biology, Ploen, Schleswig-Holstein, Germany, [2] Christian-Albrechts-University of Kiel, Kiel, Schleswig-Holstein, Germany, [3] University of Washington, Seattle, WA, USA

Glycans on mucosal surfaces play an important role in host-microbe interactions in the mammalian intestine. *B4galnt2* is a blood group-related glycosyltransferase whose two murine alleles (driving gastrointestinal- and vascular expression) are maintained by balancing selection and determine the presence/absence of *B4galnt2*-derived glycans in the intestinal tract. Metagenomic analysis of *B4galnt2*-knockout mice reveals that the loss of *B4galnt2* expression in the intestine leads to widespread alterations in intestinal bacterial community structure. This includes shifts in dominance between numerous closely-related bacterial taxon pairs, suggesting competition for a *B4galnt2*-defined niche(s). Given the signatures of selection present at *B4galnt2* and its influence on the composition of intestinal microbiota, we hypothesize that variation in *B4galnt2* expression may alter susceptibility to intestinal pathogens. To test this, we challenged mice genetically engineered to express different tissue-specific expression patterns with a *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) infection model. We found that the loss of *B4galnt2* intestinal expression is associated with increased baseline microbial phylogenetic diversity and decreased susceptibility to *S.* Typhimurium, as evidenced by decreased inflammatory cytokines and infiltrating immune cells. Further, fecal transfer experiments into previously germ-free mice confirm a role of the *B4galnt2*-dependent microbiota in conferring decreased susceptibility to *S.* Typhimurium. These results suggest that host glycan-dependent alterations in microbial communities may be an extended phenotype upon which selection can act. Finally, metagenomic and histological analysis of wild mice reveals that *B4galnt2* genotype correlates with differences in intestinal inflammation and the presence of candidate pathogens.

4.3

Invasions of the Pharynx: Microbiome of Infected Respiratory Tissue

Asha Ram, Umer Ijaz, Christopher Quince, Thomas Evans, Jan Lindström
University of Glasgow, Glasgow, UK

Increasing knowledge of microbiota is spearheading a paradigm shift in the study of infectious diseases, the focus is moving from pathogens to the ecology and evolution of microbial communities determining health and disease. As the human pharynx is an important gateway for pathogen entry, we studied the species composition and stability of the microbiome in healthy subjects over a period of time, and whether changes could be linked to upper respiratory tract infections. 18 participants provided pharynx swabs weekly for nine months; the microbiome was characterized using sequences from the V1-V2 region of the 16S rRNA gene. In the absence of infection, the bacterial pharynx microbiota were typically dominated by phylum Firmicutes and Streptococcus species, however, there were also individual differences, partly explained by sex and age. Bacterial respiratory infections were associated with low diversity communities which showed a shift from Firmicutes to Proteobacteria, whereas some viral infections (Rhinovirus and Respiratory Syncytial Virus) were not associated with significant changes in microbiota. The microbiota recovered quickly after a disturbance, although antibiotic use made this slower and more variable. This study is the first to characterize the human pharynx microbiota longitudinally; the next step is to ask whether heightened risk of infections in people who are not baseline healthy is associated with different microbiota and, in the long run, if this will impact how to treat respiratory disease through restoration of pharyngeal communities.

4.4

Rapid changes in the gut microbiome during human evolution

Andrew Moeller [1,4], Beatrice Hahn[2], Anne Pusey[3], Howard Ochman[4]
[1] Yale University, New Haven, CT, USA, [2] University of Pennsylvania, Philadelphia, PA, USA, [3] Duke University, Durham, NC, USA, [4] University of Texas, Austin, Austin, TX, USA

Humans are ecosystems containing trillions of microorganisms, but the evolutionary history of this microbiome is obscured by a lack of knowledge about microbiomes of African apes. We sequenced the gut communities of hundreds of chimpanzees, bonobos, and gorillas and developed a phylogenetic approach to reconstruct how present-day human microbiomes have diverged from those of ancestral populations. Compositional change in the microbiome was slow and clock-like during African ape diversification, but human microbiomes have deviated from the ancestral state at an accelerated rate. Relative to the microbiomes of wild apes, human microbiomes have lost ancestral microbial diversity while becoming specialized for animal-based diets. Individual wild apes cultivate more phyla, classes, orders, families, genera, and species of bacteria than do individual humans across a range of societies. These results indicate that humanity has experienced a depletion of the gut flora since diverging from *Pan*.

4.5

The landscape of epistasis in natural bacterial populations

Daniel Falush
University of Swansea, Swansea, UK

All bacteria reproduce clonally, meaning that strains carrying particularly fit combinations of interacting alleles to spread through the population. On the other hand, in many bacteria, there is sufficient recombination that in the absence of selection, pairs of alleles that are not in close physical linkage should be in approximate linkage equilibrium.

In this talk I will explore the use observations of the pattern of linkage disequilibrium between loci in genomes of natural isolates to characterize the epistatic fitness of landscape in bacteria. In order to detect true epistatic associations the algorithms need to take into account population and clonal structure of the species. I will briefly describe approaches for this and then discuss the interactions that we have detected in three quite different recombinogenic bacterial species, namely Vibrio parahaemolyticus, Helicobacter pylori and Campylobacter.

4.6

Genomic diversification in the bacterial gut symbionts of social bees

Nancy Moran
University of Texas at Austin, Austin, Texas, USA

The evolution of gut microbiota is influenced by the social contacts of hosts. In mammals, social bees and termites, frequent interactions enable direct transmission of gut microorganisms, thus enabling specialization on the gut environment and consequent restriction to this habitat. In such hosts, gut-dwelling microorganisms undergo extensive strain-level diversification. The communities in guts of honey bees and bumble bees provide a relatively simple system in which to explore the role of lateral gene transfer, sequence divergence, and recombination in this diversification.

4.7

Unorthodox transmission modes of endosymbionts in hybrids and the symbiotic origin of speciation

Wolfgang Miller [1], Daniela Schneider[1], Lee Ehrman[2]
[1] Medical University of Vienna, Vienna, Austria, [2] State University of New York, New York, USA

Although not a new idea, recent studies suggest that differences in the composition of symbiotic microbes between hosts can lead to reproductive isolation, and as a consequence also to speciation. Symbiotic bacteria of the genus *Wolbachia* are known to affect their hosts' reproduction in adaptive manners to improve the propagation of the maternally transmitted endosymbiont throughout populations. These reproductive alterations that can result in postmating isolation *via* cytoplasmic incompatibilities, have recently been shown to foster also premating isolation in some host-symbiont associations such as the *Drosophila paulistorum* species complex, giving even more reason to assume that *Wolbachia* can play a significant role in host speciation.

Here we will present most recent data on the involvement of *Wolbachia* in host speciation using two different insect systems, i.e., Neotropical Drosophila and African tsetse flies, both presently under incipient speciation in nature, carrying closely related but incompatible *Wolbachia* strains. We will also show that naturally incompatible and sterile hybrids of both systems can be rescued by means of even mild paternal *Wolbachia*-knockdown before forced mating, giving rise to fertile progeny and stable hybrid lines. Surprisingly such rescued hybrid lines show complete sexual isolation to parental lines as well as unambigous signatures of paternal inheritance of both their cytoplasmic endosymbionts*, i.e.*, of mitochondria plus *Wolbachia*.

4.8

Ecology and Evolution of Deep Sea Coral-Associated Bacterial Communities

Iliana Baums, Richard Dannenberg, Dannise Ruiz, Samuel Vohsen, Charles Fisher
Pennsylvania State University, University Park, PA, USA

Deep-sea corals form habitat for many associated organisms, including diverse microbial communities. Corals in the Gulf of Mexico are regularly exposed to natural and anthropogenic sources of complex hydrocarbons (oil). The goal of this work is to characterize the microbial communities associated with the deep black coral *Leiopathes glaberrima* (Anthipatharia), and understand their role in the response of holobionts to oil and dispersant exposure. Experimental exposure of *L. glaberrima* colonies to low concentrations of oil indicated that the holobiont may be able to metabolize oil, albeit the colonies were sensitive to dispersant and oil-dispersant exposure. 16s-Illumina tag sequencing revealed a relatively species-poor bacterial community dominated by *Endozoicomonas* and a chloroplast-relative. Illumina RNA-seq transcriptomic analysis of the holobiont point to intra-species variability of stress response with white coral colony morphotypes showing more signs of stress then red coral colonies. Liquid Chromatography-Mass Spectrometry (LC-MS) metabolomic analysis similarly yielded differences between red and white colonies exposed to oil and dispersant. Metagenome sequencing and bacterial gene expression analyses are underway to characterize the metabolic potential of the bacterial community in the two color morphotypes of this cosmopolitan foundation species. This is the first comprehensive analysis of the bacterial diversity, metabolic potential and holobiont molecular response of a black coral to oil and oil-dispersant exposure and builds the foundation for understanding the ecology and evolution of bacterial symbionts in other deep-sea corals.

4.9

TEMPORAL SCALES IN THE STUDY OF MICROBIAL ECOLOGY AND EVOLUTION

Purificacion Lopez-Garcia, Marianne Simon, David Moreira, Philippe Deschamps, Gwendal Restoux, Paola Bertolino, Ludwig Jardillier
CNRS & Université Paris-Sud, Orsay, France

The use of molecular tools based on the amplification, cloning and sequencing of marker genes, notably small subunit (SSU) rRNA genes, led to the discovery of a vast diversity of bacteria, archaea and microbial eukaryotes (protists) in natural environments. Recent high-throughput sequencing (HTS) methods, sidestepping cloning biases and providing a much deeper coverage, depict microbial community composition more accurately, making it possible to address questions on microbial biogeography and explore its underlying ecological and evolutionary determinants. In recent years the study of which parameters influence more microbial biogeography, whether environmental selection (the Delft school tenet 'everything is everywhere but the environment selects') or geographical distance (the more classical allopatric speciation) have led to varied and sometimes opposing results. However, most of those studies only focus on spatial scales. We contend that integrating temporal scales, i.e. how communities vary through time within and across sites, is essential to understand microbial biogeography, especially because of dormancy and its influence on microbial community dynamics. We will present a comparative study of protist diversity based on massive amplicon sequencing that we carried out monthly for 2 years in several shallow freshwater systems located at the Natural Regional Park of the Chevreuse Valley, France. The inclusion of temporal scales allowed us to conclude that marine-freshwater transitions are easier for some groups than previously thought and that, despite the large diversity and apparent hectic dynamics of many lineages, the relative proportion between primary producers, free-living heterotrophs and parasites remains constant in similar biotopes.

4.10

Coevolution of Gene Expression Across 322 Marine Microbial Species

Trevor Martin, Hunter Fraser
Stanford University, Stanford, CA, USA

Analysis of gene expression levels across species has revealed many insights into the evolution of gene regulation, including the finding that expression levels of functionally related genes can coevolve. However previous studies have been limited in both the number of species studied, and their phylogenetic breadth. We have developed a novel method for detecting gene expression coevolution, and applied it to RNA-seq data that was uniformly collected from a diverse set of marine microbes, comprised of 670 samples from 322 species representing all major Eukaryotic groups. Building on traditional phylogenetic profiling, our method calculates the cross-species correlation in expression levels between pairs of orthologous groups that cannot be explained by shared ancestry; this excess correlation is evidence of coevolution, and reveals functional relationships that have been previously undetectable. In addition, we leverage extensive environmental and geographic data available for the collected samples to test for associations between variables such as latitude and salinity with gene expression level across these species. In sum, we have applied a new approach for detecting coevolution to the most comprehensive cross-species gene expression data set generated to date, and discovered evidence of gene expression coevolution across the vast diversity of Eukaryotes.

4.11

Conformational strain in newly evolved protein folds

Vlad Kumirov, Emily Dykstra, Taylor Szyszka, Matthew Cordes
University of Arizona, Tucson, AZ, USA

Some protein folds have evolved from others, but it remains unclear whether such radical evolutionary transitions occur cleanly or necessarily involve awkward phases. In the Cro family of bacteriophage transcription factors, a mixed alpha-helix/beta-sheet (alpha+beta) fold evolved from an all alpha-helical (all-alpha) ancestor; the two folds conserve a similar helix-turn-helix subdomain, but differ radically in the structure of a dimerization subdomain. To better understand the evolution of the alpha+beta Cro fold, we constructed rooted phylogenetic trees of this group. We then compared the structures of an all-alpha outgroup (Xfaso 1) and representatives of basal (Pfl 6) and nonbasal (lambda) alpha+beta groups. Interestingly, the helix-turn-helix subdomain of Pfl 6 contains a very rare turn conformation, while homologous sections of Xfaso 1 and lambda have more commonly observed conformations that differ from each other. This turn has the same length in the all-alpha Cro proteins and in the basal alpha+beta group, but a single-residue deletion has occurred in the non-basal alpha+beta groups. Deletion of Thr 15 in Pfl 6 Cro increases the folding stability and reproduces the conformation observed in lambda Cro, while the corresponding deletions or insertions in Xfaso 1 or lambda, respectively, are destabilizing. We propose that the strain observed in Pfl 6 originated during the birth of the alpha+beta fold and was alleviated by a deletion in the non-basal alpha+beta groups. The emergence of a new fold may be accompanied by conformational strain, which further mutational processes may resolve.

4.12

Evolutionary connections and constraints between enzymes in the MBL superfamily

Nobuhiko Tokuriki
University of British Columbia, Vancouver, BC, Canada

The diversity of catalytic activities catalyzed by members of an enzyme superfamily never ceases astonishing us. Whereas it has been postulated that these enzymes evolved from a common ancestor, we know little about evolutionary dynamics of such molecular innovations in the enzyme superfamily. I present our studies to explore evolutionary connectivity and constraints between enzymes of the MBL superfamily. First I discuss how seemingly unrelated catalytic activities observed in the MBL superfamily are connected one to another through promiscuous enzymes. Our systematic analysis of the promiscuous activity profiles revealed that most of chemical reactions within the MBL superfamily are indeed evolutionary connected via promiscuous enzymes. Second, I discuss how these functional connections can be traversed by step-by-step accumulation of mutations. I present our laboratory evolution experiments from various starting points in the MBL superfamily toward the phosphonatase activity. I describe to which extent the same selection pressure can lead to different fitness, phenotypic and genotypic outcomes depending on genetic context of the starting points. I discuss that neofunctionalization of enzymes may be highly contingent, and highly rugged fitness landscapes underlie difference in evolvability. Finally, I discuss how natural evolution has overcome these constraints to expand their functional diversity.

4.13

Combinatorial DNA rearrangement facilitates the origin of new genes in ciliates

Xiao Chen, Laura Landweber
Princeton University, Princeton, NJ, USA

Programmed genome rearrangements in the unicellular eukaryote *Oxytricha* produce a transcriptionally active somatic nucleus from a copy of its germline during development. This process eliminates noncoding sequences that interrupt coding regions in the germline genome, and joins over 225,000 remaining DNA segments, some of which require inversion or complex permutation to build functional genes. This dynamic process of genomic reorganization permits some single DNA segments in the germline to contribute to multiple, distinct somatic genes via alternative processing. Like alternative mRNA splicing, the combinatorial assembly of DNA segments contributes to genetic variation and facilitates the evolution of new genes. In this study, we use comparative genomic analysis to demonstrate that the emergence of alternative DNA splicing is associated with the origin of new genes. Short duplications create alternative gene segments that can be spliced to a set of shared gene segments. Alternative gene segments evolve faster than shared, constitutive segments. Genes with shared DNA segments frequently have different expression profiles, permitting functional divergence. We demonstrate that alternative DNA splicing provides a mechanism for new gene origination, illustrating how programmed genome rearrangement gives rise to evolutionary innovation.

4.14

Mistranslation drives the evolution of robustness in experimental populations of TEM-1 β-lactamase

Sinisa Bratulic [1,2], Florian Gerber[4], Andreas Wagner[1,3]
[1] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [2] The Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3] The Santa Fe Institute, Santa Fe, New Mexico, USA, [4] Institute of Mathematics, University of Zurich, Zurich, Switzerland

How genotypes give rise to phenotypes is a central question in biology. On the molecular level, genotype-phenotype mapping takes place during protein synthesis, or translation. Translation, despite being a fundamental cellular process, is remarkably error-prone. Ribosomes can mistranslate mRNA, introducing phenotypic mutations into nascent polypeptides. Phenotypic errors reduce the stability of proteins and promote misfolding. Misfolded proteins can be toxic, and selection against misfolding is thought to constrain the evolution of highly expressed proteins. The cost of mistranslation can be alleviated by increased accuracy or by increased robustness. Proteins can evolve translational accuracy by adopting synonymous 'highly-fidelity' codons at sensitive site. In contrast, translational robustness can be achieved through accumulation of nonsynonymous mutations that increase protein stability and reduce misfolding. It is not known which of these two mechanisms is more important. To answer this question, we established an experimental system in which an antibiotic resistance gene TEM-1 is subject to evolution in *Escherichia coli* hosts with either wild-type, or mistranslating ribosomes. We evolved TEM-1 populations independently under weak and strong purifying selection, and analyzed them by single-molecule sequencing. We find that mistranslating lines accumulate fewer mutations, and adapt to mistranslation through a combination of expression and stability changes. Under weak purifying selection, the cost of mistranslation is alleviated by reducing the expression of TEM-1. Under strong selection, mistranslating lines increase their robustness by accumulating stabilizing and efficiently purging destabilizing mutations. Our results demonstrate that non-heritable phenotypic mutations influence the evolution of protein stability and expression.

4.15

The genomic architecture of metabolic evolutionary innovation in *Pseudomonas aeruginosa*

Macarena Toll-Riera [1,2], Alvaro San Millan[1], Andreas Wagner[2,3], Craig R. MacLean[1]
[1] Department of Zoology, University of Oxford, Oxford, UK, [2] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zürich, Switzerland, [3] The Santa Fe Institute, Santa Fe, USA

Novel phenotypes that emerge in evolution often have obscure genetic origins. Here we elucidate these origins by evolving populations of the bacterium *P.aeruginosa* in 95 different chemical environments distinguished by their carbon sources. Evolutionary adaptation to some environments requires qualitatively novel traits (innovations), but in other environments mere improvement (optimization) of existing traits suffices. Whole-genome sequencing of evolved clones revealed profound differences in the genetic architecture of these two classes of traits. Evolved clones had few mutations, which facilitates the inference of their causuality role. Innovation was characterized by mutations in regulatory and metabolic genes, and optimization by mutations in sensory and signalling systems. Whereas novel gene duplicates arose at equal frequency during innovation and optimization, mutations in existing duplicates were much more common during innovation, demonstrating that the divergence of existing gene duplicates is a key driver of evolutionary innovation. Both optimization and innovation were associated with fitness costs, but the cost associated with innovation was almost two-fold greater than the cost of optimization, suggesting that pleiotropic trade-offs constrain metabolic innovation. Pleiotropic trade-offs , however, were reduced in clones carrying mutations in recent duplicates, suggesting a link between duplication, mutational robustness and functional innovation. Our observations suggest that phenotypic innovation and optimization can be caused by different classes of genetic change.

4.16

Forecasting Evolutionary Adaptation by Mapping the Innovative Potential of Underground Metabolism

Balint Kintses [1], Richard A Notebart[2], Balazs Szappanos[1], Ferenc Pal[1], Adam Gyorkey[1], Balazs Bogos[1], Viktoria Lazar[1], Reka Spohn[1], Balint Csorgo[1], Csaba Pal[1], Balazs Papp[1]
[1] Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary, [2] Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

A central unresolved issue in evolutionary biology is how to predict evolutionary adaptation occurring through innovation. The current hypothesis is that evolutionary innovations emerge from low-level promiscuous side-reactions that are enhanced by mutations to a physiologically relevant level. Many of these side-activities have been identified in enzymes, however, which of these activities bear genuine innovative potential for an organism in novel environments has remained completely unknown, not least because this issue demands analyses at the level of the entire metabolic network. Here, we integrated all of the known promiscuous activities of *Escherichia coli* enzymes into the metabolic network to provide for the first time a comprehensive computational model for the underground metabolism of an organism. The model allowed us to conduct a genome-wide *in silico* survey to characterize the evolutionary potential of *E. coli* to adapt to hundreds of nutrient conditions. We estimate that at least ~20% of the underground reactions connected to the existing network confer a fitness advantage under specific environments when their activity is increased. Then, these hundreds of conditions were also experimentally tested for growth in a high throughput screen when all of the promiscuous activities of *E coli* were amplified. The good agreement between the *in silico* and the experimental results demonstrates that the genetic basis of evolutionary adaptation is computationally predictable via underground metabolism. The next step is to systematically detect underground reactions in the environment…

4.17

The evolutionary potential of cis-regulatory mutations for sudden changes in development

Paulo Pinto
Institute for Evolution and Biodiversity, Universität Münster, Münster, Germany

Much of the complexity and diversity of eukaryotes is attributed to an elaborate gene regulatory system. Though other mechanisms exist, transcription factors (TFs) are essential in regulating the spatio-temporal expression of genes. TFs promote or hinder the transcription of genes according to cell type, developmental stage, different intercellular signaling and even as a response to external factors.

TFs bind to short, usually between 6 and 8 nucleotides long, (cis-regulatory) DNA motifs. Mutations in these binding sites have the potential to (positively or negatively) alter the strength of binding or prevent it altogether. Using data from the UniProbe database, we construct mutational networks linking all bound sites for each of over 400 TFs divided across 6 species. These neutral networks demonstrate how TF binding is robust to mutations: many mutations can occur without the loss of binding. However, many of the binding sites display non-specificity which provides the possibility of binding other, novel TFs.

We find that the size of the binding networks span orders of magnitude for different TFs, but that there is substantial overlap even for networks of TFs with low sequence identity. We find that there is a subset of TF binding sites which is highly degenerate, binding several (often unrelated) TFs. The non-specificity of some of the cis-regulatory motifs explains how the cis-regulatory network balances itself, resist failure when one or more TFs are shut down. However, it also raises questions regarding the need for the full repertoire of different TFs found in eukaryotic organisms .

4.18

Evolution of gene expression associated with innovation and evolution of feathers

Jacob Musser [1,2], Cong Liang[1,2], Gunter Wagner[1,2], Richard Prum[1,3]
[1] Dept. of Ecology and Evolutionary Biology, Yale University, New Haven, USA, [2] Yale Systems Biology Institute, Yale University, West Haven, USA, [3] Peabody Museum of Natural History, Yale University, New Haven, USA

Feathers are an important innovation that evolved in the ancestors of birds and facilitated the evolution of flight and other facets of modern avian life. However, our understanding of innovation at the level of gene expression and in the feather genetic program is poorly understood, and the relationship of the feather genetic program with that of other skin appendages, such as scales, remains contentious. To address molecular innovation in feathers we performed mRNA-seq on different stages of skin appendage development collected from two distantly related birds, Chicken (*Gallus gallus*) and Emu (*Dromaius novaehollandiae*), and from American Alligator (*Alligator mississippiensis*), a member of the extant clade most closely related to birds. We found that in early development feathers and scutate scales, an overlapping avian scale, share similar patterns of gene expression compared to other scales and claws. This close relationship between feathers and scutate scales in early development, and subsequent unique expression in later feather development, is supported independently by both epidermal and dermal transcriptomes, and by transcriptomes from both chicken and emu. Further, we develop a new model that uses transcriptomes in a phylogenetic framework to show that feathers and scutate scales exhibit higher levels of concerted gene expression evolution then with other skin appendages. Finally, we show that many genes previously studied in feathers actually share expression across multiple skin appendages, and we identify several novel feather genes which we characterize using immunohistochemistry and in-vivo expression perturbation experiments.

4.19

Experimental evolution of increased efficiency through serial propagation in emulsion

Iraes Rabbers [1], Herwig Bachmann[1,2], Bas Teusink[1]

[1] Dept. Systems Bioinformatics, VU University, Amsterdam, The Netherlands, [2] NIZO Food Research and Top Institute Food and Nutrition, Ede, The Netherlands

From an evolutionary perspective it is advantageous for single cells to grow fast -to outcompete competitors- and for microbial populations to use substrate efficiently -to make the most offspring-. Microbial cells are thought to be metabolically efficient when growing slow and inefficient when growing fast, giving rise to the concept “Tragedy of the commons”. There is experimental evidence for this yield/rate tradeoff in i.a. yeast, however these experiments employed culturing methods where the selection pressure is on growth rate (batch) or substrate affinity (chemostat) due to resource competition. A novel method has recently been introduced that compartmentalizes individual cells in a water-in-oil emulsion, thereby privatizing the resources and allowing for yield selection upon serial propagation. In this project we applied this method to *E. coli* wildtype for ~450 generations, leading to selection of mutants with increased cell number, exhibiting increased final optical density, total protein content and biomass. This yield increase appears to result from more efficient substrate usage, mainly the full depletion of pyruvate in contrast with the ancestral strain. Characterization of these strains is ongoing, and should reveal the genetic basis for the high yield phenotype and in addition allow more insight into the existence of a yield/rate tradeoff in *E. coli*.

4.20

Origins of developmental enhancer functions in the neocortex

Deena Emera, Jun Yin, Steven Reilly, James Noonan
Yale University School of Medicine, New Haven, CT, USA

A fundamental question in biology is how new biological structures evolve. Here, we investigate the role of regulatory innovation in the evolution of the neocortex, a structure unique to mammals. Using genome-wide epigenetic profiling, we identified putative enhancers active during cortical development in both human and mouse. By assigning a phylogenetic age to every base in the human genome, we inferred when enhancers with conserved biochemical activity arose. We found that ~20% of conserved enhancers were innovations in the stem mammalian lineage, arising at the same period in vertebrate phylogeny as the neocortex itself. Using a permutation strategy that shuffles enhancers to age-matched genome background, we found that novel mammalian enhancers are overrepresented near genes that function in cell migration, cell signaling, and axon guidance. Novel mammalian enhancers are also overrepresented in modules of co-expressed genes in the cortex that are associated with these pathways, notably ephrin and semaphorin signaling. In contrast to studies in other biological systems, we find that repeat elements likely played a minor role in the evolution of novel mammalian neocortical enhancers. We also show that novel mammalian (and older) enhancers exhibit moderate to high sequence constraint, whereas younger enhancers exhibit weak constraint despite conservation of their biochemical activity. Our results provide specific insight into the regulatory elements and pathways potentially involved in the emergence of the neocortex, and general insight into regulatory innovation in mammals.

441A

Challenges in ancient microbiome reconstruction using 16S rRNA gene amplification

Kirsten A. Ziesemer [1], Allison Mann[2], Menno Hoogland[1], Hannes Schroeder[3], Krithi Sankaranarayanan[2], Corinne Hofman[1], Christina Warinner[2]

[1] Leiden University, Leiden, The Netherlands, [2] University of Oklahoma, Norman, OK, USA, [3] University of Copenhagen, Center for Geogenetics, Copenhagen, Denmark

Dental calculus, formed by periodic mineralization of dental plaque, is an ideal biomolecular reservoir as we seek to understand the ancient oral microbiome. To date, characterization of the ancient oral microbiome, as well as the ancient gut microbiome (i.e., coprolites), has primarily been accomplished through a phylotyping approach involving targeted amplification and sequencing of variable regions in the 16S rRNA gene. Specifically, the V3 region (*E coli* 341-534) of this gene has been identified through *in silico* and *in vitro* analyses as an excellent candidate for ancient DNA amplification and community reconstruction. Nevertheless, in practice this phylotyping approach often results in unusual taxonomic frequency data. We use targeted (amplicon) and non-targeted (shotgun metagenomics) sequencing methods on four archaeological dental calculus samples to better understand these discrepancies. The four samples were chosen from diverse geographic and temporal contexts: Middenbeemster, Netherlands (159 BP); Guadeloupe, Caribbean (700 BP); Samdzong, Nepal (1900 BP); and Camino del Molino, Spain (4000 BP). Through comparisons of microbial taxonomic counts from paired amplicon and shotgun sequencing datasets, we show preferential amplification of archaea and the candidate bacterial phylum TM7 and underamplification of Spirochaetes and many important bacterial genera (e.g., *Streptococcus*) in amplicon datasets. Through informatics analysis, we demonstrate that extensive length polymorphisms in the V3 region are a consistent and major cause of amplification dropout and taxonomic bias in ancient microbiome reconstructions based on amplicon sequencing. We conclude that systematic amplification bias confounds attempts to accurately reconstruct microbiome taxonomic profiles using 16S rRNA V3 amplicon data.

442B

Gut bacterial diversity in Plasmodium-infected and Plasmodium-uninfected *Anopheles minimus*

Wunrada Surat [1], Wuttichai Mhuantong[2], Duangjai Sangsrakru[2], Uraiwan Arunyawat[1], Anchanee Kubera[1], Wirulda Pootakham[2], Theeraphap Chareonviriyaphap[3]
[1] Faculty of Science, Kasetsart University, Bangkok, Thailand, [2] National Center for Genetic Engineering and Biotechnology, Pathum Thani, Thailand, [3] Faculty of Agriculture, Kasetsart University, Bangkok, Thailand

Anopheles minimus is one of the main malaria vectors in Thailand. Plasmodium transmission depends primarily on the success of the parasite survival in the mosquito's gut. Several factors affect the development of Plasmodium in the mosquito, including the gut microbiota. Nevertheless, there have been very few studies on the bacterial communities residing in the gut of *An. minimus*. Here, we used a culture-independent method, metagenomics, to identify microbiota and compare the bacterial communities in the gut of Plasmodium-infected and Plasmodium-free mosquitoes. Fifty three genera within four phyla were detected and 14 of them were discovered in malaria vectors for the first time. The bacterial diversity and the profile of the gut communities between the Plasmodium-infected and those of the uninfected mosquitoes were quite different, with the diversity in the gut of the uninfected mosquitoes being much higher than that of the infected counterpart. Gammaproteobacteria were prevalent in the infected *An. minimus* while betaproteobacteria were the most abundant in the uninfected mosquitoes. In addition, the result also showed that certain bacterial genera were present only in the gut of the uninfected (26 genera) or the infected (7 genera) mosquitoes. The data contributed that there are a relationship between plasmodium infection and bacterial communities in the mosquito's gut.

Keywords: Anopheles minimus, bacterial community, metagenomics, *Plasmodium*

443C

Characterisation of gut microbiomes in a pest fruit fly *Ceratitis capitata* reveals a heritable and potentially symbiotic microbe

Philip Leftwich, Tracey Chapman
University of East Anglia, Norwich, UK

Studies of symbiotic microbial communities in the gut have been advancing rapidly, with increasing rewards, in the last few years as the ability to analyse large sequence data grows. The Mediterranean Fruit Fly, *Ceratitis capitata* is a well-studied organism for ethology, ecology, and pest management, but there are to date no comprehensive studies of the gut microbiota of this costly agricultural pest.

We have used culture independent metagenomics analysis based on 16s rRNA gene sequences on this insect to compare microbial gut communities from wild populations through to long established laboratory populations, at all stages of development and multiple nutritional manipulations.

In contrast to other fruit flies such as *Drosophila spp.*; which have an environmentally acquired microbial community; we find instead a community dominated by a single OTU from *Klebsiella oxytoca,* which appears to demonstrate a stable and potentially heritable relationship between microbe and host organism.

This information on a putatively symbiotic relationship between gut microbe and host will be extremely useful for future studies on the interactions between *C.capitata* and *K.oxytoca* for host health, evolutionary implications and Integrated Pest management techniques such as the Sterile Insect Technique (SIT) or Release of Insects carrying a Dominant Lethal (RIDL).

444D

Unravelling the genetic structure and diversity of *Pseudomonas syringae* associated with kiwifruit leaves using MLST

Christina Straub [1], Honour McCann[1], Paul Rainey[1,2]
[1] NZ Institute for Advanced Study, Auckland, New Zealand, [2] Max Planck Institute for Evolutionary Biology, Plön, Germany

Epidemics caused by emerging plant diseases have a tremendous effect on primary food production. The plant pathogen *Pseudomonas syringae* pv. *actinidiae* (*Psa*) is the causative agent of bacterial canker on kiwifruit (*Actinidia spp.)* and in 2008 a virulent and aggressive form of *Psa* caused a global pandemic.

Psa is part of the *Pseudomonas syringae* complex, a variety of plant pathogens associated with the phyllosphere and exhibiting a diverse host range. Bacteria densely populate the phyllosphere and various commensal and/or mutualistic members of *P. syringae* can be present. This leads to questions concerning interactions between closely related members of the *P. syringae* species complex and whether they can aid the emergence of a new virulent strain.

The aim of my study is to gain an understanding of the genetic diversity and phylogenetic relations of *Pseudomonas syringae* occupying the leaf surface in kiwifruit.

Bacterial isolates from *Psa* infected and non-infected orchards from two cultivars were collected during the growing season 2013/14. Subsequently Multi Locus Sequence Typing (MLST) of four housekeeping genes was done for a total of 140 *P. syringae* isolates, as well as 90 *Pseudomonas sp.* isolates.

Preliminary data suggests that the collection of strains encompasses various representatives of four major monophyletic groups of *Pseudomonas syringae*. Curiously we have recovered a clade of pv. *phaseolicola* isolates in NZ, which has similarly been shown for a MLST analysis of a collection of strains isolated in China. Our analysis will help to understand the effect of microevolutionary processes on patterns of *Psa* diversity.

445A

Environmentally induced increase of rearrangement rates in radiation resistant prokaryotes

Jelena Repar [1,2], Fran Supek[2,3], Davor Zahradka[2], Ksenija Zahradka[2]
[1] MRC Clinical Sciences Centre, Imperial College, London, UK, [2] Ruder Boskovic Institute, Zagreb, Croatia, [3] Centre for Genomic Regulation, Barcelona, Spain

The environment shapes composition but also the evolution of microbial communities. This influence can take various forms; for example, sharing of ecologically relevant genes through horizontal gene transfer, or convergent evolution to adapt to the same niche. But the environment can also have a much more direct role on evolutionary processes. A good example are extreme environments where physical factors can directly impact DNA integrity, mutational processes and consequently, genome evolution.
Here, we show marked differences in large scale patterns of genome evolution between radiation resistant prokaryotes and their non-resistant cousins. Specifically, controlling for phylogenetic non-independence, we demonstrate that rearrangement rates are systematically higher in radiation resistant compared to non-resistant species, in phyla as diverse as *Deinococcus-Thermus*, *Proteobacteria* and *Euryarchaeota*. Our results suggest that patterns of evolution shared by members of the same ecological niche can be a direct physcial consequence of the niche and need not arise from convergent adaptive evolution.

446B

The Miridae microbiome: Comparative analysis of the bacteria found within plant bugs

Ferran Palero [1], Aurelie Blin[1], Michela Ion-Scotta[1], Bastien Quaglietti[1], Nicolas Ris[1], Jean-Claude Streito[2], Thibaut Malausa[1]

[1] Institute Sophia-Agrobiotech, Sophia-Antipolis, France, [2] CBGP, Montferrier-sur-Lez, France

Plant bugs are heteropteran insects that belong to the large family Miridae. Most of the best known mirids are either harmful to agriculture (sap-sucking insects and vectors of plant diseases) or constitute auxiliary insects that prey on agricultural pests such as mites, psyllids, aphids and thrips. The present study describes for the first time the whole-microbiome composition of several species of the biological control auxiliary Macrolophus, including populations sampled from different localities and several host plants. The microbiome plays a crucial role in the growth, development and environmental adaptation of the insect host, and it is recognized as a major genetic resource of new molecules for the bio-processing industry. Here, we investigate the microbial diversity of this important group by using 16S rDNA amplicon re-sequencing on the MiSeq platform. The main bacterial taxa found include several species of Alphaproteobacteria belonging to the families Rickettsiaceae and Anaplasmataceae (with well-known genera such as Neorickettsia and Wolbachia); Betaproteobacteria belonging to the Oxalobacteraceae; and Gammaproteobacteria belonging to the Enterobacteriaceae (a large family of Gram-negative bacteria that includes many symbionts) and Moraxellaceae. Our results indicate that bacterial community diversity and composition is influenced by both geographic and host-plant variation and expand recent studies based on PCR-cloning methods by the discovery of insect-bacteria associations previously unknown within the Miridae.

447C

The timetree of prokaryotes: new insights into their evolutionary history

Julie Marin, Sudhir Kumar, S. Blair Hedges
Temple University, Philadelphia, PA, USA

In recent years, an increasing number of studies have focused on macroevolution, almost exclusively involving eukaryotes. However the first organisms on Earth were prokaryotes and their evolutionary history can reveal the geological and climatological processes that occurred in the early history of our planet. The evolutionary history of prokaryotes is still poorly understood because of difficulties inherent to this group, such as horizontal gene transfer, rather than a lack of interest. In order to improve our understanding of their evolutionary history, we produced a comprehensive timescale of these organisms and explored patterns of speciation and extinction. Molecular information from a recently released 16S rDNA dataset, along with topological constraints from previous studies, were combined to produce a timetree of 11,170 species. The rate of diversification over time and between lineages was evaluated as well as the general model of diversification (saturated or expanding). We then focused on how speciation and recombination processes specific to prokaryotes have influenced their patterns of diversification.

448D

Phylogenetic analysis of gut bacterial communities in mammals: disentangling the differential effects of host genetics and diet.

Mathieu Groussin [1], Florent Mazel[2], Eric Alm[1]
[1] Massachusetts Institute of Technology, Cambridge, MA, USA, [2] Universite Grenoble & CNRS Alpes, Laboratoire d'Écologie Alpine, Grenoble, France

Microbial communities (microbiota) living in Mammal guts are composed of thousands of species – most of them bacterial – that are essential for host physiology, immunity and diet. It has been shown that host genetics (e.g. immune system determinants) and diet are the two major factors driving the composition of gut microbiota. However, major questions remain: (i) the relative contribution of host genetics and diet at short and long time scales is highly debated and the two processes are not well characterized, (ii) host genetics may drive the composition through co-evolution with bacterial lineages and/or through niche selection, with closely-related hosts selecting similar bacteria from the environment. Here, we show that host genetics and diet are antithetic factors and do not drive the bacterial composition at the same taxonomic scale. Diet determines what lineage is present or not at deep bacterial phylogenetic levels through gain or loss of lineages creating nested communities. Host genetics, however, selects the lineages at finer scales through true turnover of lineages, consistent with a more stringent selection of tolerated antigens. Finally, we highlight that, at the bacterial taxonomic scale where the correlation between microbiota composition variation and host distances is maximal, bacterial and host phylogenies show congruent patterns of diversification, consistent with a co-evolutionary scenario between Mammals and their gut symbionts. Our results shed light on the long-timescale evolutionary dynamic of gut bacterial communities, which are multi-layered phylogenetic structures shaped very differently by host genetics and diet.

449A

Crowded Growth Promotes Semi-Stable Coexistence on Shared Resources in Laboratory Yeast

Evgeni M Frenkel [1], Michael M Desai[1], Michael J McDonald[1], J David Van Dyken[2], Katya Kosheleva[1], Gregory I Lang[3]

[1] Harvard University, Cambridge, MA, USA, [2] University of Miami, Coral Gables, FL, USA, [3] Lehigh University, Bethlehem, PA, USA

Identifying the mechanisms that create and maintain biodiversity is a central challenge in biology. Stable diversification of microbial populations often requires the evolution of differences in resource utilization. Here we report spontaneous diversification maintained by an alternative mechanism: crowding avoidance. During experimental evolution of budding yeast populations, we observed the repeated appearance of "adherent" lineages that disperse over a greater range, reducing interference competition for nutrients among kin at the cost of a slower maximum growth rate. This strategy for coping with negative density-dependent growth leads to biphasic convergence towards a stable equilibrium frequency over repeated cycles of growth, crowding, and dispersal. However, further coevolution of the adherent and non-adherent types can perturb and eventually destroy their coexistence over longer timescales. We introduce a simple mathematical model of this semi-stable coexistence, which explains this interplay between ecological and evolutionary dynamics. Since crowded growth generally limits nutrient access in biofilms and solid tumors, the mechanism we report here may be broadly important in maintaining diversity in these natural environments.

450B

Ultrafast alignment and analysis of metagenomic DNA sequence data from the Tyrolean Iceman using MALT

Alexander Herbig [1], Frank Maixner[2], Kirsten Bos[1], Benjamin Buchfink[3], Chao Xie[4], Albert Zink[2], Johannes Krause[1], Daniel Huson[3]

[1] Department of Paleogenetics, University of Tübingen, Tübingen, Germany, [2] Institute for Mummies and the Iceman, Bolzano, Italy, [3] Center for Bioinformatics, University of Tübingen, Tübingen, Germany, [4] Life Sciences Institute, National University of Singapore, Singapore, Singapore

Modern sequencing technologies have led to the production of vast amounts of DNA sequence data from metagenomic samples revealing the complexity of microbial communities in unprecedented detail. These analyses require high-throughput computational methods that allow for an extremely fast processing of sequencing data while retaining a high level of sensitivity and precision.

Here we present MALT a program for the fast alignment of DNA sequencing reads to a database of all microbial reference genomes available on GenBank. MALT is able to process hundreds of millions of reads within only a few hours, which makes the analysis of a whole metagenomic sequencing run possible within a single day. Its tight integration with the metagenome analysis software MEGAN allows for an assignment of single reads to different taxonomic levels with precision that facilitates the identification and quantification of specific bacterial species or strains. These analyses can be performed in a comparative manner for studying the dynamics of microbial communities over time, or from different habitats or hosts. Especially the human microbiome is of major interest as it is comprised not only of a large number of commensals, but potentially also pathogens that have evolved with their human host. To gain insights into these evolutionary relationships, the field of paleogenetics aims to study ancient DNA extracted from archaeological remains.

In this context we demonstrate MALT by its application to the metagenomic analysis of two ancient microbiomes from the 5,300 year old Tyrolean Iceman based on oral cavity and lung samples.

451C

GETTING AWAY WITH ORDER − GENOME EVOLUTION OF THE LACTIC ACID BACTERIA INHABITING THE HONEYBEE GUT

Daniel Tamarit [1], Kirsten M Ellegaard[1], Johan Wikander[1], Tobias Olofsson[2], Alejandra Vásquez[2], Siv G E Andersson[1]

[1] Uppsala University, Uppsala, Sweden, [2] Lund University, Lund, Sweden

The honeybee microbiome has gained a prime focus in current symbiosis research, becoming a symbiotic community model due to its stability and probable evolutionary relevance. *Lactobacillus kunkeei* is an abundant inhabitant of the honeybee foregut and food products and has been suggested to protect its hosts via the production of antimicrobial compounds. We have sequenced 13 *L. kunkeei* genomes from different bee hosts and studied their evolution and genome structure. We found lack of co-cladogenesis between host and symbiont, and that multiple genotypes could be recovered from the same host species. We also show that recombination occurs within phylogroups but rarely between phylogroups, regardless of co-ocurrence in the same hosts. Their genomes have evolved by massive reduction, and gained genes that are good candidates for adaptations to the honeybee niche. Examining their genome organization uncovered a novel type of architecture in which vertically-inherited essential genes involved in information processing cluster near the terminus of replication, while the origin of replication harbors newly acquired genes and genes for metabolic functions, transporters and secreted proteins. The latter set includes four to five tandemly duplicated genes for a family of secreted proteins of 3,000 to 9,000 amino acids, which may be relevant for the symbiotic phenotype. We surveyed other *Lactobacillus* genomes and found that this compartmentalized nature is ancestral but remarkably intensified in two unrelated clades of bee symbionts. These results have implications for our understanding of bacterial genome evolution and host-adaptation, and shed additional light on the ecology of the bee microbiome.

452D

Metagenomic analyses of rumen microbiomes reveal functional isoform diversity driving microbial niche differentiation for nutrient acquisition and use.

Francesco Rubino[1,2], Sinead Waters[2], David Kenny[2], Matthew McCabe[2], Ciara Carberry[3], Christopher Creevey [1]

[1] Aberystywth University, Aberystwyth, UK, [2] Teagasc, Grange, Dunsany, Co. Meath., Ireland, [3] University College Dublin, Belfield, Dublin 4, Ireland

Metagenomics has provided insights into the species composition and function of microbial communities and revealed that many species even in highly competitive environments seem to share the same sets of genes for acquiring and utilising nutrients. This questions whether niche differentiation between microbes under competitive stress can exists and if so, how it is maintained.

Rumen microbes exist in a highly competitive, anaerobic environment, and compete for access to the limited nutrients available in the biomass consumed by the host. The rumen microbiome affects the efficiency and health of the host, and has been mined for novel enzymes and anti microbial peptides and is known as a potent contributor to greenhouse gas emissions. A better understanding of the structure and function of this community has large potential benefits.

We generated separate metagenomic libraries from 14 Limousin cows. Following assembly, gene prediction and taxonomic assignment, SNPs were identified and functional isoform diversity (pN/pS) was calculated for each gene. The independent calculations from each sample were used to calculate confidence intervals for testing for diversity of function between microbes.

Significant differences in functional diversity between species existed in genes involved in carbon, amino-sugar and nucleotide sugar metabolism, suggesting adaptation to utilizing different routes for nutrient acquisition in the rumen. This suggests that a greater understanding of the rumen microbiome and considering macro-ecological concepts, such as successional change and adaptation, are likely to improve strategies for increasing the efficiency and health of the host and reducing greenhouse gas emissions.

453A

Clustering large scale 16S metagenomic sequences in the cloud environment

Jeongsu Oh[1], Chi-Hwan Choi[2], Min-Kyu Park[2], Soon Gyu Hong[3], Wan-Sup Cho[2], Kyung Mo Kim[1]

[1] Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea, [2] Chungbuk National University, Cheongju, Republic of Korea, [3] Korea Polar Research Institute, Incheon, Republic of Korea

Clustering 16S rRNA sequences generates operational taxonomic units (OTUs) to which individual sequences are assigned. Since the number of OTUs is the direct measure of microbial diversity, sequence clustering becomes the most important step when studying microbial ecology of a given environment. Even though many existing programs are available for clustering 16S sequences produced by next generation sequencing platforms, CLUSTOM was recently shown to be the most accurate. However, the original version of CLUSTOM takes a lot of time to cluster a large number of 16S sequences since its performance depends on the scale-up of CPU and memory resource of a single computing node. In order to overcome this limitation, we developed a new version named CLUSTOM-CLOUD that can be deployed in a cloud environment. By using the In-Memory Data Grid (IMDG) technique that has become popular recently in the industrial field of information technology but is not adopted yet in the field of bioinformatics, CLUSTOM-CLOUD is an order of magnitude faster than the original version and enables to cluster big 16S sequences. The IMDG-based CLUSTOM-CLOUD has novel features in comparison to bioinformatics tools based on MPI or MapReduce: (I) rapid processing of large amount of data by allocating input data on IMDG that integrates random access memory of distributed computing nodes into a shared memory pool; (II) platform independent and supporting various computing environments; (III) provides highly scalable computing environment without any complicated installation or setting. These features distinguish CLUSTOM-CLOUD from existing cloud-based bioinformatics tools.

454B

High resolution genotyping of *Xanthomonas oryzae* helps resolving epidemiological and evolutionary questions

Lucie Poulin [1], Pierre Griegel[2], Maxime Magne[2], Lionel Gagnevin[1,2], Olivier Pruvost[2], Alexis Deereper[1], Pierre Lefeuvre[2], Valerie Verdier[1], Christian Vernière[2], Ralf Koebnik[1]
[1] IRD, Montpellier, France, [2] CIRAD, Saint Pierre, Reunion

Xantomonas oyzae is a pathogenic bacteria species that is a devastating rice-crop pathogen in certain tropical areas such as southern Asia and that is lately becoming of big issue in western Africa. Two different pathovars described as vascular and superficial –parenchymal- in the rice plant are known as respectively pv. *oryzae* (*Xoo*) and pv. *oryzicola* (*Xoc*). Epidemiosurveillance routine of those bacteria has to be settled in order to type and link strains at different geographic scales as well as to characterize outbreaks and epidemics. For that purpose, several high-resolution molecular typing approaches were explored. Firsty, VNTR (Variable Number of Tandem repeats)-based molecular markers were studied. For *X. oryzae*, multilocus VNTR analyses (MLVA) were developed for the three known lineages of *X. oryzae*. Large study of *X. oryzae* strains by MLVA allowed us to characterize genetic clonal complexes generally associated with new epidemics. Secondly, at the Philippines scale, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) associated spoligotyping as well as genomic SNPs, and minisatellites were explored for *Xanthomonas oryzae* pv. *oryzae* (*Xoo*). Our preliminary results have shown that the composition and the order of CRISPR arrays could be useful for a spoligotyping approach. On the other hand, minisatellites markers and SNPs revealed a significant correlation with strain races and could be further used for wide Philippine or Asian population studies in order to investigate in parallel race and population emerging and re-emerging.

455C

Parasite diversity and MHC I and II variation in three-spined stickleback (*Gasterosteus aculeatus*) populations across the Northern Hemisphere.

Naga Apoorva Kasala, Mahesh Panchal, Martin Kalbe, Manfred Milinski, Irene E. Samonte
Max Planck Institute for Evolutionary Biology, Ploen, Schlesweig Holstein, Germany

The three-spined stickleback (*Gasterosteus aculeatus*) is a good model system to investigate local adaptation mainly because of its repeated and independent postglacial colonization history that enables sticklebacks to thrive in various marine and freshwater habitats in the northern hemisphere. The different habitats harbour distinct parasite communities that are proposed to facilitate local adaptation in sticklebacks, particularly between lake and river populations. In order to advance our knowledge about parasite-mediated local adaptation in sticklebacks, we sequenced two key immune genes (MHC Class I and Class II) and screened the parasites [macro- and microparasites (bacterial community)] of three-spined sticklebacks from four lake-river pairs, two ancestral marine and one extra lake populations from Europe and North America. Specifically, we would like to determine MHC Class I and Class II diversities and relate these to the parasite (macro- and microparasites) communities thriving in the different stickleback populations. Our expectation is that selection pressure exerted by parasites to their stickleback hosts will be reflected in the allelic polymorphism of the MHC genes.

456D

The genome of the protozoan parasite *Cystoisospora suis* and its implications for vaccine discovery

Nicola Palmieri [1], Hanna Worliczek[1], Damer Blake[2], Anja Joachim[1]
[1] *Institute of Parasitology, University of Veterinary Medicine, Vienna, Austria,* [2] *Royal Veterinary College, University of London, London, UK*

Cystoisospora suis is a protozoan parasite that belongs to the phylum Apicomplexa and it is phylogenetically related to well-known protozoa of medical importance, such as *Toxoplasma* and *Plasmodium*. *C. suis* induces coccidiosis in piglets and has a worldwide economical impact on pig production. In this project we sequenced for the first time the genome of *C. suis* with the intent of finding potential vaccine candidates using an in silico based screening approach. The genome assembly has an estimated length of ~84Mb and contains about 8000 protein-coding genes, annotated by combining ab initio and orthology gene predictions. To identify proteins eligible as vaccine candidates, we applied to each predicted protein an array of programs to compute immunogenic features, such as transmembrane domains and signal peptides for membrane localization and secretion. With the aid of the software Vacceed we combined these features in a machine learning fashion and identified a preliminary list of 615 candidates, including known proteins involved in the invasion process, such as the apical membrane antigen 1 (AMA-1), proteins expressed in apicomplexan specific organelles and 131 uncharacterized proteins with no clear orthologs. We are going to refine this list by incorporating predictions of epitopes recognized by pig MHC proteins. We plan to generate RNA-Seq data for different developmental stages in order to validate the gene predictions and find novel genes with the long-term perspective to incorporate these new genomic data into the eukaryotic pathogen database EuPathDB.

457A

Single cell genomics of a rare environmental alphaproteobacterium provides unique insights into Rickettsiales evolution

Joran Martijn [1], Frederik Schulz[2], Katarzyna Zaremba-Niedzwiedzka[1], Johan Viklund[1], Ramunas Stepanauskas[3], Siv Andersson[1], Matthias Horn[2], Lionel Guy[1], Thijs Ettema[1]
[1] Uppsala University, Uppsala, Uppland, Sweden, [2] University of Vienna, Vienna, Austria, [3] Bigelow Laboratory for Ocean Sciences, East Boothbay, Massachusetts, USA

The bacterial family Rickettsiaceae includes a group of well-known etiological agents of many human and vertebrate diseases, including epidemic typhus-causing pathogen *Rickettsia prowazekii*. Due to their medical relevance, rickettsiae have attracted a great deal of attention and their host-pathogen interactions have been thoroughly investigated. All known members display obligate intracellular lifestyles, and the best-studied genera, *Rickettsia* and *Orientia*, include species that are hosted by terrestrial arthropods. Their obligate intracellular lifestyle and host adaptation is reflected in the small size of their genomes, a general feature shared with all other families of the Rickettsiales. Yet, despite that the Rickettsiaceae and other Rickettsiales families have been extensively studied for decades, many details of the origin and evolution of their obligate host-association remain elusive. Here we report the discovery and single cell sequencing of '*Candidatus* Arcanobacter lacustris', a rare environmental alphaproteobacterium that was sampled from Damariscotta Lake that represents a deeply rooting sister lineage of the Rickettsiaceae. Intriguingly, phylogenomic and comparative analysis of the partial '*Candidatus* Arcanobacter lacustris' genome revealed the presence chemotaxis genes and vertically inherited flagellar genes, a novelty in sequenced Rickettsiaceae, as well as several host-associated features. This finding suggests that the ancestor of the Rickettsiaceae might have had a facultative intracellular lifestyle. Our study underlines the efficacy of single cell genomics for studying microbial diversity and evolution in general, and for rare microbial cells in particular.

458B

Neutral Models of Microbiome Evolution

Qinglong Zeng, Jeet Sukumaran, Steven Wu, Allen Rodrigo
Duke University, Durham, NC, USA

There has been an explosion of research on host-associated microbial communities (=microbiomes). Much of this research has focused on surveys of microbial diversities across a variety of host species, including humans, with a view to understanding how these microbiomes are distributed across space and time, and how they correlate with host health, disease, phenotype, physiology and ecology. Fewer studies have focused on how these microbiomes may have evolved. In this paper, we develop an agent-based framework to study the dynamics of microbiome evolution. Our framework incorporates neutral models of how hosts acquire their microbiomes, and how the environmental microbial community that is available to the hosts is assembled. Most importantly, our framework also incorporates a Wright-Fisher genealogical model of hosts, so that the dynamics of microbiome evolution is studied on an evolutionary timescale. Our results indicate that the extent of parental contribution to microbial availability from one generation to the next significantly impacts the diversity of microbiomes: the greater the parental contribution, the less diverse the microbiomes. In contrast, even when there is only a very small contribution from a constant environmental pool, microbial communities can remain highly diverse. Finally, we show that our results are consistent with empirical patterns, and we discuss testable predictions about the types of processes that operate to assemble microbiomes over evolutionary time.

459C

Comparison of gut microbiomes of oriental river prawn (Macrobrachium nipponense) dwelling in river and lake

Daryi Wang, Yueh-Yang Pao
Academia Sinica, Taipei, Taiwan

The gut microbial community is one of the richest and most complex ecosystem on earth, the intestinal microbes play an important role in host development and health. Next generation sequencing approaches rapidly produce millions of short reads which enable the investigation on culture independence basis are now popular for exploring microbial community. However, the gut microbiome in fresh water shrimp is unexplored, to explore gut microbiomes of the oriental river prawn (Macrobrachium nipponense), and investigate the host genetic and ecotypic effects on the microbial composition, 454 pyrosequencing based on 16S rRNA gene were performed. We collected six groups of samples which consist of shrimp from two populations, shrimp obtained from rivers and lakes, and one sister species (M. asperulum) as out group. We found that Proteobacteria is the major dominant phylum in oriental river prawn, followed by Firmicutes and Actinobacteria. Compositional analysis showed microbial divergence between two shrimp species is higher than that in one shrimp species collected from river and lake. Hierarchical clustering also showed that host genetics contributed greater impact to the divergence of gut microbiome than that from host ecotypes. This finding was also supported by the functional prediction from the metagenomic data showing that the two shrimp species still shared the same type of gut metagenome function, reflecting a similar ecotype in their gut environments. In conclusion, this study provides the first investigation on gut microbiome of fresh water shrimp, and support the hypothesis of host species-specific signature of bacterial community composition.

460D

Evolution of antibiotic resistance in the absence of antibiotic exposure

Sophia Katz, Ruth Hershberg
Technion-Israel Institute of Technology, Haifa, Israel

The emergence and spread of antibiotic resistance are major challenges to human health. It was shown that the frequency of antibiotic resistant bacteria can increase following exposure to growth limiting stresses that are unrelated to antibiotic exposure, such as starvation. We have used whole-genome sequencing to show that contrary to long held assumptions, such increases in resistance frequency are not due to increased mutagenesis. Rather, certain resistance mutations to different antibiotics are favored by natural selection because they improve growth under starvation. We further show that increases in resistance frequency under starvation depend on the manner in which cells are grown. Such increases occur when cells are grown in micro-colonies, but not when they are grown in broth. Furthermore, increases in resistance frequencies and the fitness of resistant mutants depend on the master regulator of the stress response, rpoS.

461A

Identification of viruses and viroids infecting *Prunus* species using next-generation sequencing

Yeonhwa Jo, Hoseong Choi, Won Kyong Cho
Seoul National University, Seoul, Republic of Korea

Stone fruit trees—such as apricot, peach and plum—are important horticultural plant species belonging to the genus *Prunus*. So far, more than 20 viruses and viroids infecting *Prunus* species have been identified. In this study, we identified viruses and viroids infecting apricot, peach and plum by next-generation sequencing. To do so, we generated 18 libraries, which were prepared from five apricot cultivars, six peach cultivars and seven plum cultivars in Korea. The 18 libraries were sequenced by Illumina pair-end sequencing using the HiSeq 2000 system. The obtained raw data were *de novo* assembled by the Trinity program. The assembled transcripts were blasted against viral reference genome data to identify viruses and viroids. Based on the blast results, we identified various viruses and viroids. For example, nine viruses and two viroids were identified from five peach cultivars, while five viruses and one viroid were identified from eight plum cultivars. In addition, we identified several novel viruses in the order *Tymovirales*. Interestingly, many viral genomes were *de novo* assembled from transcriptome data. The numbers of infected viruses ranged from one to nine. Some viruses infected at least nine different cultivars, while others had a narrow host range. Moreover, we calculated the amount of viral RNA and copy numbers in each of the host plants, and we analysed single nucleotide variations in each sample. Taken together, this is the first comprehensive viral metagenomics analysis of stone fruit trees revealing viral communities in species in the genus *Prunus*.

462B

Viruses and viroids infecting a grapevine identified by viral metagenomics

Yeonhwa Jo, Hoseong Choi, Won Kyong Cho
Seoul National University, Seoul, Republic of Korea

The grapevine is a widely grown fruit tree providing materials for wine production. Many kinds of viruses and viroids infect grapevine cultivars, causing serious damage to the plant. In this study, we used a viral metagenomics approach with next-generation sequencing to identify viruses and viroids infecting a single grapevine. To do so, we isolated total RNA from the grapevine cultivar Cabernet Sauvignon, and a library was prepared for a transcriptome study using the Illumina HiSeq2000. We assembled sequenced reads using the Trinity program. The assembled contigs served as subjects for the blast search against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide database. We identified large numbers of contigs homologous to various viruses and viroids based on the blast results. The amount of viral RNA in the grapevine transcriptome was 2% (57,395 reads). We identified a total of six viruses and two viroids that infect the cultivar Cabernet Sauvignon. Moreover, nearly complete genome sequences for several viruses were de novo assembled. In summary, our study provides an efficient pipeline to identify plant viruses and viroids infecting a single grapevine using a viral metagenomics approach. This method can be successfully applied to reveal viral genomes and populations in other plant species.

463C

Feeding on microbiomes: how earthworms change the taxonomic and functional composition of livestock gut microbiomes

Marcos Perez-Losada [1,2], Manuel Aira[3], Seth Bybee[4], Jorge Dominguez[3]
[1] CIBIO-InBIO, Vairao, Portugal, [2] Computational Biology Institute, Ashburn, VA, USA, [3] Universidade de Vigo, Vigo, Spain, [4] Brigham Young University, Provo, UT, USA

Earthworms play a key role in nutrient cycling by interacting with microorganisms and thus accelerating organic matter turnover in soil systems. As detritivores, earthworms can ingest and digest a mixture of dead organic matter and microorganisms, like those found in animal feces. Here, we couple 454 pyrosequencing and metagenomic analysis to study how earthworm gut passage changes the bacterial composition and function of the microbial communities residing in the feces of cow, horse and pig (i.e. livestock gut microbiomes). Towards this aim, we fed earthworms of the species *Eisenia andrei* with manure from the three farm animals and compared the taxonomic and functional composition of the manure microbiomes before and after passage through the earthworm guts (i.e. earthworm microbiomes). Earthworm microbiomes showed a smaller diversity than the livestock microbiomes they fed on. Livestock microbiomes strongly differed at their taxonomic and functional composition, but these differences were markedly reduced once transformed in earthworm microbiomes. The core microbiome of *E. andrei* was comprised of 30 OTUs of the phyla Actinobacteria and Proteobacteria, while the livestock core microbiomes were composed mainly of Firmicutes and Bacteroidetes. Our results suggest that earthworms build up their gut microbiome by selecting from the pool of ingested bacteria. As a consequence, earthworms enhance rates of decomposition and nutrient turnover twice: first during gut transit and then once the microbiomes are released as earthworm casts.

464D

Microbial communities from microbialites: phylogenetic composition and metabolic potential assessed by metagenomics

Aurélien Saghaï [1], Yvan Zivanovic[2], David Moreira[1], Karim Benzerara[3], Paola Bertolino[1], Purificación López-García[1]

[1] Unité d'Ecologie, Systématique et Evolution, Centre National de la Recherche Scientifique (CNRS) UMR8079, Université Paris-Sud, Orsay, France, [2] Institut de Génétique et Microbiologie, CNRS UMR8621, Université Paris-Sud, Orsay, France, [3] Institut de Minéralogie et de Physique des Matériaux et de Cosmochimie, CNRS UMR7590, Université Pierre et Marie Curie, Paris, France

Microbial mats are structures made of phylogenetically and metabolically diverse microbial communities. Particular mats, called microbialites and including the well-known stromatolites, are abundant in the fossil record. Their distribution span over all the history of life on Earth, the oldest fossil identified so far being also the oldest indisputable trace of life (3.5 Ga old). Contemporary microbialites can thus be used as models to study the early evolution of life, even though the earliest microbialites derived only from prokaryotic activity, their eukaryotic counter parts having only emerged some two billion years ago. Besides, the close interaction (both spatial and metabolic) between diverse microbial groups represents a favourable environment for genetic exchange and subsequent evolutionary innovations.

The alkaline (pH~9) crater lake Alchichica, located at 2,300 m a.s.l on the Mexican Central Plateau, harbours conspicuous microbialites. Preliminary microbial diversity surveys by PCR amplification, cloning and Sanger sequencing of 16S/18S rRNA genes revealed the presence of complex microbial communities composed of diverse bacterial and eukaryotic lineages. We have applied a metagenomic approach to assess the phylogenetic composition and the metabolic potential of microbial communities associated to Alchichica microbialites. The objectives of our study were to (i) characterise the diversity of microorganisms from the three domains of life at the macroscale, (ii) to have access to the relative abundance of the various microbial groups and (iii) explore the gene content of the metagenomes to identify the dominant metabolic functions.

465A

Structuring genetic and taxonomic diversity in gut microbes of lizards affected by a quick dietary change.

Chloe VIGLIOTTI [1], Eric Bapteste[2], Michel Habib[3], Anthony Herrel[1], Philippe Lopez[2]
[1] MNHN UMR 7179, Paris, France, [2] UPMC UMR 7138, Paris, France, [3] LIAFA, Universite Paris Diderot - Paris 7 UMR 7089, Paris, France

Podarcis sicula is a species of insectivorous lizards, present in different countries, including Croatia. At the end of the sixties, Nevo & al. (1972) wanted to sudy the competition between *Podarcis sicula* and *Podarcis melisellensis* on islands. They introduced 10 *Podarcis sicula* from Pod Kopiste on Pod Mrcaru, and 10 *Podarcis melisellensis* from Pod Mrcaru, on Pod Kopiste. Scientists came back 35 years later, and observed that *Podarcis melisellensis* had disappeared, and *Podarcis sicula* on Pod Mrcaru had became omnivorous (80 % herbivorous). Many morphological changes correlated with this dietary shift, but changes at the level of the gut microbiome and microbiota were not investigated.
Here, we have sequenced 16 samples from gut of insectivorous and omnivorous lizards by Illumina Miseq. We obtained between 3,544,700 and 7,397,327 of paired ends reads (about 300bp) by sample.
We tested two hypotheses, about evolutions of genes content and taxonomic composition associated with the dietary shift: Can we notice the acquisition or loss of gene families ? Did taxa appear or disappear in the gut microbial populations of omnivorous lizards ?
We used different approaches (pca and clustered heatmap on present taxa, abundance analysis of taxa and gene content using reads annotation results, networks based on reads similarities...) to study the diversity of these lizards gut microbiomes and microbiota, and to find some structures correlated with their diet. Preliminary analyses suggest that abundances of taxa and functions changed, without functional or taxonomic gains or losses.

466B

Detection of composite genes in large similarity networks

Jananan Pathmanathan, Philippe Lopez, Eric Bapteste
UMR 7138 Evolution Paris Seine, Université Pierre et Marie Curie, Paris, France

Evolutionary combinatorial processes, such as fusion and recombination of DNA segments derived from (un)related gene families, are involved in the creation of composite genes. These saltatory mechanisms are a source of new genes, involved in adaptations and phenotypic changes in organisms. These processes have been well established in eukaryotic genomes but little is known about their impact on soil, marine or gut microbial communities. Moreover despite their high adaptive potential, where and how composite genes are created in the environment is poorly understood.
An increasing amount of molecular data with a considerable genetic diversity is now available from metagenomics projects, allowing to address these fundamental issues in uncultivable microorganisms.
Bioinformatics methods, like FusedTriplets 2.0 and MosaicFinder, are available to detect composite genes and families of composite genes, respectively, in sequence similarity networks, where each node represents a unique sequence and each edge represents a similarity between connected sequences. Here we present a new algorithm, which is more memory-efficient and faster, for the detection of composite genes and families of composite genes in very large similarity networks, e.g. with several millions of nodes and hundreds of millions edges. Furthermore, we also developed a user-friendly tool that computes a score for each family of composite genes allowing one to select the most conserved composite genes from metagenomic data.

467C

The cryptic origin of denitrification pathway in benthic foraminifera

Christian Woehle [1], Alexandra-Sophie Roy[1], Julia Weissenbach[1], Tanita Wein[1,2], Philip Rosenstiel[3], Joachim Schönfeld[2], Tal Dagan[1]

[1] Institute of Microbiology of Christian-Albrechts-University Kiel, Kiel, SH, Germany, [2] GEOMAR – Helmholtz-Zentrum für Ozeanforschung Kiel, Kiel, SH, Germany, [3] Institute of Clinical Molecular Biology of Christian-Albrechts-University Kiel, Kiel, SH, Germany

Foraminifera are a group of amoeboid protists commonly found in benthic habitats. Several species have been reported to perform complete denitrification, which is a rare metabolic capacity among eukaryotes. This property and their high density in certain marine sediments pinpoint them as important players in the oceanic nitrogen cycle. The mechanisms behind foraminiferal denitrification are still scarcely understood and studies corroborate a possible contribution of bacterial endosymbionts.

To characterize the genetic mechanisms associated with foraminiferal denitrification and their regulation, we sequenced the transcriptome of denitrifying *Globobulimina turgida* and additionally explored its endosymbiotic community structure using a 16S marker. The resulting quantitative transcriptome data confirm that under hypoxia conditions *G. turgida* uses nitrate driven respiration as an energy source rather than fermentative processes. The data furthermore revealed the presence of a nitrogen metabolism pathway. This includes an unusual putative nitrite reductase (*NIR*) gene that appears phylogenetically distinct from already known *NIR* genes. Additionally, we detected a nitric oxide dismutase (*NOD*) homolog whereas no nitric oxide reductase (*NOR*) was found. These findings suggest that in contrast to commonly known denitrifying species, the *G. turgida* denitrification pathway may not include an intermediary N_2O product. Our investigation of associated bacterial community revealed four dominant gamma-proteobacteria that are related to foraminifera-associated bacteria described in previous studies.

Our study thus yields new insights into the underlying mechanisms of foraminiferal denitrification. However, a clear distinction between a eukaryotic or endosymbiotic origin of the nitrogen metabolism pathway remains challenging. Future efforts are aimed at solving this enigma.

468D

Environmental *Vibrio* represent a source of antibiotics that inhibit pathogenic *Vibrio cholera* .

Stephen Norris[1], Abigail Joy[1], Kathryn Kauffman[2], Martin Polz[2], Hans Wildschutte[1]
[1] Bowling Green State University, Bowling Green, OH, USA, [2] MIT, Cambridge, MA, USA

In coastal waters world-wide, diverse bacterial species belonging to the Vibrionaceae are both ubiquitous and persistent within the marine column. Through repeated sampling efforts of estuary waters, we routinely find non-pathogenic *Vibrio* associated with suspended particles of detritus, chitin, and gassy filaments. A likely trait contributing to such fitness effects is the ability to antagonize nearby competitors through production of antimicrobial factors, and the ability to resist their affects. In an environment where multiple groups of closely related *Vibrio* are in competition, we predict such antagonistic interactions lead to a social structure in which genotypic groups of Vibrio are resistant to specific factors produced by select members of the same group, resulting in a fierce out-competition of some strains including pathogenic isolates. We hypothesize that in the absence of outbreaks, dynamic interactions among proximal vibrios result in decreased the fitness of human pathogenic strains. Through a pair-wise analysis involving 100 *Vibrio* strains we were able to show that some isolates, we termed ‘super-killers' (SKs), inhibited the growth of many other strains suggesting that SKs produce potent inhibitory compounds. Furthermore, numerous environmental derived strains displayed inhibitory activity against clinical pathogens of *Vibrio cholerae* and *Vibrio parahaemolyticus*, and we are in the process of identifying the toxigenic compounds. As a results of their inhibitory actions, these environmental isolates represent natural sources of novel drug discovery against human pathogens.

469A

Oral bacteria in historical human remains from Finland -a metagenomic study

Kerttu Majander [1], Alexander Herbig[2], Verena Schuenemann[1], Kati Salo[3], Florian Aldehoff[1], Daniel Huson[4], Päivi Onkamo[5], Johannes Krause[2,1]

[1] University of Tübingen, Institute for Archaeological Sciences, Tübingen, Baden-Württemberg, Germany, [2] Max Planck Institute for the Science of Human History, Jena, Thüringen, Germany, [3] University of Helsinki, Department of Anthropology, Helsinki, Finland, [4] University of Tübingen, Department of Computer Science, Tübingen, Baden-Württemberg, Germany, [5] University of Helsinki, Department of Biological and Environmental Sciences, Helsinki, Finland

Microbial communities undergo a continued co-evolution with their host. Whereas pathogens are in a constant arms-race with the human immune system, commensal bacteria are adapting to changes in their environment, such as nutrition, and are often essential for the normal function of their host. Some commensal bacteria can, however, cause infections for a host with a weakened immune system. The oral cavity of humans hosts several such bacterial species, which in lack of dental care may cause inflammation.

Palaeogenetic research can help to directly study changes in bacterial species and communities that are involved in human diseases throughout time. The quality of ancient DNA is however compromised by degradation and contamination from external sources. Tooth bone and dental pulp, enclosed in hard enamel, are relatively protected against these type of effects.

Here we present a metagenomic study for human dental remains from Finnish archaeological collections, that carry evidence for various inflammations and dental diseases. We use a high-throughput shotgun DNA sequencing approach in order to study the composition of oral bacterial species found in teeth from individuals diagnosed with dental diseases. The analyses, based on alignments against bacterial genome databases, identified pathogenic and opportunistic bacterial taxa that are consistent with the preliminary osteological diagnosis.

Paleogenetic studies may thus help to confirm observed dental pathologies and contribute to our understanding of past population's vulnerability to certain diseases. Metagenomic studies also advance our understanding of temporal changes in the human microbiome, and give insights in to host-pathogen co-evolution.

470B

High-throughput pipeline for the identification of micro evolution of E. coli during an infection from whole genome sequencing

Adrien LAUNAY [1,2], Alexandre BLEIBTREU[1], Olivier TENAILLON[1]
[1] inserm, Paris, France, [2] UPMC, Paris, France

Being the cause of intra and extra intestinal infection responsible for more than 1 million death per year, *E. coli* is not just a model organism, it is also a public health concern. It is therefore important to understanding the evolutionary forces driving the diversification of this species. We know that the *E. coli* species shows an important variability with less than half of the genes observed in a strain being part of the ubiquitous core genome. This characteristic can prove to be problematic when trying to assemble a newly sequenced genome by comparing it to a reference as we cannot guess what *E. coli* genes will be present. Here we used a newly developed pipeline combining de novo assembly with reference mapping to a database of 128 fully sequenced and annotated E. coli genomes. We show that this approach can be used to reconstruct rearrange and annotate newly sequenced *E. coli*, allowing for the identification of micro evolution during an *E. coli* infection. We sequenced 8 coli clones from a patient suffering from peritonitis. We were able to identify 232 different mutation among all strains, identifying a mutator strain, and 8 genes that were mutated independently in the different samples, indicating evolutionary convergence.

471C

CRISPR-based Herd Immunity In Spatially Structured Bacterial Populations

Pavel Payne, Jonathan Bollback
Institute of Science and Technology Austria, Klosterneuburg, Austria

One of the major challenges for bacteria is the ubiquitous threat from bacteriophages in their environments. In order to survive this battle, bacteria have evolved multiple defense systems such as restriction-modification systems, abortive infection systems, BREX, and the Cas/CRISPR system. In this study, we explored the dynamics of bacteriophage spread in spatially structured populations consisting of bacteriophage-susceptible and bacteriophage-resistant individuals. Our system consists of *Escherichia coli* cells with CRISPR-based immunity to T7 bacteriophage. We show that even a relatively small fraction (~10-20%) of the resistant cells in a population can effectively halt the spread of a bacteriophage epidemic, i.e. a herd immunity effect. However, during these epidemics resistant bacteria are also being killed by bacteriophage, we therefore examined the penetrance of the immune system. We show that there is a critical threshold, of approximately seven bacteriophages per cell, that once exceeded, even the resistant cells become susceptible to lysis. Therefore, the lysis of susceptible cells locally increases the multiplicity of infection above the critical threshold, resulting in lysis of both susceptible and resistant neighboring bacteria and therefore failing to halt the spread of the epidemic.

472D

Seasonal variation in bacterial communities associated with Drosophila melanogaster

Fabian Staubach [1], Dmitri Petrov[2], NESCent consortium Tracking the biotic response to global climate cha[3]

[1] Freiburg University, Freiburg, Germany, [2] Stanford University, Stanford, CA, USA, [3] NESCent, Durham, NC, USA

There is recent evidence that D. melanogaster is subject to cycles of seasonal natural selection, but the direct sources of seasonal selection pressures are largely unknown. Microbes are known to exert strong selection pressures on metazoans. D. melanogaster lives on rotting fruit, an environment that is rich in microbes. Because climatic changes between seasons might favor the growth of different microbes in the environment of fruit flies, they could be an important source of seasonal selection pressure. If selection pressure exerted by microbes played a role in the observed patterns of seasonal adaptation, we would expect to find different microbes associated with D. melanogaster between seasons. In order to find out if fruit flies are indeed associated with different microbes through the seasons, we analyzed bacterial communities of D. melanogaster collected in spring and fall across different locations.

473A

Finding novel needles in a haystack: reconstructing genomes from metagenomes

Lionel Guy [1,2], Anja Spang[1], Jimmy H. Saw[1], Katarzyna Zaremba-Niedzwiedzka[1], Thijs J.G. Ettema[1]

[1] Department of Cell- and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden, [2] Department of Medical Biochemistry and Microbiogy, Uppsala University, Uppsala, Sweden

Exploration of the prokaryotic dark matter, i.e. the 99% of bacteria and archaea that cannot be grown on plates, has started to transform our view of the early evolution of living organisms. Although metagenomics and single-cell genomics have allowed us to study this hidden majority of prokaryotes, extracting complete genomes is still highly challenging.

A recent metagenomic survey of sediments sampled close to Loki's Castle's hydrothermal vents revealed that up to 10% of the biodiversity belonged to the Deep Sea Archaeal Group (DSAG). DSAG is a deep-branching group of the TACK superphylum, from which the putative archaeal ancestor of Eukaryotes is thought to originate.

To extract the complete DSAG genome from this metagenome, we developed an innovative approach to use supervised binning of metagenomic contigs without available reference genomes. To constitute training sets, 59 robust taxonomic markers were selected, their homologs identified in the metagenome, and single-gene trees of these complemented with a hundred reference sequences were inferred. Trees were visually inspected, paying a special attention to the placement of markers coming from the metagenome. Supervised binning could be performed using as reference the contigs on which the markers were located. Reassembly of the reads from one bin (Lokiarchaeum) yielded a near-complete (92%) composite genome. Phylogenomic analyses involving the 59 markers mentioned above showed that the newly defined Lokiarchaeota phylum is the closest archaeal relatives to eukaryotes.

The procedure here will be developed and automated. Its application to other datasets will allow reconstructing more novel genomes from uncultivable species.

474B

The quest for a unified view of bacterial land colonization

Hao Wu [1], Yongjun Fang[2], Jun Yu[1], Zhang Zhang[1]

[1] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, [2] Key Lab of Rubber Biology, Ministry of Agriculture and Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, China

Exploring molecular mechanisms underlying bacterial water-to-land transition represents a critical start toward a better understanding of the functioning and stability of the terrestrial ecosystems. Here, we perform comprehensive analyses based on a large variety of bacteria by integrating taxonomic, phylogenetic and metagenomic data, in the quest for a unified view that elucidates genomic, evolutionary and ecological dynamics of the marine progenitors in adapting to nonaquatic environments. We hypothesize that bacterial land colonization is dominated by a single-gene sweep, that is, the emergence of dnaE2 derived from an early duplication event of the primordial dnaE, followed by a series of niche-specific genomic adaptations, including GC content increase, intensive horizontal gene transfer and constant genome expansion. In addition, early bacterial radiation may be stimulated by an explosion of land-borne hosts (for example, plants and animals) after initial land colonization events.

475C

The Effect of Varying Concentrations of Select Carbon sources on Biofilm Growth and Adhesion Patterns of *Staphylococcus aureus* 305A Static Biofilms

Cherysa Anselm [1], Sarah Sutrina[1]
[1] *University of the West Indies, St. Michael, Barbados*

Biofilms were grown in glass assay tubes, in tryptic soy broth (TSB), LB, and a modified LB broth (LBM) supplemented with potassium dihydrogen phosphate (KH2PO4) media, and quantified via crystal violet staining. *S. aureus* 305A formed weak biofilm rings at the gas/liquid interface when grown in LB despite supplementation with various carbon sources. However, stimulated biofilm growth and two distinct adhesion patterns were observed when cells were grown in supplemented LBM. We hypothesized that due to the stimulation of biofilm growth observed the biofilm would spread downward from an initial ring formed at the gas/liquid interface to cover the tube surface. However, biofilm staining patterns monitored over hourly time points indicated that initial adhesion occurred at the bottom of the tube, and spread upward over the surface toward the gas/liquid interface as biofilm density increased.

5 Open Symposium

5.1

Genomics of adaptation and species cohesion in ecologically divergent forest trees (Populus spp.)

Margot Paris [1], Kai N. Stölting[1], Cecile Meier[1], Berthold Heinze[2], Stefano Castiglione[3], Denes Bartha[4], David Macaya Sanz[5], Santiago Gonzalez-Martinez[5], Christian Lexer[1]
[1] University of Fribourg, Fribourg, Switzerland, [2] Austrian Federal Research Centre for Forests (BFW), Vienna, Austria, [3] University of Salerno, Salerno, Italy, [4] West-Hungarian University, Sopron, Hungary, [5] Center of Forest Research, CIFOR-INIA, Madrid, Spain

Understanding the mechanisms underlying adaptation, and species cohesion is central to population and conservation genomics, especially in taxa that function as foundation species for entire communities of animals and plants. In Eurasia, several phenotypically differentiated forms within the tree genus *Populus* have diverged in the face of gene flow. We explore the determinants of divergence and cohesion along this continuum, ranging from differentiated populations and phylogeographic lineages to divergent species with fairly strong postzygotic barriers. We report a whole genome sequencing effort involving multiple populations of Eurasian *Populus alba* and *P. tremula* including multiple post-glacial recolonization lineages within the former. *P. alba* is an ecologically important model tree species with a very wide Eurasian distribution range, large population sizes and can adapt to drought and/or elevated salinity. We find a complex genomic architecture of adaptive differentiation from both, standing genetic variation and new mutations. In Central European *P. alba* alone, screening of >13 million sequence polymorphisms reveals highly differentiated outlier markers in >430 predicted genes mostly involved in soil adaptation. Genomic cohesion in hybridizing species appears to be maintained by widespread coupling effects, while around 4% of the studied genome windows do not exhibit any fixed between-species polymorphism. Some of these regions permeable to introgression also showed intraspecific signature of selection, suggesting that they could be promising candidates for introgressive adaptation in *Populus*.

5.2

Comparative Analyses of Four Snake Genomes Provide Important Insights Into Trajectories of Snake Evolution

Qi Zhou [1], Zongji Wang[2], Wei Ying[4], Guojie Zhang[2,3], Guangmei Yan[4]
[1] University of California,Berkeley, Berkeley, USA, [2] Beijing Genomic Institute, Shenzhan, China, [3] University of Copenhagen, Copenhagen, Denmark, [4] Zhongshan School of Medicine, Guangzhou, China

Snakes have evolved a limbless body plan, female heterogametic sex chromosomes and diverse venom systems, of which the underlying genomic changes from their reptile ancestor are unclear. We present here comparative analyses of genome and transcriptomes of *Deinagkistrodon acutus* (the 'hundred pacer' viper), to genomes of three other species that together span the entire snake phylogenetic range. We identify venom or olfactory gene families that have undergone adaptive sequence evolution or family expansion. Only venom genes specific to viper show a biased expression in *D. acutus* venom glands, implying their functional specialization. On the other hand, many genes (e.g., *Hox*) involved in organ development (e.g., eye and lung) show sequence signatures of relaxed selective constraints or gene family contraction. We date such genes underlying forelimb development to the ancestor of snakes, but find different genes underlying hindlimb development in different snake species. This reflects that hindlimb loss has occurred independently after forelimb loss occurred in the common ancestor of snakes. Finally, we show snake sex chromosomes have undergone at least two punctuated times of recombination suppression, with repetitive elements preferentially enriched at the putative sex-determining regions. Overall, we use *D. acutus* as model and address various unique aspects of snake genome evolution regarding its genes, repetitive elements and sex chromosomes.

5.3

From genome to function: Timing adaptations in the intertidal insect *Clunio marinus*

Tobias S. Kaiser
Max F Perutz Laboratories, Vienna, Austria

The marine midge *Clunio marinus* (Diptera: Chironomidae) lives in the intertidal zone of the European Atlantic coast. It's development and adult emergence are precisely timed to the rhythm of the tides by circadian and circalunar clocks. As the pattern of the tides differs along the coastline, *C. marinus* populations show a variety of local genetic adaptations in circadian and circalunar timing. Population genetic analysis suggests that timing adaptations evolved within the last 20,000 years. QTL mapping indicates that the timing adaptations are controlled by few major effect loci.

In order to pinpoint the genes and evolutionary processes underlying timing adaptation, we sequenced, assembled and annotated a highly contiguous *C. marinus* reference genome (N50: 1.9 Mb). Genetic mapping lead to a full reconstruction of the three chromosomes and identified specific features of *C. marinus*' genome architecture, such as chromosome arms, large chromosomal re-arrangements, heterogeneous recombination rates and variable synteny to other dipterans.

Next we re-sequenced pools of 300 individuals from five populations of *C. marinus*, which differ in circadian and circalunar timing. Genome-wide we detected timing-associated genes based on genetic divergence and its correlation to timing differences. The distribution of genetic variation in the timing QTLs suggests that timing adaptation happened from standing genetic variation and primarily involves regulatory changes.

Subsequent molecular analysis substantiated that adaptation in circadian timing relies on modulating alternative splicing of a metabolic enzyme.

5.4

Genomic imprinting and its systematic perturbation in abortive interspecific tomato seeds

Ana Marcela Florez-Rueda[1,3], Margot Paris[1,3], Anja Schmidt[2,3], Alex Widmer[1,3], Ueli Grossniklaus[2,3], Thomas Städler [1,3]

[1] Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland, [2] Institute of Plant Biology, University of Zurich, Zurich, Switzerland, [3] Zurich-Basel Plant Science Center, Zurich, Switzerland

Hybrid seed failure represents an important postzygotic barrier to interbreeding among species of wild tomatoes (*Solanum* section *Lycopersicon*) and other angiosperm groups. We studied this phenomenon in the closely related *Solanum peruvianum* and *S. chilense*; hybrid crosses between these species yield very high proportions of inviable seeds due to endosperm failure and arrested embryo development. Based on seed size differences in reciprocal hybrid crosses and developmental evidence implicating endosperm failure, we hypothesized that (perturbed) genomic imprinting might be involved in this strong postzygotic barrier. Consequently, we surveyed endosperm transcriptomes obtained via laser-assisted microdissection of developing seeds representing both intra- and interspecific pollinations. We implemented a novel approach to estimate parent-of-origin–specific expression using both homozygous and heterozygous nucleotide differences between the two parents and identified hundreds of candidate imprinted genes. Importantly, we uncovered systematic shifts of the 'normal' (intraspecific) maternal:paternal transcript proportions in hybrid endosperms; the average maternal proportion of gene expression increases in both directions of the hybrid cross but is strongly negatively correlated with 'normal' maternal proportions. This genome-wide shift almost entirely eliminates imprinted paternal gene expression in hybrid endosperms but also affects maternally expressed imprinted genes (which on average shift to lower maternal proportions) and all other genes. In addition to these profound changes in parental expression proportions in hybrid endosperms, we found expression-level changes in methyltransferases, histones, MADS-box genes and other transcription factors. These results resemble known perturbations of transcriptional regulation in *Arabidopsis*, suggesting a likely role for small interfering RNAs in hybrid seed failure.

5.5

Joint estimation of contamination, sequencing error and demography for nuclear DNA from ancient humans

Fernando Racimo [1], Gabriel Renaud[2], Montgomery Slatkin[1]
[1] University of California, Berkeley, Berkeley, CA, USA, [2] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

When sequencing an ancient DNA sample from a hominin fossil, DNA from modern humans involved in excavation and extraction will be sequenced along with the endogenous material. This type of contamination is problematic for downstream analyses as it will introduce a bias towards the population to which the contaminating individuals belong. Quantifying the extent of contamination is a crucial step as it allows researchers to assess the validity of conclusions based on downstream population genetics analysis. Here, we present an algorithm to co-estimate the contamination rate, sequencing error rate and demographic parameters - including drift times and admixture rates - for an ancient nuclear genome obtained from a diploid individual, when the putative contaminating DNA comes from a population related to the sample. We assume we have a large panel representing the putative contaminating population (e.g. European, Asian or African). The method is implemented in C++, and can also be used to determine the most likely population to which the contaminant DNA belongs. We have applied it to simulations and Neanderthal genome data, and we recover accurate estimates of all parameters, even when sequencing coverage is low and contamination is high.

5.6

Human adaptation to life in the high arctic

Ida Moltke [1], Matteo Fumagalli[2], Niels Grarup[1], Fernando Racimo[3], Peter Bjerregaard[4], Marit E Jørgensen[5], Thorfinn S Korneliussen[1], Pascale Gerbault[2], Emilia Huerta-Sanchez[6], Oluf Pedersen[1], Torben Hansen[1], Anders Albrechtsen[1], Rasmus Nielsen[3,1]

[1] University of Copenhagen, Copenhagen, Denmark, [2] University College London, London, UK, [3] UC Berkeley, Berkeley, CA, USA, [4] University of Southern Denmark, Copenhagen, Denmark, [5] Steno Diabetes Center, Gentofte, Denmark, [6] UC Merced, Merced, CA, USA

The native people of Greenland, the Inuit, are of special interest to anthropologists and evolutionary biologists because of their extreme adaptations to life in the high arctic. In particular, Inuit traditionally have a diet based primarily on hunting marine mammals, which have a high content of polyunsaturated fatty acids. Inuit provide an opportunity to investigate genetic adaptations to such a diet. We analyze SNP chip data and exome sequencing data from 191 and 18 Inuit, respectively, in order to identify regions targeted by positive selection specifically in Inuit related to metabolic processes. We find extreme selection signals in several loci and using association mapping, we show that alleles associated with positive selection in these loci are strongly associated with multiple phenotypes, including weight.

This study illustrates the use of small understudied populations for understanding the genetic basis of human biological variation.

5.7

Strong *cis* -regulation of allele-specific DNA methylation in jewel wasp reciprocal F1 hybrids

Xu WANG [1], John Werren[2], Andrew Clark[1]

[1] *Cornell University, Ithaca, NY, USA,* [2] *University of Rochester, Rochester, NY, USA*

Parasitoid wasps of the genus *Nasonia* are emerging as a model for DNA methylation in insects. To study evolutionary changes in DNA methylation during species divergence, we scored allele-specific methylation (ASM) genome wide in both parental and reciprocal F1s of *N. vitripennis* (NV) and *N. giraulti* (NG). We found 4364 differentially methylated CpGs (DMCpGs) with >8x coverage for each parental allele in reciprocal F1s, and they clustered in 150 genes (DM-genes). If methylation changes were due to changes in *cis*-regulatory sequences, F1 ASM will resemble the parental status. If the methylation changes were due to *trans* factors, F1 ASM will be ~50% for both alleles. Surprisingly, in F1s, all 150 DM-genes are differentially methylated between the two alleles, indicating that they "remember" the parental methylation status with nearly 100% fidelity, consistent with strong *cis*-elements driving DNA methylation. We confirmed the absence of *trans*-effects by quantifying ASM using PyroMark assays in introgression lines that had only the DMCpG region coming from NG in an otherwise NV background. The largest DM-gene cluster is in a pericentromeric region, and a homozygous introgression line containing this region is sterile, suggesting a potential epigenetic component in speciation. In addition, we discovered that allele-specific expression is positively correlated with ASM in a subset of the 150 DM-genes. Unlike mammals, there is no parent-of-origin methylation or epigenetic reprogramming of global de-methylation and re-methylation in *Nasonia*. These results will shed light on the mechanism and evolution of DNA methylation as well as epigenetic changes in speciation.

5.8

A small impact of generation times on mutation rates in apes

Carlos Eduardo G. Amorim [1], Priya Moorjani[1], Ziyue Gao[2], Minyoung Wyman[1], Molly Przeworski[1]

[1] Columbia University, New York, NY, USA, [2] University of Chicago, Chicago, IL, USA

Mammals with shorter generation time tend to have higher substitution rates. This association has been extensively studied and is widely believed to reflect an increased number of cell divisions per unit time in species with shorter generation times. In humans, the possibility of a generation time effect has recently been the subject of considerable interest because of the finding that contemporary rates measured in pedigrees are two-fold lower than those estimated from phylogenetic relationships. This observation raises the possibility that yearly mutation rates have decreased during the course of human evolution with the lengthening of the generation time, and calls into question the constancy of the molecular clock. Here, we revisit the relationship between generation time and mutation rates in primates using whole-genome multi-species alignments. We confirm earlier findings of a "Hominoid rate slowdown", where Great Apes, whose generation time is about twice that of Old World Monkeys, have ~25% lower substitution rates. This generation time effect is also observed in New World Monkeys and Lemurs. Among the great apes, however, and considering consistently collected data, it is possible to observe only a minor effect. Moreover, as we show, these findings are consistent with what is expected from simple models of generation time effect. Important implications are that changes in the mean age of reproduction alone are likely to have had only subtle effects on the yearly mutation rate in primates and are unlikely to explain the difference between phylogenetic and pedigree based estimates of human mutation rates.

5.9

Human coding RNA editing is generally nonadaptive

Jianzhi Zhang
University of Michigan, Ann Arbor, Michigan, USA

Impairment of RNA editing at a handful of coding sites causes severe disorders, prompting the view that coding RNA editing is highly advantageous. Recent genomic studies expanded the list of human coding RNA editing sites by over 100 times, raising the question of how common advantageous RNA editing is. Analyzing 1783 human coding A-to-G editing sites, we show that (i) both the frequency and level of RNA editing decrease as the importance of a site or gene increases, (ii) edited As are more likely than unedited As to be replaced with Gs but not with Ts or Cs during evolution, and (iii) among nonsynonymously edited As, those that are evolutionarily least conserved exhibit the highest editing levels. These and other observations reveal the overall nonadaptive nature of coding RNA editing, despite the presence of a few sites where editing is clearly beneficial. We propose that most observed coding RNA editing results from tolerable promiscuous targeting by RNA editing enzymes whose original physiological functions still remain elusive. We also devised a method to identify the small fraction of potentially advantageous sites from the sea of largely nearly neural editing sites.

5.10

A genomic study of the contribution of DNA methylation to regulatory evolution in primates

Julien Roux [1,2], Hernando-Herraez Irene[3], Nick Banovich[2], Raquel Garcia Perez[3], Claudia Chavarria[2], Amy Mitrano[2], Jonathan Pritchard[4,5], Tomas Marques-Bonet[3], Yoav Gilad[2]
[1] University of Lausanne, Lausanne, Switzerland, [2] University of Chicago, Chicago, USA, [3] Institute of Evolutionary Biology (UPF-CSIC), PRBB, Barcelona, Spain, [4] Howard Hughes Medical Institute, Stanford University, Stanford, USA, [5] Departments of Biology and Genetics, Stanford University, Stanford, USA

A long-standing hypothesis is that changes in gene regulation play an important role in adaptive evolution, notably in primates. Yet, in spite of the evidence accumulated in the past decade that regulatory changes contribute to many species-specific adaptations, we still know remarkably little about the mechanisms of regulatory evolution. In this study we focused on DNA methylation, an epigenetic mechanism whose contribution to the evolution of gene expression remains unclear.

To interrogate the methylation status of the vast majority of cytosines in the genome, we performed whole-genome bisulfite conversion followed by high-throughput sequencing across 4 tissues (heart, kidney, liver and lung) in 3 primate species (human, chimpanzee and macaque). In parallel, we collected gene expression profiles using RNA-seq from the same tissue samples, allowing us to perform a high resolution scan for genes and pathways whose regulation evolved under natural selection.

We integrated these datasets to characterize better the genome features whose methylation status leads to expression changes, and we developed a statistical model to quantify the proportion of variation in gene expression levels across tissues and species. We discovered that, in contrast to the confirmed negative association between gene expression and methylation changes across tissues, the correlation was greatly reduced across species. Our study questions the importance of epigenetic modifications as a mechanism causing regulatory changes and adaptations in primates.

5.11

Evolution at a single gene causes a difference in recombination rates between *Drosophila* species

Cara Brand [1], M. Victoria Cattani[1,2], Sarah Kingan[1], Lori Wright[1], Daven Presgraves[1]
[1] *University of Rochester, Rochester, NY, USA,* [2] *New York University, New York, NY, USA*

Meiotic recombination ensures the proper segregation of homologous chromosomes and increases the general efficacy of natural selection. Crossing over, however, presents the risk of ectopic exchange among repetitive sequences in the genome— such as transposons— generating deleterious duplications and deletions. The rate of crossing over may therefore evolve to balance the benefits of recombination against the risks of ectopic exchange. Crossover rates differ among *Drosophila* species, providing an opportunity to investigate the genetic basis and population genetic forces involved. We performed an evolutionary screen of genes with meiotic phenotypes in *Drosophla melanogaster*. We identified one gene, known to be essential for crossover formation, that has a striking history of recent and long-term recurrent positive selection. Using transgenic flies, we show that molecular evolution at this gene can account for the evolved difference in the rate of crossing over between *D. melanogaster* and its closely related species, *D. mauritiana*.

5.12

The fossilised birth-death process applied to the total-evidence approach for dating with fossils

Alexandra Gavryushkina[1], David Welch [1], Tracy Heath[2], Alexei Drummond[1], Tanja Stadler[3]
[1] University of Auckland, Auckland, New Zealand, [2] Iowa State University, Iowa, USA, [3] ETH, Zurich, Switzerland

Evolutionary studies greatly rely on phylogenetic trees. Although there have been developed a number of methods to infer phylogenies there are still problems that require attention. One such problem is the dating of species phylogenies using fossil evidence. A promising method is the `total-evidence' approach, where molecular and morphological data of extant and fossil species are jointly used to infer species divergence times and macroevolutionary parameters. However current implementations of the method need to be improved in several aspects. In a Bayesian framework, an important component is the tree prior model which describes the tree branching process. Previous attempts to apply the total-evidence approach have used tree prior models that do not account for the possibility of fossil samples to be direct ancestors of other samples, that is, other fossils or extant species. Recently, Heath et al and Gavryushkina et al applied the fossilized birth-death model that explicitly models the sampling process and naturally allows for sampled ancestors to estimate divergence times based on molecular data and fossil occurrence dates. Here we present a method to analyse morphological and molecular data in a unified Bayesian framework with models that account for sampled direct ancestors. We apply this method to extant penguins and their fossil ancestors

5.13

The relationship between *dN/dS* and scaled selection coefficients: elucidating the properties, limitations, and capabilities of codon-based models

Stephanie Spielman, Claus Wilke
University of Texas at Austin, Austin, TX, USA

Numerous computational methods exist to assess the mode and strength of natural selection in protein-coding sequences, yet how distinct approaches relate to one another remains largely unknown. Here, we derive a precise mathematical relationship between the focal parameters of two such complementary modeling frameworks: *dN/dS* models and mutation-selection (MutSel) models. While *dN/dS* models estimate the relative rate of nonsynonymous to synonymous changes, MutSel models infer scaled selection coefficients among all amino acid and/or codon changes, indicating the selective response to specific mutations. The formal link we establish between these models provides a uniquely rigorous platform from which we gain deeper insight into the behaviors, limitations, and capabilities of each framework. For example, we prove that, when synonymous changes are neutral, MutSel models correspond only to $dN/dS < 1$, meaning this model is only suitable under purifying selection. We additionally show that synonymous selection alone can produce $dN/dS > 1$, indicating that *dN/dS* cannot clearly distinguish between purifying selection on synonymous codons and positive selection on amino acids. Finally, we demonstrate that the widely-used Goldman-Yang-style *dN/dS* model parameterizations yield negatively-biased *dN/dS* estimates on realistic sequences, while Muse-Gaut-style models display substantially less bias. Strikingly, the models giving the least-biased and most-precise *dN/dS* estimates are never those with the best fit to the data, measured through both AIC and BIC scores. Thus, selecting models based on goodness-of-fit criteria can be highly misleading, particularly if the models considered do not precisely correspond to the underlying mechanism that generated the data.

5.14

Heterogeneous characteritics of Conserved noncoding sequences in Eukaryotes

Nilmini Hettiarachchi [1], Naruya Saitou[1,2]

[1] *Graduate university for advance studies, Kanagawa, Japan,* [2] *University of Tokyo, Tokyo, Japan*

Conserved noncoding sequences (CNSs) are enriched in regulatory sequence elements. We conducted a whole genome analysis on plant CNSs and identified them to be GC rich (Hettiarachchi et al. 2014). Babarinde and Saitou (2013) reported mammalian CNSs to be GC poor. This heterogeneity in GC content might be related to varying sequence features of regulatory elements in different lineages. Since animals and fungi are sister groups, in order to determine the evolutionary origin of low GC content of mammalian CNSs we investigated the features of fungi lineage common CNSs. This investigation was further extended to discover the sequence features of lineage common CNSs of invertebrates, non-mammalian vertebrates with the intension to answer varying regulatory features of different lineages. Currently we have identified that plant, fungi, invertebrate lineage CNSs are predominantly GC rich where as vertebrates are GC poor. We also found that this GC content feature is directly related to their location in the genome. High GC CNSs showed a tendency to be found in heterochromatin regions, whereas low GC CNSs shows a tendency to locate in open chromatin. The transition of high GC content of CNSs from the majority of mulicellular eukaryotes to low GC content in vertebrates and the structural architecture of CNSs with its function are some of the questions we intend to answer in the future.

5.15

Deep sequencing of natural and laboratory populations of Drosophila melanogaster reveals new insights into the spectrum of de novo deleterious mutations

Zoe June Assaf, Dmitri Petrov
Stanford University, Stanford, CA, USA

A substantial fraction of new mutations are deleterious, yet detecting them is difficult because purifying selection rapidly removes the vast majority of deleterious mutations from natural populations. One way to overcome this problem is via mutation accumulation experiments, for example in Drosophila melanogaster, which allow mutations to accrue in the genomes of laboratory populations. Yet again, our ability to characterize deleterious mutations is limited by the total number of events which occur during the experiment. However, the past several years has seen a growing body of literature reporting DNA sequence data from both wild and experimental populations. For example there is population-level sequence data for Drosophila reported by groups across multiple continents, which collectively represent >17,000X coverage of >5,000 strains. Sequencing depth like this allows detection of low frequency mutations, a class of sites enriched for new deleterious mutations that rarely rise to higher frequencies. Similarly, recent mutation accumulation studies, including new data reported here, collectively give >2,000 mutation events. Combined, these data sets represent a rich untapped source of information for characterizing the deleterious mutation spectrum. Here we leverage these data to sample mutations at low frequencies in natural populations, accounting for confounding factors like sequencing error, and compare our results with mutation accumulation experiments. We show that mutational biases present in the common class of polymorphisms, including statistics like Transition-to-Transversion ratio, Pn/Ps, and GC content, among others, are removed in the low-frequency variant class, and in fact approach the de novo spectrum as revealed by mutation accumulation experiments.

5.16

An Approximate Bayesian Computation approach to reconstruct demography from sequence data.

Flora Jay [1], Simon Boitard[2], Frédéric Austerlitz[1]
[1] UMR 7206 Eco-anthropologie et Ethnobiologie, Muséum National d'Histoire & CNRS & Université Paris Diderot, Paris, France, [2] UMR 1313 Génétique Animale et Biologie Intégrative, Institut National Agronomique & AgroParisTech, Jouy en Josas, France

Approximate Bayesian Computation (ABC) has proven to be useful for inferring demography from microsatellite or SNP data. Whole-genome data are expected to be extremely rich in information about past demography but, because simulations were, until recently, computationally too costly, ABC methods have not been thoroughly tested on such very long sequences. Dense polymorphism data contain extra information that is not available from unlinked site polymorphisms, and should improve the reconstruction of demographic history. We investigate how summary statistics computed from sequences, such as the lengths of haplotypes shared between individuals, or the decay of linkage disequilibrium with distance, can be combined with classical statistics (eg heterozygosity, Tajima D) and efficiently integrated to an ABC framework. We then quantify their influence on the inference of demographic parameters, particularly in the presence of expansions or bottlenecks in a single population. Furthermore, we describe how errors that are usually more frequent in sequence than SNP data impact the inference, and show that modeling the error process in the ABC framework increases accuracy. Finally, we apply our method to publicly available European and African complete genomes.

5.17

Physical and functional plastome reduction coincides with major shifts in substitution rates and relaxation of purifying selection in photosynthetic and nonphotosynthetic parasitic Orobanchaceae

Susann Wicke [1,2], Kai F. Mueller[2], Claude W. dePamphilis[3], Gerald M. Schneeweiss[1]
[1] University of Vienna, Vienna, Austria, [2] University of Muenster, Muenster, Germany, [3] Pennsylvania State University, State College, PA, USA

Parasitic plants, such as those of the broomrape family (Orobanchaceae), possess strongly reconfigured plastomes due to convergent losses of photosynthesis and housekeeping genes, making them excellent model systems for studying genome evolution under relaxed selective pressures. Phylogenomic and phylostatistical analyses of 20 complete plastomes of nonparasitic and parasitic Orobanchaceae reveal that the establishment of obligate parasitism triggers the relaxation of selective constraints. Following the loss of photosynthesis, functional genome reduction proceeds rapidly, accompanied by the accumulation of recombinogenic factors, which foster severe genomic reconfigurations. The eventual physical plastome reduction by deletion of non-essential regions is strongly influenced by their proximity to genes under selection and the co-occurrence with those in operons. This indicates complex constraints beyond gene function that determine the evolutionary survival time of plastid regions after the loss of selective pressures. By using a Random Effects Likelihood framework we show that relaxation of purifying selection occurs in several photosynthesis and housekeeping complexes along or after the transition to an obligate heterotrophic lifestyle. Maximum Likelihood models crossed with multivariate Brownian diffusion processes and parametric bootstrapping reveal a correlation between lifestyle changes and substitution rate heterogeneity in the Orobanchaceae. We show that bursts of gene loss and genome reconfiguration coincide with shifts in nucleotide substitution rates, most notably so in non-synonymous substitution rates. However, the vast majority of genes (incl. those for the photosynthetic ATP synthase) retained in nonphotosynthetic Orobanchaceae evolve under purifying selection, suggesting that those elements are still of some functional relevance for these parasites.

5.18

Enhancing Reproducibility in Bioinformatics for Microbiology

Paul Walsh [1], Brian Kelly[1], Michael Bekaert[1], Roy Sleator[1,2], John Carroll[1], Xiangwu Lu[1]
[1] NSilico Life Science, Cork, Ireland, [2] Cork Institute of Technology, Cork, Ireland

Reproducibility is one of the main principles of the scientific method and is defined as the ability of an entire study to be replicated, either by the researcher or by someone else working independently, so as to corroborate the original results. Achieving reproducibility in life science is particularly challenging as even the most basic entities are composed of highly dynamic and interactive complex systems. Issues of reproducibility in life science are now being recognised more widely in both the scientific community and popular press. Recently biotech company Amgen attempted to replicate the findings of 53 landmark articles published by reputable labs in top journals, but only 6 of the 53 studies were reproduced. The field of bioinformatics is not immune to this issue as many researchers are not proficient in computing and take a trial and error approach by downloading tools from the Internet and fail to track parameters used in published studies. We address this issue in the field of bioinformatics as applied to the domain of microbiology and address the concerns of usability, provenance, traceability and reproducibility. We demonstrate this by fully sequencing and analysing the genome of bacterial pathogens implicated in clinical cases of neonatal sepsis. We demonstrate how all bioinformatics analysis related to this clinical study is fully reproducible through the use of a novel cloud based framework.

5.19

Evolutionary and functional impact of polymorphic inversions in the human genome

Mario Cáceres [1,2], Sergi Villatoro[1], David Castellano[1], Magdalena Gayà-Vidal[1], Carla Giner-Delgado[1,3], Meritxell Oliva[1], David Izquierdo[1], David Vicente-Salvador[1], Lorena Pantano[1], Marta Puig[1]

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain, [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain, [3] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

For a long time *Drosophila* inversion polymorphism has been a big paradigm in evolutionary biology. However, due to the difficulty of their study, little is known about the role of inversions in other organisms. Here we present the final results of INVFEST, an ambitious project towards the complete characterization of polymorphic inversions in the human genome. First, we have determined the distribution of 45 inversions in 550 individuals from seven populations of the 1000 Genomes Project, which represents the largest population genetics study of human inversions so far. Inversion frequency spectrum showed considerable variation (MAF=0.5-49.7%), with a bias towards intermediate frequencies and significant differences among populations (Fst=0.01-0.49) in several cases. In particular, different tests indicated that distribution patterns of some inversions are not consistent with a neutral scenario and suggest events of positive or balancing selection. Second, the analysis of the nucleotide variation within the inverted region revealed that inversions mediated by inverted repeats (N=26) show an unexpectedly high degree of recurrence, with most of them occurring on different haplotypes in humans and showing also different orientations in chimpanzees and gorillas. This contrasts with inversions with simple breakpoints (N=19), which are unique and can be tagged by SNPs. Finally, we have identified different functional effects of the inversions, ranging from gene breakage, inversion of alternatively spliced exons, and generation of new fusion transcripts. Our integrative analysis therefore illustrates the dynamic nature of the genome and represents a key step in defining the evolutionary impact of this type of structural variants.

5.20

Energy efficiency trade-offs drive nucleotide usage in transcribed regions

Wei-Hua Chen[2], Peer Bork[3], Martin Lercher [1]

[1] Heinrich Heine University, Düsseldorf, Germany, [2] University of Geneva, Geneva, Switzerland, [3] EMBL, Heidelberg, Germany

Efficient nutrient usage is a trait under universal selection. A substantial part of cellular resources is spent on making nucleotides. We thus expect preferential use of cheaper nucleotides especially in transcribed sequences, which are often amplified several thousand-fold compared to genomic sequences. To test this hypothesis, we derived a mutation-selection equilibrium model for nucleotide skews (strand-specific usage of *A* versus *T*, and *G* versus *C*). The model explains the distribution of site type-specific nucleotide skews across 1,550 prokaryotic genomes as a consequence of ubiquitous selection on efficient resource usage. Transcription-related selection favors the cheaper RNA nucleotides *U* and *C* at synonymous sites in the vast majority of genomes. However, the information encoded in mRNA is further amplified through translation, which adds another level of selection on efficient nutrient usage. Due to unexpected trade-offs in the codon table, cheaper nucleotides encode on average energetically more expensive amino acids; these trade-offs apply both to strand-specific nucleotide usage (*AT* and *GC* skews) and to *GC* content. The trade-offs cause a universal bias towards the more expensive nucleotides *A* and *G* at non-synonymous coding sites, and are consistent with selection for reduced investment into nucleotides in low-GC genomes.

5.21

Prevalence of episodic positive selection in immune genes of ants

Lumi Viljakainen [1], Heli Havukainen[2], Iiro Salminen[1], Seth Barribeau[3], Yannick Wurm[4], Hua Yan[5], Sanne Nygaard[6], Morten Shiott[6], Eyal Privman[7]
[1] University of Oulu, Oulu, Finland, [2] University of Helsinki, Helsinki, Finland, [3] East Carolina University, Greenville, NC, USA, [4] Queen Mary University London, London, UK, [5] Howard Hughes Medical Institute, New York, NY, USA, [6] University of Copenhagen, Copenhagen, Denmark, [7] University of Haifa, Haifa, Israel

How social interactions affect pathogen evolution and evolution of host immunity is unclear. The social insects represent ideal study systems to contrast with solitary animals. However there are two competing predictions. On one hand, large colony size and frequent interactions between colony members, should lead to rapid pathogen evolution and frequent immune challenges. This would lead to extensive positive selection of immune genes. On the other hand, social insects have evolved so called 'social immunity' that lowers the risk of infection, which could result in lesser importance for molecular immunity and relaxed purifying selection on immune genes. Both of these alternative hypotheses have gained support in taxonomically restricted studies on small numbers of genes. Here, we investigated patterns of evolution in more than 80 innate immune genes of seven ant species. We focused on taxonomically conserved genes involved in innate immunity to answer the following questions: i) are immune genes of ants common targets of positive selection, as found for the homologous genes in Drosophila? ii) are the immune genes involved in an ongoing evolutionary arms race and continuous positive selection, or is their evolution characterized by episodic lineage-specific selection? We addressed these questions by investigating evolutionary rates in immune genes using codon-based models of sequence evolution and related the results to functional domains on the protein structures of the positively selected genes. We found frequent episodic positive selection in immune genes of ants - the selection has affected different sites on different branches indicating short-term associations with various pathogens.

5.22

Local DNA topography predicts genomic mutation rates

Premal Shah, Joshua Plotkin
University of Pennsylvania, Philadelphia, PA, USA

Random mutations are the primary fuel for evolutionary processes. Recent high-throughput mutation accumulation experiments have shed light on both the rate and spectrum of biases in mutation across several species ranging from bacteria to drosophila; and yet we still do not understand the underlying molecular mechanisms for the observed rates and biases.

Current approaches to understand mutation biases in genomes have focussed largely on comparative analyses. By contrast, we propose a mechanistic explanation — that the local topography of DNA structure influenced by its nucleotide composition affects binding affinities of proteins involved in DNA methylation and repair, and thereby patterns of mutation biases. We quantify openness in DNA structure of a given sequence based on both hydroxyl-cleavage experiments and structural information from NMR images. By testing this hypothesis in several organisms – from bacteria to fruit fly – we show that openness of DNA structure can explain up to 70% of variation in random mutations in mutation accumulation experiments. Our ability to predict regions of genomes that are subject to differential mutation rates raises several exciting questions about the predictability of evolution, and the evolution of genomic nucleotide composition.

5.23

Co-evolution of sexy proteins: characterizing the structural interactions of egg and sperm fertilization proteins by NMR

Damien Wilburn, Willie Swanson
University of Washington, Seattle, WA, USA

Pervasive co-evolution between sperm and egg fertilization proteins is a common theme among sexually reproducing organisms to yield high affinity conspecific interactions. Particularly for external fertilizers, a secondary evolutionary challenge may exist to restrict unfavorable heterospecific crosses. This challenge can be addressed with changes in timing or location of reproduction, or through molecular refinement of gamete recognition proteins. For the marine gastropod abalone, seven sympatric species with similar breeding seasons live off the coast of California, yet hybrids are rarely observed. During fertilization, male abalone sperm secrete a 16 kDa acrosomal protein, lysin, which specifically binds to repeat domains in the egg coat protein VERL. Lysin-VERL interactions are species specific, and molecular evolutionary studies demonstrate strong signatures of positive selection and rapid co-evolution between the two proteins. However, lysin has acquired approximately five times as many non-synonymous substitutions as VERL, and the molecular mechanism of how these mutational effects contribute to high-affinity, species specific interactions remains unclear. Using multidimensional NMR, we determined the 3D structures of lysin and VERL from red abalone (*Haliotis rufescens*). NMR perturbation analyses and molecular docking studies were used to characterize the lysin-VERL interface, with molecular dynamics assayed by NMR relaxation analysis. Homology modeling of lysin and VERL from additional abalone species provided an evolutionary framework to help understand the forces driving high affinity conspecific interactions.

5.25

Spore killer and its consequences on *Neurospora* populations

Ruxandra I. Molnar, Hanna Johannesson
Uppsala University, Uppsala, Sweden

Conflicts caused by selfish genetic elements (SGE) are a driving force for evolutionary innovation, and hence, of fundamental importance for all aspects of biology. In many ascomycete fungi, meiotic drive, one form of SGE, is observed as spore killing. This phenomenon is evident in crosses between killer and non-killer strains, in which case half of the spores produced die and all offspring are killers. However, in contrast to the mammalian system, this apparent fitness cost may limit the impact of the meiotic drive element on population demography (i.e. the relative proportions of killer, non-killer and resistant alleles). In *Neurospora*, the *Spore killer* element and its resistance factor have been mapped to an approximately 2 Mbp-large region of suppressed recombination on LG3. Here, we use *N. intermedia* to study the costs and benefits of killer relative to non-killer and resistant strains. We have introgressed four killer elements and two resistance alleles into two different genetic backgrounds to study the phenotypic consequences of carrying the element and the influence of the genetic background on killers' performance. This empirical work will expand upon existing models used to explain the population dynamics of this SGE in *Neurospora.*

518A

Evolution of two prototypic T cell lineages: TCRα/TCRδ locus signature in the genomic organization of the adjacent VLRA and VLRC loci in jawless vertebrates

Sabyasachi Das [1], Jianxu Li[1], Stephen Holland[2], Lakshminarayan Iyer[3], Masayuki Hirano[1], Michael Schorpp[2], L Aravind[3], Thomas Boehm[2], Max Cooper[1]

[1] Emory University, Atlanta, Georgia, USA, [2] Max-Planck Institute of Immunobiology and Epigenetics, Stuebeweg, Freiburg, Germany, [3] National Center for Biotechnology Information, Bethesda, Maryland, USA

Lampreys and hagfish, the only extant jawless vertebrates, are a focus of interest in the search for evolutionary origin of adaptive immunity, because of their pivotal position in chordate phylogeny. Whereas jawed vertebrates use Ig-domain based T-cell receptors (TCRs) and B-cell receptors (BCRs) to recognize antigens, the jawless vertebrates employ leucine-rich repeat (LRR)-based variable lymphocyte receptors (VLR) for this purpose. During the assembly of VLR genes, several types of donor LRR genomic cassettes are copied in a piecewise and step-wise fashion into incomplete germ-line genes to generate the mature forms of antigen receptor genes. Both lampreys and hagfish have three types of VLR-bearing lymphocytes: VLRA- and VLRC-producing cells share many similarities with the two principal T-cell lineages of jawed vertebrates that express the αβ or γδ TCRs, whereas VLRB-producing cells are B-cell like. Despite the mutually exclusive expression of VLRA and VLRC, some donor LRR cassettes are shared during assembly. The genomic structure of the adjacent VLRA and VLRC loci in lampreys is reminiscent of the interspersed nature of the TCRα/TCRδ locus in jawed vertebrates, an organization which facilitates the sharing of some variable gene segments during the recombinatorial assembly of TCRα and TCRδ genes. Comparative analysis of the VLRA/VLRC loci using sea lamprey and japanese lamprey genome sequences suggests that the evolutionary dynamics of repertoire development in VLRA/VLRC loci is similar to that of TCRα/TCRδ locus in jawed vertebrates. Our findings offer new insight into the evolution of anticipatory receptors for adaptive immunity in vertebrates.

519B

Anthropogenic secondary-contact of cryptic penguin species

Stefanie Grosser [1], Christopher Burridge[2], Amanda Peucker[3], Nicolas Rawlence[1], Paul Scofield[4], Ian Smith[5], Jonathan Waters[1]

[1] Allan Wilson Centre for Molecular Ecology and Evolution, Department of Zoology, University of Otago, Dunedin, New Zealand, [2] School of Biological Sciences, University of Tasmania, Hobart, Australia, [3] School of Life and Environmental Sciences, Warrnambool, Australia, [4] Canterbury Museum, Christchurch, New Zealand, [5] Department of Anthropology and Archaeology, University of Otago, Dunedin, New Zealand

The use of genetic and genomic tools has facilitated the discovery of new species that, due to morphological similarities, were previously indistinguishable using traditional approaches. These cryptic species complexes have even been found within some conspicuous and well-studied vertebrate taxa and it is has been suggested that many well-known and widespread species could harbour substantial unrecognised species diversity. Here we present genetic evidence for two cryptic, reproductively isolated species within the world's smallest penguin. We used mitochondrial, microsatellite and intron markers to reveal two highly divergent *Eudyptula* lineages, one Australian and one New Zealand lineage, which co-occur in southern New Zealand, with minimal hybridisation. Coalescent modelling, using fast molecular rates, indicates that the Australian lineage likely expanded into southern New Zealand within recent centuries, supporting the hypothesis of an anthropogenic extinction-replacement event. Ancient DNA analysis and carbon dating of fossil and archaeological *Eudyptula* remains confirm the recent timing of colonisation. As well as documenting cryptic species diversity, our temporal genetic analyses strongly support the time-dependency of molecular rates.

520C

Characterization of the Uncertainty of Divergence Time Estimation under Relaxed Molecular Clock Models Using Multiple Loci

Tianqi ZHU [1], Mario dos Reis[2], Ziheng Yang[2,1]
[1] *Beijing Institute of Genomics, CAS, Beijing, China,* [2] *University College London, London, UK*

Genetic sequence data provide information about the distances between species or branch lengths in a phylogeny, but not about the absolute divergence times or the evolutionary rates directly. Bayesian methods for dating species divergences estimate times and rates by assigning priors on them. Because times and rates are confounded, our posterior time estimates will not approach point values even if an infinite amount of sequence data are used in the analysis. Uncertainty in posterior time estimates is partitioned into three sources: Sampling errors in the estimates of branch lengths in the tree for each locus due to limited sequence length, variation of substitution rates among lineages and among loci, and uncertainty in fossil calibrations. Using a simple but analogous estimation problem involving the multivariate normal distribution, we predict that as the number of loci (L) goes to infinity, the variance in posterior time estimates decreases and approaches the infinite-data limit at the rate of 1/L, and the limit is independent of the number of sites in the sequence alignment. We then confirmed the predictions by using computer simulation on phylogenies of two or three species, and by analyzing a real genomic data set for six primate species. Our results suggest that with the fossil calibrations fixed, analyzing multiple loci or site partitions is the most effective way for improving the precision of posterior time estimation. However, even if a huge amount of sequence data is analyzed, considerable uncertainty will persist in time estimates.

521D

Hydrogen peroxidase responses differ in Triticum aestivum NILs in the presence of *Diuraphis noxia* resistance gene *Dn1*

Thia Schultz [1], Anna-Maria Botha[2]
[1] *University of Zululand, KwaZulu-Natal, South Africa,* [2] *Stellenbosch University, Western Cape, South Africa*

Plants defend themselves with a complex set of resistance interactions. The *Dn1* gene conveys an antibiosis-type resistance to *Diuraphis noxia* (Russian wheat aphid) in *Triticum aestivum* (wheat). Hydrogen peroxide production was studied in two *Triticum aestivum* (bread wheat) near isogenic lines (NILs), Tugela and TugelaDN under wounding conditions in order to further characterize the action of the *Dn1* gene. In order to ascertain differences in the activities of key detoxification enzymes in TugelaDN plants under wounding conditions, thylakoid-associated ascorbate peroxidase (*tAPX*), glutathione-*S*-transferase (*GSTF6*), and superoxide dismutase (*SOD*) were cloned into barley stripe mosaic virus (BSMV) to trigger silencing utilizing the virus-induced gene silencing (VIGS) technique. Silencing of *GSTF6* adversely affected H_2O_2 production. DAB staining results from *tAPX* show hydrogen peroxide production is limited to the areas around the wound sites. Silencing of *SOD* in wounded plants incurred tissue damage and chlorosis. Comparing the NILs peroxidase activity levels over time showed TugelaDN showed an earlier up-regulation in peroxidase activity when compared to Tugela. In Tugela; at 48 hours peroxidase activity assays showed high up-regulation, followed by a decrease in activity, whereas the TugelaDN plants sustained the higher level of peroxide production. In conclusion: the TugelaDN cultivar showed disparate antioxidant and peroxidase activity under wounding conditions when compared to Tugela which can be attributed to the presence of the *Dn1* gene. TugelaDN plants were found to sustain H_2O_2 concentrations over time without sustaining damage.

522A

Determining the Ploidy Level in four populations of *Trifolium tumens* (Fabaceae) using flowcytometry, chromosome counting and chloroplast number

Giti Barzin [1], Fahimeh Salimpour[2]

[1] Department of Biology, Islamic Azad University, Islamshahr Branch, Islamshahr, Tehran, Iran, [2] Department of Biology, Islamic Azad University, North Tehran Branch, Tehran, Iran

Trifolium L., is one of the largest genera in the family, with about 255 species. Some of these taxa are important forage and pasture plants. T. tumens is one of the species in Vesicaria section that has variation in morphological characters. The number of chromosomes in the mitotic metaphase stage in the meristematic cells of the root ends were counted in four populations of this taxon (Fabaceae). The base chromosome number (x = 8) in the populations of this species confirmed the views of the previous researchers. Moreover, observation of the tetraploidy state in the species T. tumens for the first time through the methods of flowcytometry, chloroplast count of stomata guard cells, morphological studies, and comparison of pollen grain size in diploid and tetraploid populations of this species was confirmed. Also, this study showed that chloroplast scoring methods could be used in Trifolium specially in T. tumens. The observations showed an increase in the ploidy level in this species. Based on the results, this new cytotype is reported from Iran for the first time.

523B

Reconstructing a fine scale recombination rate map of the great tit genome and its implication in genome evolution

Toni Gossmann, Jon Slate, Kai Zeng
University of Sheffield, Sheffield, UK

A special feature of birds is the very high karyotype conservation - bird genomes usually consists of around a dozen large macrochromosomes and numerous tiny microchromosomes. This heterogeneity in chromosome size causes a substantial variation in recombination rate between chromosomes. There is also known variation within macrochromosomes. Nonetheless, despite the stable karyotype, intrachromosomal rearrangements were common during avian genome evolution. However, it is unclear how this affected the long-term recombination environment in bird genomes. Unlike other commonly studied bird species, the great tit (*Parus major*) has a large, panmictic population, making it an ideal system to study genome-wide patterns of recombination rate variation and its effects on patterns of polymorphism and divergence. So far, little focus has been given on fine-scale recombination variation in the avian lineage and therefore we use resequencing data from 29 great tit individuals to reconstruct a fine scale recombination map. We also scan the genome for traces of positive and purifying selection and compare sequence divergence to other available bird genomes in protein coding genes as well as noncoding regions to illustrate how recombination has facilitated the action of natural selection due to Hill-Robertson-Interference. In sum, this study reveals the fundamental role of recombination and its impacts on the composition of genomes.

524C

Evolutionary change in thermoperception between *Xenopus* species inhabiting different thermal environments: from molecules to behavior

Shigeru Saito [1,2], Masashi Ohkita[3], Claire Saito[1], Toshio Ohta[3], Makoto Tominaga[1,2]
[1] Okazaki Institute for Integrative Bioscience (NIPS), Okazaki, Aichi, Japan, [2] SOKENDAI (The Graduate University for Advanced Studies), Okazaki, Aichi, Japan, [3] Tottori University, Tottori, Tottori, Japan

Thermoperception must have shifted among species inhabiting different thermal niches. However, molecular evolutionary mechanisms for thermoperception remain elusive. Recent molecular understanding of thermoperception enables us to clarify how animals evolved their thermoperception by experimental approaches. Here we focused on two species of clawed frogs (*Xenopus laevis* and *Xenopus tropicalis*) inhabiting different thermal environments. We first compared behavioral responses and found that *X. laevis* was much more sensitive to heat stimulation than *X. tropicalis*. Primary cultured sensory neurons also exhibited similar difference between the two species. Thus, we compared thermal responses of two ion channels (TRPV1 and TRPA1) which serve as heat sensors by electrophysiological experiments. As a result, both channels exhibited clear species differences in different manners. In the case of TRPA1, temperature threshold for activation in *X. laevis* TRPA1 was lower than that of *X. tropicalis* TRPA1. In the case of TRPV1, although temperature thresholds for activation were almost the same between the two species, clear species difference was observed with repeated heat stimulation. *X. laevis* TRPV1 exhibited almost full activity in the first heat stimulation, while *X. tropicalis* TRPV1 exhibited only a partial activity. These results indicate that functional changes of two different heat sensors contributed to the evolutionary shift in thermoperception in *Xenopus* lineages. Moreover, we identified a single amino acid substitution that is largely responsible for the species difference of TRPV1 thermal responses, which may suggest that subtle amino acid substitutions can cause functional changes of thermal sensors.

525D

The impact of antibiotic selection on fitness of *E. coli* O157:H7 on plants.

Oliver George [1], Martin Woodward[1], Liz Shaw[1], Simon Andrews[1], Penny Hirsch[2], Tim Mauchline[2], Robert Jackson[1]

[1] University of Reading, Reading, UK, [2] Rothamsted Research, Harpenden, UK

Escherichia coli O157:H7 is an enteric bacterium that resides harmlessly in the rectal passage of cows, but is capable of causing serious disease and death in humans. Several hundred cases of food poisoning and disease are reported annually. Transmission of the bacterium occurs through meat contamination or through application of contaminated slurry to fruit and vegetables that are eaten raw. Several independent studies have clearly shown that *E. coli* O157:H7 can establish both epiphytic and endophytic plant colonisation.

This study has been investigating the impact of antibiotic resistance on bacterial performance. Antibiotic resistant strains can emerge from animals treated with antibiotics, yet we know very little about the fate of mutants in the environment. We therefore used experimental evolution to generate mutants resistant to rifampicin, nalidixic acid and streptomycin; tiamulin-resistant strains have not yet been obtained. In vitro growth analysis detected no change in performance of any of the nalidixic acid or streptomycin strains, although there was some variation with rifampicin-resistant strains. Motility tests found several strains with reduced spreading motility. These strains are being assayed against their parent strains *in planta* and genome sequences determined to identify any genomic changes and therefore providing insight to the systems used to survive. It will also provide key information on the adaptability of the pathogen to changing environments, an important attribute in helping growers to prevent crop contamination.

526A

Statistical analysis of ribosomal DNA loci distribution on animal chromosomes

Jana Sochorová [1], Sonia Garcia[2], Aleš Kovarik[1]

[1] Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic, [2] Laboratori de Botànica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Catalonia, Spain

Eukaryotic ribosomal DNA (rDNA) loci vary both in positions and numbers. rDNA encodes the four critical genes needed for ribosome function: the 5S, 5.8S, 18S and 26S rRNA. In animals, the 18S-5.8S-28S genes (45S) are always linked in a single 45S rDNA unit. The 5S genes occur as separate tandems or linked to the 45S units. Here, we describe a database containing information about the chromosomal position, number of 5S and 45S and their relative position (linked /co-localized/ separated) in more than 790 animal species. The data based on in situ hybridization of metaphase chromosomes have been collected from more than 320 research articles. Species representation is descending in the following order: Fishes (47% of species), arthropods, mammals, mollusks, reptiles and amphibians. Fishes also have the greatest variability in the number of rDNA loci (1 to 21 5S sites and 1 to 23 45S sites per haploid genome) which may be explained by overrepresentation of these genera in the database, but other factors such as ploidy, karyotypic divergence and genome plasticity should also be considered. Average number of rDNA loci is 2.6 per haploid genome. The 5S and 45S loci are organized as separate tandems in most of genomes (79%). Thirty percent of species have 5S rDNA in a single locus, whereas 45% of species have 45S rDNA in a single locus. The animal database will become accessible via a web-based interphase as a mirror to recently issued Plant rDNA database (http://www.plantrdnadatabase.com/).

527B

Decrypting plant evolution in the Azores archipelago

Mónica Moura [1], Mark Carine[2], Elisabete F. Dias[1], Miguel Menezes de Sequeira[3], Hanno Schaefer[4], Lurdes Silva[1], Carlos Melo[5], Ana Hipólito[6], Sérgio Ávila[1,7], Luís Silva[1]
[1] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Ponta Delgada, Portugal, [2] Department of Life Sciences, The Natural History Museum, London, UK, [3] Madeira Botanical Group, Life Science Centre, University of Madeira, Madeira, Portugal, [4] Plant Biodiversity Research, Technische Universität München, München, Germany, [5] Departamento de Geociências, Universidade dos Açores, Ponta Delgada, Portugal, [6] *Centro de Vulcanologia e Avaliação de Riscos Geológicos, Ponta Delgada, Portugal, [7] Faculdade de Ciências da Universidade do Porto, Porto, Portugal*

Knowledge of the Azores flora has historically relied on the study of collections obtained during sporadic and brief visits conducted by collectors and naturalists coming from abroad. However, the far oceanic location of the Azores made it more inaccessible to these early studies than other Atlantic archipelagos, such as the Canary Islands and Madeira. This has led to an early bias in the botanical knowledge within the Macaronesian region, with the Canary and Madeira archipelagos historically possessing the most well-studied floras while the Azores (and Cape Verde), remained of a more restricted interest.
In the Azores, research of the flora was limited not only by its remote geographic location but also by the small number of endemic plant taxa documented in early checklists, which led to the impression that the archipelago did not seem to have much to offer in terms of biodiversity, in comparison with the multitude of strikingly different taxa in the Canary Islands.
A recent series of studies at the molecular level have started to unravel the real amount of plant biodiversity within the Azores. These studies revealed new species and subspecies and the existence of intra-archipelago genetic diversity patterns previously unknown. Building on this amount of new, archipelago-wide, molecular data available for most Azorean endemic plant lineages, obtained from population genetics and phylogenetic markers, we present a meta-analysis of these evolutionary diversity patterns in the light of the geological evolution of the Azores archipelago.

528C

The genomic analysis of the Andaman islanders gives a new insight on the spread of modern humans in Asia

Mayukh Mondal [1], Ferran Casals[1], Jaume Bertranpetit[1], Analabha Basu[2], Partha Majumder[2]
[1] University of Pompeu Fabra, Barcelona,Catalonia, Spain, [2] National Institute of Bio-Medical Genomics, Kalyani, West Bengal, India

Motivation:
Andamanese is a native-population living in Andaman Island, India. Andamanese and other "Negrito" populations in South-East Asia are phenotypically different from other Indian populations and resemble to African pygmies. Andamanese have been isolated from other populations with limited admixture with other populations. Thus they are a key population to study the dispersal of humans into South-East Asia.

Objectives:
Using whole-genome-sequence data of Andamanese we tested (i) whether these populations share an origin with other Asian populations, (ii) quantify the presence of other ancient hominids in Andamanese, and (iii) how natural selection has shaped the genome of these populations.

Methods:
We generated whole-genome-sequence data of 10 individuals from Andaman islands, and 60 individuals from Mainland India.

Results:
Using D-stat our analysis suggest: (i) Andamanese share a common ancestry with other Asian populations, originated in a single out of Africa expansion. (ii) They have similar amount of Neanderthal ancestry to other Out-of-Africa populations, though they lack Denisova ancestry. We also find, in the Andaman genomes, traces of ancient genomes that are neither Neanderthal nor Denisovan. (iii) Selection have acted strongly on height related genes in Andamanese populations.

Conclusions:
Our findings suggest that Andamanese populations don't have a different origin than other Asian populations,contrary to the hypothesis of a first Out-of-Africa that would populate the Andaman Islands, where they would remain, having been substituted in mainland Asia. Their phenotypical differences are mainly due to strong selection on specific type of genes (i.e. height) which might be the result of convergent-evolution producing the Negrito phenotype.

529D

Genotype-frequency estimation from high-throughput sequencing data.

Takahiro Maruki, Michael Lynch
Indiana University, Bloomington, IN, USA

Rapidly improving high-throughput sequencing technologies provide unprecedented opportunities for carrying out population-genomic analyses in various organisms. To take full advantage of the new technologies, we have developed a maximum-likelihood (ML) method for estimating allele and genotype frequencies from sequence data from a population with any type of mating system and internal population structure. Furthermore, we provide systematic frameworks for testing the statistical significance of polymorphisms and their genotypic deviations from Hardy-Weinberg equilibrium (HWE). Examination of the performance of the proposed method using computer simulations demonstrates essentially unbiased estimates of allele and genotype frequencies and sampling variances that asymptotically approach the theoretical minimum values. The statistical tests for detecting polymorphisms and HWE deviation are conservative and have reasonably high power. To examine the performance of the proposed method when applied to real data, we analyzed low-coverage human data on chromosome 6 from the CEU population in the 1000 Genomes Project. Our method identified a number of polymorphic sites harboring rare alleles with reasonably high power. Consistent with previous results, the vast majority of polymorphic sites did not significantly deviate from HWE. However, we found a striking excess of polymorphisms with high minor allele frequencies that show highly negative inbreeding coefficients. We examined the cause of the observation by studying spatial patterns of heterozygosity and inbreeding coefficient estimates along the chromosome. Our results show the importance of relaxing the HWE assumption for subsequent analyses examining signatures of natural selection and population demography.

530A

The distribution of scientific effort among genes: correlations with evolutionary conservation and human disease relevance

Travis Struck, Brian Mannakee, Ryan Gutenkunst
University of Arizona, Tucson, AZ, USA

Extensive effort in molecular biology has been devoted to uncovering and studying interesting protein-coding genes. Those efforts have been biased, however, by a propensity to study genes that have been previously studied, due to familiarity or the availability of research tools. The advent of whole-genome sequencing provides the opportunity to assess how the distribution of scientific effort aligns with unbiased measures of gene relevance based on evolutionary conservation or association with disease. We measured scientific effort by number of publications and analyzed correlations with evolutionary conservation in multiple species and with genome-wide association study (GWAS) results in humans. We found a positive correlation between evolutionary conservation and publications for genes in S. cerevisiae, but no correlation for D. melanogaster, C. elegans, and H. sapiens. We suspect this difference is, in part, due to the use of the multi-cellular organisms as models for human disease. On the other hand, we found a strong correlation between publications associated with a gene and its probability of being a significant hit in at least one GWAS study. Surprisingly, we found no correlation between the effect size of a hit and the number of publications associated with that gene, but we found a negative correlation between the p-value of the hit and publications. Our results raise interesting questions about how biases due to prior research interact with genome-wide results in guiding future research efforts.

531B

Population genomic analysis of the infectious and integrated *Wolbachia pipientis* genomes in *Drosophila ananassae*

Jae Young Choi, Jaclyn Bubnell, Charles Aquadro
Cornell University, Ithaca, NY, USA

Coevolution between *Drosophila* and its maternally inherited bacterial endosymbiont *Wolbachia pipientis* has many intriguing aspects. For example, *Drosophila ananassae* hosts two forms of W. pipientis genomes where one originates from the infectious bacterial genome, and the other is found integrated into the host nuclear genome. Here, we have characterized the genomics of the infectious and integrated genomes of *W. pipientis* infecting *D. ananassae* (wAna), by whole genome sequencing of several strains of *D. ananassae* from around the world that have either the infectious or the integrated forms of the wAna genomes. Our results indicate evolutionarily stable maternal transmission for the infectious wAna genome since the species' initial infection, suggesting a stable long-term coevolution with its host. On the other hand, the integrated wAna genome shows pseudogene-like characteristics accumulating many variants that are predicted to have deleterious effects if present in an actively transcribed and/or infectious bacterial genome. Additional strain genotyping through PCR of structural variations indicated several substrains of infectious wAna variants, while suggesting a single wAna strain originally integrated into the host *D. ananassae* genome. Most surprisingly, our analysis of the copy number and segregating variants of the integrated wAna genome indicates the presence of close to two complete wAna genomes integrated into the host genome in all strains of *D. ananassae* sampled that contain an integration. This is predicted to have had only minimal negative fitness consequence as evidenced by its presence across a wide geographical range for *D. ananassae*.

532C

Detecting convergent molecular evolution in eusocial insects

Jenny Greenwood [1], Tristan Bitard-Feildel[1], Carsten Kemena[1], Nicolas Terrapon[2], Florentine Schaub[3], Erich Bornberg-Bauer[1], Judith Korb[3]
[1] Institute for Evolution and Biodiversity, University of Muenster, Muenster, Germany, [2] Architecture et Fonction des Macromolécules Biologiques, Université Aix-Marseille, Marseille, France, [3] Institute of Biology I, University of Freiburg, Freiburg, Germany

Convergence is a pervasive phenomenon in evolution, and nowhere is it better exemplified than in insect societies. Eusociality has evolved multiple times within insects and is characterised by division of labour and cooperative care of offspring. Determining how such complex behavioural and morphological changes arose multiple times is a challenging problem.

With increasing knowledge about the genomes of insects, including numerous recently sequenced, eusocial species, the molecular underpinnings of eusociality are beginning to be elucidated. Already, numerous gene family expansions have been implicated in the evolution of eusociality and cis-regulatory changes have been identified which appear similar across eusocial taxa[1]. Such results offer the possibility that some of the molecular mechanisms underlying eusociality are convergent.

Using genomes from multiple eusocial hymenopterans and isopterans (including an unpublished termite species), as well as solitary species, we investigated whether eusocial insects exhibit convergent evolution at the protein sequence level, with a special focus on those genes with known roles in eusociality. We also investigated whether these genes are under selective pressure. We present the first results of these analyses.

1. Simola et al., Genome Research 23:1235-1247.

533D

Estimating Identical-By-Descent tracts from low coverage NGS data

Filipe G Vieira [1], Anders Albrechtsen[2], Rasmus Nielsen[3]

[1] Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [2] Department of Biology, University of Copenhagen, Copenhagen, Denmark, [3] Department Integrative Biology, University California Berkeley, Berkeley, USA

Genome-wide patterns of Identical-By-Descent (IBD) tracts and their variation across individuals provide a valuable insight into human genetic diversity and evolutionary history. Methods have been developed to infer these tracts but they are based on marker/genotype data, due to the low error rates. Next Generation Sequencing (NGS) technologies have revolutionized research in evolutionary biology by both increasing speed and reducing costs. However, these data typically have high error rates due to multiple factors (from random sampling of homologous alleles, to sequencing or alignment errors) and, furthermore, many studies rely on low coverage sequence data (< 5× per site per individual), causing SNP and genotype calling to be associated with considerable statistical uncertainty.

Recent methods rely on probabilistic frameworks to account for these errors, integrating the base quality score together with other error sources to calculate an overall "genotype likelihood". This likelihood function can then be combined with a prior to calculate a posterior probability for the genotype. Here, we present a new Hidden-Markov-Model based method to estimate IBD tracts, specially suited to low coverage NGS data since it takes the uncertainty in the data into account by working with genotype likelihoods. Furthermore, and apart from the IBD tracts, this new method also estimates genome-wide inbreeding coefficients that can be used as priors in other analyses. We assess its performance both on simulated data and a subset of the 1000 genomes, looking into several combinations of sample size and coverage, and show accurate inferences for sequencing coverages as low as 2x.

534A

Genomic mechanisms of thermoadaptation : Methanococcales as a case study

Michel Lecocq [1], Mathieu Groussin[2], Manolo Gouy[1], Céline Brochier-Armanet[1]
[1] Laboratoire de Biométrie et Biologie Évolutive (UMR 5558) CNRS/UCBL 1, Villeurbanne, France, [2] Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Temperature is a major environmental factor influencing prokaryotic molecular evolution. In particular, temperature impacts the amino acid composition and evolutionary rates of proteins (Zeldovich, 2007; Stetter, 2006). However the genomic processes driving proteome adaptation to temperature are not well characterized, and the role played by horizontal gene transfers (HGT) is still an ongoing debate (Nelson et al., 1999; Omelchenko et al., 2005).
Methanococcales represent an order of methanogenic euryarchaeota. They are widespread, being present in various environments from marine deep ocean hydrothermal vents to salt marshes. Members of this order have optimal growth temperature ranging from 35°C up to 95°C. However, they are homogeneous with respect to other factors suspected to influence substitution processes, such as metabolism, genome size and genomic G+C contents. It makes Methanococcales the ideal biological model to decipher the molecular mechanisms related to the temperature adaptation of proteomes (Haney, 1999). We conducted an in depth analysis of 18 publicly available Methanococcales proteomes. We showed that all proteins are significantly impacted regardless of their function and we detected moderate rates of HGT, suggesting that proteome adaptation to temperature is mainly due to mutational processes rather than to massive gene acquisition. The use of temperature predictive models (Boussau et al., 2008; Lartillot, 2014) coupled with a recently developed time-heterogeneous substitution model for ancestral protein sequence reconstruction (Groussin et al., 2013) allowed us to determine the evolution of optimal growth temperature all along the diversification of Methanococcales and to disclose shifts in the substitution patterns underlying temperature adaptation.

535B

The determinants of protein evolution are highly conserved across Drosophila species

David Castellano [1,2], Marta Coronado[1,2]

[1] Institut de Biotecnologia i de Biomedicina, Barcelona, Spain, [2] Departament de Genetica i Microbiologia, Barcelona, Spain

Proteins evolve at different rates due to a myriad of interacting factors, from protein structure and function, to the genomic position of the encoding genes or their expression patterns. In this work we wonder whether the proportion of variation explained by those factors depends on the time since divergence from the ancestral species. To do this we compile a dataset with more than 6000 one-to-one orthologous genes in eight Drosophila species, from D. melanogaster to D. grimshawi. Then we estimate for each gene and pair of species the number of nonsynonymous changes per nonsynonymous site (dN). Finally, a machine learning algorithm is used to predict dN according to fifteen gene features retrieved from D. melanogaster genomic data. We would expect a reduction in our predictive accuracy over time if the determinants of protein evolution evolve. Surprisingly we do not find any correlation between divergence time and the amount of explained variation in dN. After extensive cross-validation we are able to explain on average ~53% of the variance in dN for all pairs of species. We find that features related to gene expression explain ~22% of the variation in dN. Gene architectonic properties, frame-preserving indel substitutions and the rate of synonymous substitutions explain each one around 9%, while features related to protein structure and gene context explain altogether less than 3%. We conclude that protein evolution is highly predictable across Drosophila species due to the high degree of conservation of their genetic determinants.

536C

Molecular characterization of MHC class II in the Australian invasive cane toad reveals multiple splice variants

Mette Lillie [1], Richard Shine[2], Katherine Belov[1]
[1] Faculty of Veterinary Science, University of Sydney, NSW, Australia, [2] School of Biological Sciences, University of Sydney, NSW, Australia

The cane toad (*Rhinella marina*) invasion of Australia has encompassed more than 1 million square kilometers and continues to spread. This invasion has accelerated over time due to the rapid adaptation for dispersive traits observed in toads of the invasion front. This has come at the cost of the immune system, with lower investment in some immune functions by these highly dispersive toads. To investigate the cane toad's immunogenetics, we characterized the MHC class II in the Australian cane toad. We characterized four class IIA and three class IIB loci, with preliminary observations suggesting very low allelic diversity at all loci. Bottlenecking as a result of the cane toad's complex history of introductions en route to Australia would have contributed to this. We also observed two splice isoforms at two class IIA and two class IIB loci. One isoform lacks the domain essential to peptide binding and presentation. The other isoform lacks the connecting peptide, transmembrane domain and the cytoplasmic region, and is likely to be a soluble MHC product. These MHC class II splice variants in the cane toad is the first documented in any anuran species, with the only other case being in three caudate amphibian species. These results suggest a significant role of alternative splicing of MHC loci in the Australian cane toad, and may provide the first indication that alternative splicing of the MHC plays a wider functional role within Amphibia.

537D

Detecting periods of rapid evolution on large phylogenetic trees

Pablo Duchen, Christoph Leuenberger, Daniel Wegmann
University of Fribourg, Fribourg, Switzerland

While it is now widely accepted that phenotypic evolution may not necessarily happen at a constant rate, the frequency and phylogenetic position of periods of rapid evolution remains unclear. In his highly influential view of evolution, GG Simpson (1944) postulated that such evolutionary jumps occur when organisms transition into new adaptive zones, for instance after dispersal into a new geographic area, after rapid climatic changes, following the appearance of an evolutionary novelty allowing species to exploit resources differently, or when an ecological niche is freed by extinction of another lineage. Only recently, large, accurate and well calibrated phylogenetic trees have become available that allow testing this hypothesis directly, yet inferring evolutionary jumps on such trees remains challenging. Here, I propose to model evolutionary jumps as a Lévy process and to develop and implement a computationally efficient algorithm to accurately infer the parameters of this model. I will then make use of this development, as well as recent methods to infer ecological niches from remote sensing data, to infer the rate of evolutionary jumps and their phylogenetic location on the recently published phylogeny encompassing close to all birds.

538A

Molecular signatures of balancing selection - a closer look at population genetic data from Human, Drosophila and Arabidopsis

Alexander Klassmann
University of Cologne, Cologne, Germany

Although the theory of long-term balancing selection is fairly well understood, the relevance of balancing selection as an important evolutionary force in natural populations - besides textbook examples of immune and bloodgroup genes - has been under debate for long. Detection of long-term balancing selection in molecular data is notoriously difficult because recombination events restrict its effect to short chromosomal fragments. Nevertheless several approaches have been proposed and applied to whole-genome population genetic data. I re-analyzed a selected set of biologically plausible loci from three species (human, Drosophila, Arabidopsis). I report results from various statistical tests for balancing selection on these loci in order to suggest a benchmark set for future tests.

539B

Ancient Taíno genome sheds light on the peopling of the Caribbean

Hannes Schroeder [1,2], Maria Ávila-Arcos[3], Marcela Sandoval[1], Jason Laffoon[2], Robert Carr[4], Willam Schaffer[5], Rachel Hopkins[6], Tom Higham[6], Carles Lalueza-Fox[7], Thomas Gilbert[1], Carlos Bustamante[3], Corinne Hofman[2], Eske Willerslev[1]

[1] Center for Geogenetics, University of Copenhagen, Copenhagen, Denmark, [2] Faculty of Archaeology, Leiden University, Leiden, The Netherlands, [3] Department of Genetics, Stanford University, Stanford, CA, USA, [4] Archaeological and Historical Conservancy Inc., Davie, FL, USA, [5] Center for Bioarchaeological Research, Arizona State University, Tempe, AZ, USA, [6] Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford, UK, [7] Institute of Evolutionary Biology, CSIC, University Pompeu Fabra, Barcelona, Spain

The Caribbean was the last part of the Americas to be explored and occupied by humans. But how was it settled, and where did the people come from who Columbus encountered in 1492? These questions have been the source of heated debate in Caribbean archaeology since at least the 1950s. Theoretically, there are three entry points to the Caribbean: Florida in the north, the Yucatan peninsula in the west, and Trinidad and Tobago in the south. Based on archaeological evidence, the first people to enter the Caribbean are thought to have originated from Yucatan around 4000 BC. Later, successive waves of migration are thought to have reached the Caribbean from South America, expanding rapidly northward into the Lesser Antilles from around 500 BC. These various migrations of people ultimately culminated in the well-known Taíno culture that Columbus encountered when he reached the Caribbean in 1492. We sequenced the genome of a Lucayan-Taíno from the Bahamas and use it to trace the origins of the Taíno people to South America. When compared with modern genome-wide data from the Americas, the Lucayan genome clusters most closely with Arawakan speakers living in present-day Colombia and Venezuela, providing independent evidence for the early peopling of the Caribbean.

540C

Genetic factors affecting EBV load in transformed LCLs from the 1000 genome project: a GWAS on transformation

Rajendra Mandage, Gabriel Santpere, Arcadi Navarro
Institute of Evolutionary Biology (Universitat Pompeu Fabra-CSIC),PRBB, Barcelona, Spain

Worldwide, >90% of the adult population is infected by the Epstein-Barr virus (EBV), which has been an evolutionary companion of our lineage for millions of years. EBV remains within the host after primary infection and can cause mononucleosis. There is also evidence linking it with multiple sclerosis and different types of tumors. Although the EBV has been the focus of extensive research work, much remains unknown about what makes some individuals more sensitive to EBV infection and to adverse outcomes as a result of infection.

The EBV is used to transform B-cells into lymphoblastoid cell lines (LCLs). We hypothesized that differences among individual LCLs in the EBV load resulting from EBV transformation may reflect different genetic susceptibility to EBV infection. To test this hypothesis, we retrieved whole-genome sequenced LCL reads 2215 samples sequenced within the 1000 Genome Project and derived from 26 different populations worldwide. We subjected these samples to *in silico* viral load estimation, and the accuracy was validated by RT-PCR.

Our results showed considerable differences in viral load among populations, while no significant difference was observed between males and females. The proper estimation of EBV load has made it possible to perform a genome wide association analysis (GWAS) between estimated EBV load and genetic variants determined within the 1000 genome project samples. GWAS yielded many putative candidate genes that necessitate further evaluation to reveal the biological mechanisms underlying EBV load and EBV associated diseases.

541D

Uncertainty in ancestry inference

Suyash Shringarpure, Carlos Bustamante
Stanford University, Stanford, CA, USA

Likelihood-based ancestry inference methods such as STRUCTURE, ADMIXTURE, frappe are used to examine population structure in admixed genomic samples. These methods can recover admixture proportions, i.e., fractional contributions from multiple populations to the genomes of the individuals being studied. They can estimate the uncertainty in admixture proportion estimates using bootstrapping. However, bootstrapping is computationally expensive and resulting estimates have no genetic interpretation.

We developed closed-form expressions for the uncertainty in admixture proportion estimates. We show that the uncertainty depends directly on the number of SNPs and the genetic distance between source populations. We developed uncertainty results for two-population and multi-population admixtures. We created tests for estimating whether a specific ancestry contribution to an individual's genome is nonzero. We also developed results quantifying the impact of using proxy populations, which is common for ancient DNA analysis.

Using data from the 1000 Genomes project, we show that our results are accurate when 1000 or more unlinked SNPs are used for analysis. We demonstrate the accuracy of our method for between-continents admixture (CEU+YRI for two-population, CHB+CEU+YRI for multi-population,) and within-Europe admixture (GBR+CEU for two-population, FIN+GBR+CEU for multi-population). Our method is nearly 40 times faster than the bootstrap-based uncertainty estimation in ADMIXTURE.

Our results explain the rationale underlying some of the empirical best practices in ancestry inference. They can be used to efficiently incorporate uncertainty into ancestry analyses of large genomic samples. By providing guidelines about marker density required to obtain desired levels of uncertainty, they can also guide design of future studies.

542A

Phylogenetic origins of the Avian MHC Class IIB

Julien Goebel [1], Reto Burri[2], Marta Promerová[3], Luca Fumagalli[1]

[1] University of Lausanne, Department of Ecology and Evolution, Laboratory for Conservation Biology, Lausanne, Switzerland, [2] Uppsala University, Department of Evolutionary Biology, Uppsala, Sweden, [3] Uppsala University, Department of Medical Biochemistry and Microbiology, Biomedical Centre, Uppsala, Sweden

Gene duplication is thought to be one of the primary sources of adaptive evolution. The major histocompatibility complex (MHC) multigene family is a well-known example of an important ecological novelty coming from duplication events. It encodes genes involved in the vertebrates' adaptive immune response and stands for the most polymorphic genetic system known to date in vertebrates. Its frequent gene duplications and losses result in a marked variation in gene number and genomic organization between vertebrates. Mammalian MHC Class II β chain (MHCIIB) paralogs evolve independently and their duplication history can be traced back over tens of millions of years. Recent studies on birds report also strong evidence for the persistence of two ancient MHCIIB lineages (DAB1 & DAB2) over at least 100 million years, unequally distributed between bird species. We isolated and recovered MHCIIB genes from species all over the avian phylogeny. We figured out mechanisms that broke the duplication signal by reconstructing avian MHCIIB recombination and selection histories, and assessing its concerted evolution pattern. Thus, we could identify gene regions that reflect the history of the duplication. We reconstructed the phylogenetic history of avian MHCIIB and provided an accurate estimate for the origin of the two ancestral lineages.

543B

Analysis of a nonhuman primate *Mycobacterium leprae* strain: implications for zoonotic transmission of mycobacterial pathogens

Tanvi Prasad Honap [1], Luz-Andrea Pfister[2], Gideon Erkenswick[3], Mrinalini Watsa[4], Anne Stone[2]
[1] School of Life Sciences, Arizona State University, Tempe, AZ, USA, [2] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA, [3] Department of Biology, University of Missouri St. Louis, St. Louis, MO, USA, [4] Department of Anthropology, Washington University in St. Louis, St. Louis, MO, USA

The majority of emerging or re-emerging infectious diseases are zoonoses. For example, tuberculosis, which is caused by members of the *Mycobacterium tuberculosis* complex (MTBC), shows bidirectional transmission among humans and several other mammalian species including nonhuman primates. *M. leprae* causes leprosy, primarily affecting humans and nine-banded armadillos, with some isolated cases of natural leprosy observed among nonhuman primates. Here, we present an analysis of whole-genome sequence data for an *M. leprae* strain isolated from a wild-born sooty mangabey from Nigeria. The animal was imported to the USA for research purposes and developed leprosy without any contact with a known source. Our analyses determined that the mangabey *M. leprae* strain has 22 unique single-nucleotide polymorphisms and is closely related to human *M. leprae* Branch 4 strains commonly found in West Africa. This suggests that the mangabey may have acquired leprosy from a human in Nigeria. In regions endemic for tuberculosis or leprosy, these pathogens might be transmitted between human and nonhuman primates that live in close contact. It is necessary to broadly screen wild and captive nonhuman primate populations, living in proximity to humans, for the presence of mycobacterial pathogens. A recent study suggests the MTBC may be present in captive nonhuman primates in Peru. Here, we also present the results of a phylogeographic screening, for the MTBC and *M. leprae*, of wild nonhuman primate populations from the Peruvian Amazon.

544C

The maternal history of the sable antelope (*Hippotragus niger*) inferred from the genomic analysis of complete mitochondrial sequences

Joana L. Rocha [1,2], Pedro Vaz Pinto[1,3], Hans R. Siegismund[5], Matthias Meyer[6], Bettine Jansen vanVuuren[4], Nuno Ferrand[1,2], Raquel Godinho[1,2]

[1] CIBIO/InBio – Research Center in Biodiversity and Genetic Resources, Vairão, Portugal, [2] Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal, [3] ISCED-Instituto Superior de Ciências de Educação da Huíla, Huíla, Angola, [4] Department of Zoology, Faculty of Sciences, University of Johannesburg, Johannesburg, South Africa, [5] Department of Biology, University of Copenhagen, Copenhagen, Denmark, [6] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

The sable antelope (*Hippotragus niger*) is one of the most majestic mammals scattered throughout the African savanna woodlands. Although previous phylogeographic analyses using mtDNA fragments have contributed to our understanding of the evolutionary patterns of *H. niger*, conflicting and unresolved relationships remain. In this study we provide the most comprehensive dataset of *H. niger* to date covering the species whole geographic range. To this end, we made use of available target-enrichment strategies for next-generation sequencing to generate 215 complete mitogenomes from modern and ancient samples. With our extensive sample size and geographic coverage, we have been able to elucidate the phylogeographic patterns and the timing of divergence of the different *H. niger* haplogroups. This allowed explaining a previously described highly divergent haplogroup as the result of a mitochondrial introgression event from an extinct lineage that diverged from *H. niger* around 1.7 mya. We also hypothesize that climatic and geomorphological changes occurring during the Pleistocene might be responsible for most of the species maternal history. This allowed us to explain the close phylogenetic relationship between the iconic giant sable antelope of Angola and sables from Tanzania and Malawi. We believe that ancestral population(s) to these sables might have been expanded in savanna habitat regions in between their current distribution, followed by dispersal into Angola. Ultimately, the detailed phylogeographic scenario provided for the species can be used to illuminate the evolutionary history of other species in the regions of central, southern and eastern Africa, which are still poorly understood.

545D

Sideways demographic inference with Gaussian processes

Julia Palacios [1,2], John Wakeley[1], Sohini Ramachandran[2]
[1] *Harvard University, Cambridge, MA, USA,* [2] *Brown University, Providence, RI, USA*

Sophisticated inferential tools coupled with the coalescent model have emerged in the last decade for estimating past population sizes from genomic data. However for samples of more than size two, current Bayesian methods only apply to single locus data or assume independence among loci; recent maximum-likelihood-based methods that model recombination make constraining assumptions to analyze multilocus data and do not report measures of uncertainty for estimates. Here, we develop a Gaussian process-based Bayesian nonparametric method coupled with a sequentially Markov coalescent model (SMC') that accounts for recombination and allows us to accurately infer effective population size as function of time from a set of local genealogies. As opposed to current methods, our approach considers a broad class of recombination events, including those that do not change local genealogies. We show that our Bayesian nonparametric method outperforms likelihood-based methods that rely on discretization of the parameter space such as Multiple Sequentially Markovian Coalescent (MSMC) and the sequentially Markov conditional sampling distribution approach (diCal). We illustrate the application of the method to multiple demographic histories including population bottlenecks and exponential growth, and show that trajectories of population sizes can be estimated accurately in a variety of sampling scenarios. In simulation, our Bayesian approach produces point estimates four times more accurate than an MLE approach; furthermore, our method's credible intervals cover at least 90% of the truth 85% of the time across multiple demographic scenarios, enabling formal hypothesis testing about population size differences over time.

546A

"A genealogical interpretation of multidimensional scaling (MDS)"

Ivan Levkivskyi[1], Anna-Sapfo Malaspinas [2]
[1] *Institute for Theoretical Physics, Zuerich, Switzerland,* [2] *Centre for GeoGenetics, Copenhagen, Denmark*

Population structure plays an important role in determining the evolutionary history of a group. Single nucleotide-polymorphism (SNP) array data has already greatly informed the history of populations across the tree of life. In recent years, the unprecedented increase in sequencing data has opened up a wide range of possibilities to investigate population histories - provided one can handle such large amounts of data.

Methods based on non-parametric multidimensional statistics (more specifically principal components analysis, PCA) were first applied to genetic data more than 30 years ago. PCA has since become a standard tool in population genetics owing in particular to the low computational demand of such analyses.

In this work, we investigate a related statistical approach, namely multidimensional scaling (MDS). Following a recent study by McVean (PLoS Genet 5(10): e1000686), we first derive analytical results for two and three populations that relate the Euclidean distance between points on the MDS plots with pairwise coalescent times, for a specific genetic distance. We then perform coalescent simulations and derive the rate at which simulated data converges towards theoretical predictions as a function of the number of SNPs. Finally, we assess the impact of several parameters specific to high-throughput sequencing data (such as the depth of coverage) on PCA and MDS for the projection of the samples onto the first two dimensions.

547B

The role of selection in the evolution of grass homospermidine synthases

Anne-Maria Wesseling, Dietrich Ober
Botanical Institute, CAU Kiel, Kiel, Germany

Gene duplications are at the basis of the evolution of many new enzymatic functions and various scenarios of selection following gene duplication have been proposed. Our study focuses on the evolution of **h**omo**s**permidine **s**ynthases (HSS) in grasses (Poaceae). HSS is a key enzyme in the biosynthetic pathway of **p**yrrolizidine **a**lkaloids (PAs), a class of secondary metabolites found in a great variety of flowering plants and known for their deterring effects against herbivores. HSS derives from a gene duplicate of **d**eoxy**h**ypusine **s**ynthase (DHS), an enzyme which is found in all eukaryotes and essential for basic cell functions. By combining phylogenetic and selection analyses as well as functional assays we characterized the molecular evolution of grass HSS. We found that within the Poaceae two independent gene duplication events led to the evolution of functional *hss* genes. We compared the results of the *in silico* and *in vitro* analyses between these two lineages to investigate the extent of parallel evolution and discovered different signatures of selection in each lineage. We only found relaxed functional constraints in one of the identified *hss* lineages while the other *hss* cluster clearly experienced positive selection.

548C

Evidence of Vomeronasal Function and Subsequent Loss from Trpc2 in New World Leaf-nosed Bats (Family: Phyllostomidae)

Laurel Yohe, Christina Giordano, Liliana Dávalos
Stony Brook University, Stony Brook, NY, USA

The mammalian vomeronasal system (VNS) is a chemosensory system that detects pheromones in conspecific social interactions, and recent studies show the VNS may also respond to olfactory cues. The *transient receptor potential ion channel* (*Trpc2)* maintains the signal transduction from the VNS to the brain and is indispensable for vomeronasal organ (VNO) function. Bats and primates are the only mammals to show variation in vomeronasal function, as all other mammals have a functioning VNO. Most bats have completely lost VNO function, and in many cases their *Trpc2* gene is disrupted. Within the New World (NW) Leaf-nosed bats (Phyllostomidae), however, many species possess a well-developed VNO. Here, we analyze a 500-bp exon of *Trpc2* among > 100 species of NW bats, including phyllostomids. We find the *Trpc2* exon 2 is functional in nearly all phyllostomid lineages, and several closely related families. The exception is a small clade of nectar-feeding bats that shares several recent frameshift deletions in *Trpc2*. Selection tests show most phyllostomids are under strong purifying selection to retain a functioning vomeronasal signal transduction channel, while most other bats have a pseudogenized *Trpc2* gene. Intriguingly, a subgroup of nectar-feeding phyllostomid has experienced relaxed selection, and additional anatomical and behavioral data are needed to illuminate the complete loss of function in this group. We argue that the VNO may have served as a key adaptation enabling NW Leaf-nosed bats to diversify into novel dietary niches.

549D

Alternative translation (AT) in wild barley (*Hordeum spontaneum*) and its possible role in canalization against changing temperature

Efrat Laiba[1], Dhruv Srivastava[1,4], Katia Wostrikoff[2], Wojciech Wietrzynskib[2], Keren Bendalak[3], Francis-André Wollman[2], Tamar Ziv[3], Arie Admon[3], Eyal Fridman [1]

[1] Volcani Agricultural Research Organization (ARO), Bet Dagan, Israel, [2] Biologie Physico-Chimique (FRC-CNRS550), Paris, France, [3] Technion, Haifa, Israel, [4] The Hebrew University, Jerusalem, Israel

Modification of the canonical relationship between DNA and RNA to protein sequences by alternative translation (AT) is recently realized as a mechanism by which different organisms utilize to cope with environmental changes. To investigate the possible occurrence and significance of AT in modulating high environmental canalization, the wild barley (*Hordeum vulgare* sps. *spontaneum*) collection (B1K) is studied for photosynthetic activity and growth under different temperatures. Comparison of AT between accessions with differential phenotypic stability (i.e., environmental canalization) has identified a clear trend of specific AT in a group of highly canalized accessions. I will present this first reported case of AT in plants and will illustrate the consequences of this novel variation on protein function *in vitro* and in heterologous *Chlamydomonas reinhardtii* system. Mapping-by-sequencing and allele editing is utilized to unravel the causal mechanisms and question the evolution of AT in this non-classical model plant.

550A

The impact of reconstruction methods, phylogenetic uncertainty and branch lengths on inferences on chromosome number evolution.

Gerald M. Schneeweiss, Jamie McCann, Tod F. Stuessy, Hanna Weiss-Schneeweiss
Dep. of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

Chromosome number change plays an important role in eukaryotic evolution in general and in plant diversification and speciation in particular. Several methods to investigate character evolution within a phylogenetic framework can be chosen from, but little is known on the impact of reconstruction methods, phylogenetic uncertainty or branch lengths on inferences about chromosome number evolution. Herein, we study chromosome base number evolution in the plant genus *Melampodium*, which comprises five chromosome base numbers (x = 9, 10, 11, 12, 14). We used Bayesian methods for phylogeny inference implemented in BEAST and MrBayes, where branch lengths are proportional to time and to the amount of molecular evolution, respectively. Chromosome number reconstructions were done using general approaches of ancestral character state reconstruction (maximum parsimony in Mesquite, maximum likelihood and a Bayesian method in BayesTraits) and a likelihood method developed specifically to infer chromosome number evolution (ChromEvol). The outcome of the analyses (in terms of reconstructed chromosome number and the estimated number of gains and losses and thus the prevalence of ascending or descending dysploidy) is strongly affected by data source and method of analysis. As models implemented in ChromEvol, arguably the best suited method for inferring chromosome number evolution, may differ insignificantly with respect to their likelihoods; therefore, model averaging is advocated. Except for inferences under maximum parsimony, maximum clade credibility (MCC) trees may be poor representatives of the reconstructed character histories and thus should not be used as sole representative in chromosome number (and likely generally in character state) reconstruction.

551B

The genomic basis of ecological adaptation in a Drosophila agricultural pest

Sukanya Ramasamy [1,3], Santosh Revadi[1,2], Lino Ometto[1], David Horner[3], Davide Pisani[4], Teun Dekker[2], Gianfranco Anfora[1], Omar Rota-Stabelli[1]

[1] Research and Innovation Centre, Fondazione Edmund Mach, San Michele All Adige, Italy, [2] Chemical Ecology Unit, Swedish University of Agricultural Sciences, Alnarp, Sweden, [3] Università degli studi di Milano, Milan, Italy, [4] School of Biological Sciences and School of Earth Sciences, University of Bristol, Bristol, UK

The switch from a fermenting to a fresh fruit reproductive habit characterizes the emerging model pest *Drosophila suzukii*, but the genetic basis of this new trait is still widely unexplored. In our work, we mined the genome of two populations of D. suzukii and that of its closely related but a non-pest species, D. biarmipes; manually annotated the full repertoire of chemosensory genes; and performed thorough evolutionary studies on a 14 Drosophila phylogenetic framework. Compared to most other Drosophila species, the odorant and gustatory receptors of *D. suzukii* are characterized by an increased turnover rate, and a non-randomly distribution of evolutionary events (duplications and positive selection). In *D. suzukii*, odorant receptors that respond to some of the odours typical of ripening fruit have undergone duplication and show signs of positive selection; the most represented volatiles eliciting a response in these receptors include isoamyl-acetate, for which we could confirm a functional role in *D. suzukii* using ad-hoc behavioral assays. Conversely, some of the key receptors used to detect volatiles produced during fermentation were lost or experienced neo-functionalization in *D. suzukii*. Interestingly, the comparison of two *D. suzukii* populations showed that genes that underwent functional diversification are fairly divergent, suggesting they underwent ancient multiple neo-functionalization. Overall, our comparative analyses reveal unusual genome evolutionary events in *D. suzukii* that can be associated with adaptations to new ecological behaviors, and unveil key genes and ligands that might become target of applied control strategies.

552C

An analysis of the Leucine432Valine single nucleotide polymorphism of the cytochrome P450 1B1 gene and uterine leiomyomata in a Barbadian population

Virgil Bideau, Angela Alleyne
University of the West Indies, Cave Hill, Bridgetown, Barbados

Objective: This study assessed the association of Leu432Val polymorphism (C/G) in the cytochrome P450 1B1 (CYP1B1) gene with uterine leiomyomata risk in black Barbadian women.

Materials and Methods: The case-control study investigated 25 women; 16 clinically diagnosed with uterine leiomyomata (cases) and 9 women who were not diagnosed with the disease (controls). Genotyping was done using the polymerase chain reaction-restriction fragment length polymorphism method of the Leucine432Valine single nucleotide polymorphism. Odds ratio and Hardy-Weinberg population frequency statistics were calculated.

Results: Genotypic frequency for Leucine/Leucine, Leucine/Valine and Valine/Valine genotypes were 4, 32 and 64% respectively in the sample population. Allele C was twice as common in cases as in controls and allele G frequency was similar in both populations. Overall, a higher proportion of allele G was observed in the sample population (A=.80). The Leucine432Valine polymorphism was not associated with uterine leiomyomata (OR=0.37 95% CI: 0.057-2.35, P = 0.290) and the population was in Hardy-Weinberg equilibrium.

Conclusion: Among the cases and controls the homozygous Valine432 polymorphism was found in 80% of the sample population, suggesting that the SNP is not associated with uterine leiomyomata risk in the black Barbadian population studied. This is the first measure of CYP1B1 in a Caribbean population and it agrees with previous studies in African Americans.

553D

Replication N-domains are conserved since amniota divergence

Chun-Long Chen [1], Benjamin Audit[2], Antoine Baker[2], Nataliya Petryk[3], Malik Kahli[3], Yan Jaszczyszyn[1], Yves d'Aubenton-Carafa[1], Olivier Hyrien[3], Alain Arneodo[2], Claude Thermes[1]
[1] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, UMR 9198, Gif-sur-Yvette, France, [2] Laboratoire de Physique, ENS de Lyon, CNRS UMR5672, Lyon, France, [3] Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS), and Inserm U1024, and CNRS UMR 8197, Paris, France

Growing number of evidences have revealed that the replication program is an important factor shaping the mutation landscape of the human genome during evolution. Our previous genome studies have suggested that replication induces different mutation rates on the leading and lagging replicating strands. During evolution, this has generated strong asymmetries of genome nucleotide composition, namely the skew S=(G-C)/(G+C)+(T-A)/(T+A). Analysis of the S profile along the human genome revealed large ~1Mb-domains exhibiting a characteristic N-shaped pattern and covering >1/3 of the genome. The linear decrease of S along these "N-domains" suggested a progressive inversion of replication fork directionality RFD=R–L (R and L are the proportions of rightward- and leftward- moving forks). Recently, we have measured directly RFD in human cells using a new experimental method based on sequencing of highly purified Okazaki fragments (OK-Seq). *OK-Seq data fully verified the predicted N-shaped RFD gradients within N-domains.* Skew N-domains were observed in all studied mammals and birds but not in amphibians and fishes. Remarkably, extremities of N-domains observed in different genomes were located at homologous regions. This indicated that N-domain position and the associated replication program have been conserved during evolution since the amniota divergence (~300 MY). During this period, *the specific replication program associated with RFD gradients has shaped the N-profile of the compositional skew.* This is in full agreement with the skew profile computed from measured substitution rates. These data provide unique opportunities to determine the links between replication and genome structure and to study their evolution in vertebrates.

554A

Interspecies convergence detection by a genomic scale approach.

Carine Rey [1,2], Marie Sémon[2], Bastien Boussau[1]

[1] Laboratoire de Biologie et Biométrie Evolutive, University of Lyon, Villeurbanne, France, [2] Institut de Génomique Fonctionnelle de Lyon, ENS de Lyon, Lyon, France

Convergent phenotypes are widespread in nature, yet we only have a fragmented grasp of the genetic mechanisms underlying their evolution. Recently a controversy erupted around studies aiming at finding genome-scale convergent evolution in species with convergent phenotypes: some studies find vast amounts of genomic convergence, while others do not. We considered that the differences between those studies may come from differences in the methods used, and decided to build a robust and accurate pipeline for detection of genomic convergence. We investigated existing methods to assemble high-quality transcripts from RNA-seq reads, took special care to build accurate sequence alignments, avoided potential artifacts by using reconciliation methods to build accurate gene trees, and used substitution mapping to find cases of convergent genomic evolution. We report the results we obtained, like the study of 5393 gene families to address the controversy surrounding genomic convergence to echolocation between bats and whales.

555B

Convergent Evolution of the Human Antibody Repertoire in Response to Influenza Vaccination

Nicolas Strauli [1,2], Ryan Hernandez[1,3]

[1] University of California, San Francisco, San Francisco, CA, USA, [2] Biomedical Sciences Graduate Program, San Francisco, CA, USA, [3] Bioengineering and Therapeutic Sciences, San Francisco, CA, USA

Understanding how the conglomerate of antibody (Ab) expressing B cells in an individual (the antibody repertoire, AbR) responds to vaccination is of utmost importance to vaccine design. When an individual is vaccinated, there occurs extreme temporal dynamism within the AbR, as different lineages of B cells proliferate and apoptose in response to the antigenic stimulus in a process akin to natural selection. While these changes form the basis of adaptive immunity, they are not well characterized, nor understood. Here we analyze a high-resolution longitudinal dataset of 5 patients' response to the same tri-valent influenza vaccine (TIV). To view the AbRs in these individuals, we developed a novel bioinformatic pipeline to extract AbR information from an RNA-seq dataset. We use functional data analysis to leverage the time-series component of these data to identify the Abs that are putatively targeting TIV. We then test for convergent evolution of the AbR across individuals, and find evidence for convergence. This suggests that different individuals can use similar Abs to target the same antigen despite starting with divergent AbR. Together, these results establish a novel methodology for extracting the AbR from RNA-seq data, demonstrate that individuals' AbRs undergo a robust response to TIV, and that the underpinnings of these responses tend to involve the use of similar Abs.

556C

The driving forces of cultural complexity: Neanderthals, Modern Human, and the question of population size

Laurel Fogarty [1,2], Joe Wakano[3], Marcus Feldman[2], Kenichi Aoki[3]
[1] University College London, London, UK, [2] Stanford University, California, USA, [3] Meiji University, Tokyo, Japan

Cultural factors such as marriage practices, subsistence strategies, or tool use can alter selective pressures on genetic traits and influence the process of genetic evolution. The ecological and demographic success of humans is often attributed to their capacity to generate increasingly complex technologies, and to accumulate large amounts of useful information that can be passed between individuals and between generations. However, exactly what has driven this cultural accumulation is hotly debated. At the heart of the debate is the question of what genetic, demographic, or cultural features of behaviourally modern humans (as opposed to, say, early modern humans or Neanderthals) allowed culture to accumulate to this unprecedented level. Answering this question would allow us to explain the emergence of human behavioural modernity, the changing rate of cultural evolution observed across human evolutionary history, and the variability in cultural complexity among modern human groups. Here we show, using both an analytical model and an agent-based simulation model, that a full understanding of the emergence of behavioural modernity, and the cultural and biological evolution that has followed, depends on understanding and untangling the complex relationships between culture, genetically-determined cognitive ability, and demography.

557D

Clarifying the approximations inherent in sequentially Markov coalescent models

Peter Wilton [1], Shai Carmi[2], Asger Hobolth[3]
[1] Harvard University, Cambridge, MA, USA, [2] Columbia University, New York, NY, USA, [3] Aarhus University, Aarhus, Denmark

Sequentially Markov coalescent models have enabled the development of coalescent-based inference methods capable of analyzing entire genomes. Two sequentially Markov coalescent models are available as approximations to the ancestral recombination graph (ARG): one (the SMC) yields simpler expressions at the cost of accuracy, while the other (the SMC') is more accurate but less mathematically tractable. We derive the joint distribution of pairwise coalescence times at two fixed points along a pair of sequences evolving under the SMC', and we use this distribution to calculate a number of new quantities that illuminate the closeness between the SMC' and the ARG. We show that the SMC' is, in a particular well-defined and intuitive sense, the best possible sequentially Markov approximation to the ARG, and we use this result to demonstrate an asymptotic bias that is inherent in population size estimates obtained under the SMC and absent when using the SMC'.

558A

Global analysis of human polymorphic inversions from the InvFEST database

Sònia Casillas [1,2], Alexander Martínez-Fundichely[1], Isaac Noguera[1], Mario Cáceres[1,3]
[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain, [2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain, [3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Bellaterra, Barcelona, Spain

The newest genomic advances have uncovered an unprecedented degree of structural variation throughout genomes. However, compared to insertions and deletions, inversion prediction presents unique challenges. In order to get a reliable estimate of the real number of inversions in the human genome we have developed the InvFEST database (http://invfestdb.uab.cat). InvFEST automatically merges predictions into different inversions taking into account the resolution of each specific study, refines the breakpoint locations, and finds associations with genes and segmental duplications (SDs). It includes data on experimental validation, population frequency, functional effects, and evolutionary history. Recently incorporated features include the implementation of BreakSeq to automatically predict the generation mechanism for each inversion, analyze the DNA properties at the breakpoints, and determine the ancestral orientation. We have performed a meta-analysis of validated inversions from InvFEST to uncover the main characteristics of inversion polymorphisms in humans. Compared to simulated inversions generated between all intrachromosomal inverted SDs, inversions generated by non-allelic homologous recombination (NAHR) between inverted repeats tend to appear between the most identical, physically-close SDs, are enriched in chromosome X, and tend to break and invert genes less often than expected. Inversions not generated by homology-related mechanisms appear in regions that are more flexible and less stable than NAHR inversions, are genetically longer, and display a larger distance to the closest gene. All in all, while InvFEST aims to represent the most reliable set of human polymorphic inversions, these and further analyses underway should contribute to understand their functional and evolutionary impact in the human genome.

559B

A polymodal chemo/photoreceptor cell type from the Cnidaria of evolutionary significance

David Plachetzki, Molly Hartley, Connor Bell, Janine Appleton
the University of New Hamphsire, Durham, NH, USA

Taste 1 receptors (T1R) are class C G-protein-coupled receptors that mediate sweet and savory taste perception in vertebrates. Current evidence suggests that T1Rs originated in the lineage leading to gnathostomes, but are absent in all agnathan and invertebrate lineages examined to date. Here we present evidence from phylogenomic analyses of 36 holozoan whole genome sequences that T1Rs are present in several non-bilaterian genomes, demonstrating a pre-bilaterian origin for this sensory gene family. A corollary of this finding is that T1Rs were independently lost in several lineages including those leading to protostomes, cephalocordates and agnathans. We also describe a polymodal sensory-motor neuron (PSMN) cell type that coordinates cnidocyte-firing behavior in cnidarians. We show that transcripts of T1Rs, opsins and several components of their signal transduction cascades co-localize to PSMNs in the cnidarian *Hydra magnipapillata*. In addition, studies of cnidocyte discharge behavior demonstrate that T1Rs and opsins play opposing roles in mediating cnidocyte discharge where T1R signaling is excitatory and opsin-mediated phototransduction is inhibitory. Our findings reorder the current view of the evolutionary history of T1Rs and suggest that this sensory gene family was in fact present prior to the major diversification events in animals, but lost in several lineages independently. In addition, the nature of cnidarian PSMN cell types, where both T1Rs and opsins contribute to function, suggests novel hypotheses for the origins of animal sensory neurons and their transduction cascades.

560C

Molecular evolution of *Caenorhabditis* meiosis genes

Victoria Cattani, Matthew Rockman
New York University, New York, USA

Though many meiotic pathways and genes are conserved across eukaryotes, there is variation in meiotic recombination rates within and between species at different levels, from single recombination hotspots to large genomic regions. Recent studies have also shown that some meiotic genes can diverge quite rapidly in a variety of organisms. Meiotic drive has been proposed as a mechanism for the rapid evolution of meiotic genes, particularly for those involved with chromosome segregation. This idea might explain the observed rapid evolution of centromeric DNA in plants and animals. However, centromeric drive has been mainly explored in monocentric taxa; little is known about taxa with holocentric chromosomes. We took advantage of the extensive molecular and cytological knowledge of meiosis in *C. elegans*, the most studied organism with holocentric chromosomes, to study the evolution of meiotic genes in related *Caenorhabditis* species. We asked whether 75 meiotic genes in eight *Caenorhabditis* species show evidence of positive selection, and whether genes involved in chromosome segregation have preferentially evolved rapidly. Strikingly, we found that more than a quarter of the analyzed genes show evidence of positive selection. Contrary to expectations from a simple meiotic drive model, we found that only four out of 21 genes involved in chromosome segregation show evidence of positive selection. In fact, most genes with evidence of positive selection are involved in homolog pairing and in double strand break formation and resolution, which may be a signature of the distinctive role of crossover position in orienting the spindle in these animals.

561D

The multipartite mitochondrial genome of *Enteromyxum leei* (Myxozoa)

Dayana Yahalomi[1], Tamar Feldstein[2], Arik Diamant[3], Dorothee Huchon [1]
[1] Tel-Aviv University, Tel-Aviv, Israel, [2] The Steinhardt National Collections of Natural History Tel Aviv University, Tel-Aviv, Israel, [3] National Center for Mariculture, Israel Oceanographic and Limnological Research, Eilat, Israel

The typical animal mitochondrial genome is a compact and circular molecule of 13-16 kb. This is true for anthozoan cnidarians (e.g., corals, anemones and gorgons) but not for medusozoan cnidarians (e.g., jelly fishes, box jelly, and hydras), which have linear and fragmented mitochondrial genomes. Here we aimed to determine the sequence and structure of the first mitochondrial genome of Myxozoa, a large group of poorly characterized parasites, which have been found to be highly derived cnidarians.

Total genomic DNA of *Enteromyxum leei* was sequenced using the Illumina platform and assembled with de novo transcriptome assembly programs. In addition, large PCR fragments were amplified and sequenced to confirm the genome structure.

Our results indicate that *E. leei* possesses a unique mitochondrial genome organization compared to other cnidarians. The mitochondrial genome was found to be fragmented into six circular maxichromosomes of ~23 kbp. Each chromosome was found to harbor a large non-coding region (~15 kb) which was almost identical between chromosomes and one or two coding gene regions. The mitochondrial chromosomes show an unusual high rate of sequence evolution, and protein coding genes possess little similarity to their cnidarian homologs. Only five protein coding genes could be identified using Blast searches. No tRNA or rRNA could be identified.

These observations confirm the remarkable plasticity of mitochondrial genome organization within cnidarians. They also support the view that Myxozoa are probably not members of the Meduzosoa.

562A

Analysis of the rare variant burden in the exomes of candidate HIV-target genes in relation to HIV-acquisition and AIDS-progression

Michael Turchin [1], Sudhir Penugonda[2], Eun-Young Kim[2], Kevin Kunstman[2], Matthew Stephens[1], Steven Wolinsky[2]

[1] University of Chicago, Chicago, IL, USA, [2] Northwestern University School of Medicine, Chicago, IL, USA

Multiple genome-wide association studies (GWAS) have been conducted attempting to link common human genetic variation (minor allele frequency, MAF, >5%) to various aspects of HIV and AIDS pathology. Despite using large sample sizes (up to 2,000s) and samples representing ancestries beyond Western Europeans, the majority of associations have only been found within the human leukocyte antigen (HLA) region. Here, we present the results of a more focused gene-exome sequencing study utilizing ~550 genes whose protein-products have previously been implicated as significantly associating with the 18 HIV-1 proteins (Jäger et al. 2012). These ~550 genes were sequenced in over 700 individuals, the majority of which are of Western European descent. Individuals in this sample were classified as either sero-positive or sero-negative for associations involving HIV-acquisition, with HIV-positive individuals further being classified as slow/very-slow progressors or rapid/very-rapid progressors (where applicable), thus allowing us to look at associations regarding AIDS-progression as well. Analyses were conducted focusing on both the rare variant burden (MAF <5%) in the exomes of these genes as well as variants that are classified as 'damaging' as designated by multiple functional-prediction algorithms. Overall, we find a handful of genes outside the HLA region that appear to contain an excess of rare, 'damaging' variants in one class of individuals versus another for both HIV-acquisition and AIDS-progression. These results provide further insight into the biology of HIV-human interactions, and with proper functional follow-up and validation, may provide future directions for the development of HIV-drug targets.

563B

Genetic variability of the common barbel (*Barbus barbus*) in the main river catchment areas of Germany

Sascha Krenek, Thomas Schiller, Daniel Wehnert, Thomas U. Berendonk
TU Dresden, Dresden, Germany

Conservation of biodiversity and the use of biological resources in a sustainable manner are not only an objective of the Convention on Biological Diversity, but a common concern of humankind. To achieve these goals and to better assess the anticipated impact of human-induced global change on biodiversity, it is necessary to gather information about the genetic diversity of species.
I will present data of a project estimating the genetic diversity of the common barbel (*Barbus barbus*) in German river basins. *Barbus barbus* is a riverine cyprinid species widely distributed across Europe that performs up to 100 km wide spawning migrations. Their stocks have significantly declined due to water pollution as well as habitat loss and destruction in the second half of the 20th century. To assess the genetic diversity within and between 30 populations of this tetraploid species originating from the main river catchment areas in Germany, we are using the mitochondrial control region as well as nuclear microsatellite markers. Based on these data we could e.g. show that investigated species from river Elbe exhibit a lower haplotype diversity compared to the Rhine population, or that diversity between nearby populations is significantly different while it is not between highly distant populations from the same river catchment area. Further, in order not to harm the investigated fish and to evade animal protection laws during sampling campaigns, we have established a fast, easy and non-invasive DNA sampling method for barbels using buccal cell and body mucus swabbing.

564C

The biophysical effects of mutations in protein evolution

Tobias Sikosek [1,2], Erich Bornberg-Bauer[2], Hue Sun Chan[1]
[1] *University of Toronto, Toronto, Ontario, Canada,* [2] *University of Münster, Münster, Germany*

Certain aspects of protein evolution can only be understood in the light of structure and the biophysical rules that determine it. One such rule is the formation of a stable hydrophobic core in many if not most proteins that we see today. Another rule is the efficient folding of an extended peptide chain into a functional native form, avoiding the misfolding and misinteracting with other molecules.

While the core of a protein fold is strongly conserved, the surface can tolerate many mutations, leading to the widely observed properties of mutational robustness and adaptability that are crucial for evolution. Adaptability can arise from mutations stabilizing hidden or excited protein folds that may compete with the dominant native fold during folding and dynamics. If these hidden folds provide a new biological function, they may become relevant for adaptation, creating adaptive conflicts where native and non-native folds of the same protein molecule compete and give rise to a multi-functional protein. In fact, there is reason to believe that ancestral proteins had a multitude of competing folds and functions at the cost of reduced stability and efficiency. We have addressed these questions using simple lattice model theory as well as all-atom folding simulations.

565D

Identification of a tentative new species from the genetic limitation of sesame

Pattana Srifah Huehne [1,2], Jutamaad Satayavivad[2], Wasana Wongyai[1], Akarapong Swatdipong[1], Bhawat Wongkhamprai[2], Kisana Bhinija[2]

[1] Kasetsart University, Jatuchak, Bangkok, Thailand, [2] Chulabhorn Research Institute, Laksi, Bangkok, Thailand

Sesame (*Sesamum indicum* L.) is an ancient oilseed crop which has been widely cultivated and consumed in tropical regions such as Asia and Africa. The origin of major germplasms containing highly conserved genetic materials were presumably contributed from India and Africa countries. In this study, a tentative new species naturally emerged in Thailand is described and illustrated by molecular and morphological investigations. In particular a co-dominant marker, 12 simple sequence repeats (SSR) markers and seed morphology present adequate information regarding to this adaptive landraces compared with the cultivated sesame species growing in Thailand. The genetic changes recognized by the combination of different analyses might be a reliable tool for the taxonomy of sesame species distinction.

566A

Improvement of the Bias-Variance Tradeoff for Estimators of the Recombination Rate by Approximate Bayesian Computation

Kerstin Gärtner [1], Andreas Futschik[2,1]
[1] Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria, [2] Department of Applied Statistics, Johannes Kepler University Linz, Linz, Austria

Having good measures of the recombination rate is important for the analysis of population genetic data as well as for understanding the process of recombination itself and the distribution of linkage disequilibrium in natural populations. However, the estimation of recombination from population genetic data is challenging, as not all recombination events leave traces in the data. Thus even the best estimation methods available provide estimates that still exhibit a considerable amount of uncertainty. We investigate possible improvements of already existing estimators of the recombination rate in terms of the mean squared error by approximate Bayesian computation. In particular we focus on the popular software packages *LDhat* (McVean and Auton, 2007) and *LDhelmet* (Chan, Jenkins, and Song, 2012), which estimate the population recombination rate from DNA sequence data using composite likelihood methods. In our approach, we take the estimate of *LDhat* (and *LDhelmet* respectively) as summary statistic to perform approximate Bayesian computation. Hereby the programs *msms* (Ewing and Hermisson, 2010) and *ms2dna* (Haubold and Pfaffelhuber, 2013) are used to simulate DNA sequence data with specified recombination rates. We compare different prior distributions for the recombination rate and different methods of approximate Bayesian computation as a rejection algorithm or regression correction.

When using a uniform prior, our approach delivers estimators with a smaller mean squared error compared to *LDhat* (and *LDhelmet* respectively) when the population recombination rate is larger than approximately 0.013/bp. It seems to be even possible to construct uniformly better estimates with an appropriate choice of the prior distribution.

567B

Evolution of cycad plastomes: genomic stasis and existence of biased gene conversion toward elevated GC content

Chung-Shien Wu, Shu-Miaw Chaw
Biodiversity Research Center, Taipei, Taiwan

Cycads (Cycadophyta), possessing some 300 living species, are a basal clade of seed plants. They were extremely common during the Jurassic and have since changed little morphologically. However, a comprehensive study of cycad plastomes has never been conducted previously. Therefore, we have analyzed the plastomes of nine cycad species, including representatives from all three families and nine of ten genera. The cycad plastomes are highly conserved in genome architecture (gene order) and nucleotide composition. To evaluate the effects of plastomic structure on genomic stasis, we estimated mutation rates of synonymous (ds) sites and noncoding loci in single copy (SC) and inverted repeat (IR) regions, separately. We found that mutation rates between SC and IR regions differ significantly and show greater correlation in noncoding loci ($r=0.831$, $p=0.011$) than in ds sites ($r=0.578$, $p=0.133$), implying that plastomic structural effects are more influential on noncoding loci than on ds sites. Our analysis of equilibrium GC content (GC_{eq}) further indicated that noncoding loci are biased toward AT richness in SC regions, but toward GC richness in IR regions. These contrasting trends in GC_{eq} between SC and IR regions strongly suggest the existence of biased gene conversion toward GC (BGC_{GC}) in cycad plastomes. Additionally, in both SC and IR regions mutational rates and GC_{eq} are negatively correlated, indicating that BGC_{GC} is able to correct plastome-wide mutations in cycads. We therefore propose that BGC_{GC} might counteract the effect of spontaneous AT-biased mutations during the evolution of cycad plastomic GC content.

568C

Molecular mechanisms underlying the phenotypic convergence of inflorescence architecture in domesticated rice species

Thomas Harrop [1], Veronica Gregis[2], Israr Ud Din[2], Martin Kater[2], Stefan Jouannic[1], Hélène Adam[1]
[1] IRD (Institut de Recherche pour le Développement), Montpellier, France, [2] Dipartimento di Bioscienze, Universita degli Studi di Milano, Milano, Italy

During the independent domestications of African and Asian rice, artificial selection for increased yield resulted in convergent evolution of properties of inflorescence architecture that affect grain output, such as overall size and branching complexity. However, little is known about the molecular mechanisms involved in the phenotypic convergence in the domesticated species. In the developing inflorescence a switch occurs from the production of lateral branches in the indeterminate phase to the production of spikelets, which develop into flowers in the mature inflorescence, in the determinate phase. It is likely that the timing of this switch is related to the branching complexity of the inflorescence and hence grain yield.

To investigate this we are studying five accessions of rice, including the *japonica* and *indica* varieties of Asian rice and domesticated African rice, along with their wild relatives. Phenotyping results indicate that the accessions are different in terms of grain yield and inflorescence architecture. We are currently using strand-specific, whole-transcriptome RNA sequencing of indeterminate and determinate stages of young inflorescences to compare gene expression and regulatory networks and study the conserved and disparate mechanisms that effect the phenotypic variation between species in the context of plant domestication.

569D

Patterns of Evolution of MHC Class II Genes of Crows (Corvus)Suggest Trans-species Polymorphism

John Eimes [1,2], Irem Sepil[5], Yoko Satta[2], Andrea Towsend[3], Isao Nishiumi[4]
[1] Seoul National University, Seoul, Republic of Korea, [2] Graduate Universty for Advanced Studies, Hayama, Japan, [3] Hamilton College, New York, USA, [4] National Museum of Nature and Science, Tsukuba, Japan, [5] University of Oxford, Oxford, UK

A distinguishing characteristic of genes that code for MHC is that alleles often share more similarity between, rather than within species. Two likely mechanisms explain this pattern: convergent evolution and trans-species polymorphism (TSP). Distinguishing between these mechanisms has major implications in how we view adaptation of immune genes. In this study we analyzed exon 2 of the MHC class IIB in three passerine bird species in the genus *Corvus*: jungle crows (*Corvus macrorhynchos japonensis*) American crows (*C. brachyrhynchos*) and carrion crows (*C. corone orientalis*). Carrion crows and American crows are recently diverged, but allopatric, sister species, whereas carrion crows and jungle crows are more distantly related but sympatric species, and possibly share pathogens linked to MHC IIB polymorphisms. These patterns of evolutionary divergence and current geographic ranges enabled us to test for trans-species polymorphism and convergent evolution of the MHC IIB in crows. Phylogenetic reconstructions of MHC IIB sequences revealed several well supported interspecific clusters containing all three species, and there was no biased clustering of variants among the sympatric carrion crows and jungle crows. The topologies of phylogenetic trees constructed from putatively selected sites were remarkably different than those constructed from putatively neutral sites. In addition, trees constructed non-synonymous substitutions from a continuous fragment of exon 2 had more, and generally more inclusive, supported interspecific MHC IIB variant clusters than those constructed from the same fragment using synonymous substitutions. These phylogenetic patterns suggest that recombination, especially gene conversion, has partially erased the signal of allelic ancestry in these species.

570A

Rethinking questions about evolution

David Penny
Massey University, Palmerston North, New Zealand

With the rise of relatively rapid DNA-sequencing, and consequently of whole genomes, they are increasingly becoming standard. However, we need to be much more careful about our understanding of the role of key biological processes during evolution. For example, in a recent study of 48 bird genomes the authors undertake a good phylogeny, but then just accept an old archaic story about modern birds arising only after the extinction of dinosaurs, despite significant evidence to contradict this. There is extensive evidence for birds beginning to radiate earlier, and that the smaller dinosaurs were also declining. Some of the issues we face are important evolutionary questions, such as whether key biological processes have improved during evolution. Why do we get a continued turnover of larger organisms? We need to know better the role of mutation in the continued rise of new combinations, is there really a population size effect. However, we are going to have to (unfortunately) think - there is no easy answer. So perhaps we are getting to a new set of questions as we obtain whole genomes. The rise if new sequencing methods mean that it is now relatively routine to sequence entire genomes, but the real underlying evolutionary questions may not always be so obvious

571B

The role of the Major Histocompatibility Complex and genome-wide relatedness on cryptic female choice in Chinook salmon (*Oncorhynchus tshawytscha*)

Cornelia Gessner [1], Patrice Rosengrave[1], Shinichi Nakagawa[2], Neil J. Gemmell[1]
[1] Allan Wilson Centre, Department of Anatomy, University of Otago, Dunedin, New Zealand, [2] Department of Zoology, University of Otago, Dunedin, New Zealand

Cryptic female choice (CFC) is a post-copulatory process that enables females to bias fertilization outcomes towards certain males. In Chinook salmon, an external fertilizer, the ovarian fluid surrounding the ova differently enhances the sperm velocity of males. Sperm velocity is a key ejaculate trait that determines fertilization success in externally fertilizing fishes, thus its differential enhancement might bias male fertilization outcomes and represent a mechanism of CFC.

Another mechanism of CFC might exist at the egg surface, whereby sperm-egg interactions could bias fertilization outcomes.

Here, we explored the potential genetic basis of both possible mechanisms by examining whether i) the genotypic combinations of females and males at the Major Histocompatibility Complex (MHC) class I and II and ii) the genome-wide relatedness of mates, assessed with a 6000 Chinook salmon SNP chip, explain the variation in sperm velocity and/or male fertilization success.

We employed paired-male fertilization trials with males exhibiting different sperm velocities in the focal female. Male fertilization success was evaluated using microsatellite based paternity assignment.

We showed that relative sperm velocity was positively correlated with fertilization success, confirming that the differential enhancement of sperm velocity may be a mechanism of CFC in Chinook salmon. The variation in sperm velocity was MHC-independent, but explained by relatedness between mates on two SNP Linkage Groups, indicating the presence of genes involved in sperm-ovarian fluid interactions in these regions. Besides relative sperm velocity, the MHC class II similarity of mates is correlated with fertilization success, indicating that this locus might influence sperm-egg interactions.

572C

Balancing selection at a polymorphic colour locus in Gouldian Finches

Ben Jackson, Kang-Wook Kim, Terrence Burke, Kai Zeng
University of Sheffield, Sheffield, UK

We investigate the population genetics of a locus underlying a striking, naturally occurring colour polymorphism in the Gouldian Finch, Erythrura gouldiae. Coalescent simulations and a modified implementation of the HKA test show that patterns of nucleotide diversity at the locus of interest are incompatible with the neutral model of molecular evolution, which suggests a role for balancing selection in maintaining colour polymorphism in these birds. In order to reach this conclusion we account for non-random sampling at our focal locus when trying to i) infer key population genetic parameters (e.g., the recombination rate), and ii) detect signatures of selection. Our results demonstrate the importance of correctly accounting for sampling bias before making conclusions about the evolutionary history of functionally important parts of the genome.

573D

Perturbation of Iron Homeostasis Promotes the Evolution of Antibiotic Resistance

Orsolya Méhi, Balázs Bogos, Bálint Csörgő, Ferenc Pál, Ákos Nyerges, Balázs Papp, Csaba Pál
Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary

Antibiotic resistance is an increasingly urgent, present-day medical issue, that frequently develops in microbes via the acquisition of spontaneous mutations during antimicrobial therapy. Here, we demonstrate that inactivation of a central transcriptional regulator of iron homeostasis (Fur) facilitates laboratory evolution of ciprofloxacin resistance in *Escherichia coli*. As a first step in elucidating the underlying molecular mechanisms we performed a global transcriptome analysis and demonstrated that the set of genes regulated by Fur changes substantially in response to antibiotic treatment. We hypothesized that the impact of Fur on evolvability under antibiotic pressure is due to the elevated intracellular concentration of free iron and the consequent enhancement of oxidative damage-induced mutagenesis. In agreement with expectations, overexpression of iron storage proteins, inhibition of iron transport, or anaerobic conditions drastically suppressed the evolution of resistance, whereas inhibition of the SOS response-mediated mutagenesis had only a minor effect. By applying a cell permeable iron chelator the evolution of resistance could be drastically reduced, enforcing our results. In sum, our work revealed that iron metabolism represents an important factor in the evolution of antibiotic resistance, a pattern that could influence the development of novel antimicrobial strategies.

574A

Gynodioecy and mitochondrial sequence evolution in the ribwort and Buck's-horn plantain (*Plantago*)

Nicholas Levsen [1], Deborah Charlesworth[2], Kirsten Wolff[1]
[1] *Newcastle University, Newcastle Upon Tyne, UK,* [2] *The University of Edinburgh, Edinburgh, UK*

Gynodioecy is a sex phenotype polymorphism in which individuals are either hermaphroditic or female (male sterile). Male sterility is caused by disruptive mutations in the nuclear or cytoplasmic genomes, whereas male fertility may be recovered via the inheritance of nuclear restorer alleles. Cytoplasmic male sterility (CMS) perpetuates under a model of inter-genomic conflict that has implications for mitochondrial and nuclear DNA sequence evolution. For two congeneric species of flowering plants, we investigate genetic evidence of mitochondrial recombination, heteroplasmy, and long-term balancing selection.

575B

Molecular phylogeny of Trias (Bulbophyllinae, Orchidaceae) based on ITS, rbcL and matK

Passorn Wonnapinij, Ajaraporn Sriboonlert
Kasetsart University, Bangkok, Thailand

Trias, a unique epiphytic orchid with distinct triangle-shaped flowers. The habitat of this orchid genus is limited to Indo-China, India, and Australia. The phylogenetic relationship among Trias species and their related taxa Bulbophyllum has not been revealed mainly due to scarcity of the samples. This orchid genus had been classified in subtribe Bulbophyllinae; however, recent evidences suggested that this orchid should be placed as one section in the genus Bulbophyllum. In this study, we aimed to assess the relationship of Trias and Bulbophyllum and investigate interspecific relationships of this group of orchid using DNA sequences from two plastid genes, ribulose 1,5-bisphosphate carboxylase (rbcL) and maturase K (matK) and one nuclear internal transcribe spacer (ITS) region. All phylogenetic trees constructed in this study showed that all Trias species were embedded in Bulbophyllum supporting the notion that the Trias should be included in the Bulbophyllum. All trees also revealed that the Indo-China Trias form a monophyletic group. These Trias species could be divided into four clades that coincide with their vegetative appearances. This study further shows that, as an individual marker, nuclear ITS region sequences can be sufficient for inferring phylogenetic relationship among Trias species; though this marker and the combination of this marker with plastid genes is not able to distinguish the differences between closely related Trias species. The results of this study reveals the relationship between Trais and Bulbophyllum as well as the relationship among some Trias species.

576C

Simulating realistic multiple sequence alignments

Avigayel Hefter, Eli Levy Karin, Haim Ashkenazy, Oren Avram, Tal Pupko
Tel Aviv University, Tel Aviv, Israel

A wide variety of analyses for the detection of important biological phenomena rely on a multiple sequence alignment (MSA) as input. Such analyses include, for example, phylogenetic tree reconstruction, the inference of selective forces, and the identification of traits affecting the rate of evolution. The accuracy and power of such inference methods are often evaluated through simulation studies. Moreover, parametric bootstrap methods use simulations as part of their inference procedure. Ideally, sequence simulators should generate MSAs that resemble the characteristics of the original real MSA in terms of number and length of indels. However, to date, there is no available methodology for inferring relevant parameters regarding indel dynamics, and thus, simulated MSAs often differ drastically in their indel characteristics from real MSAs. In this study we present a novel methodology to infer indel dynamics parameters from real MSAs. Using simulations, we demonstrate that our methodology can accurately infer the indel dynamics parameters for a large variety of plausible settings. We next study the impact of using different alignment programs on the ability to infer those parameters. Finally, we apply our methodology to study the distribution of those parameters in two genomics datasets, one from mammalian sequences and one from the COG database. In summary, our methodology offers a novel approach towards realistic simulations of sequence data, and should thus impact a wide range of phylogenetic analyses.

577D

Molecular evolution of *TOLL* immune genes in *Anopheles dirus* complex mosquitoes

Uraiwan Arunyawat [1], Prin Phunngam[1], Wasitthee Kongkachana[1], Theeraphap Chareonviriyaphap[2]
[1] Evolutionary Genetics and Computational Biology Research Unit, Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand, [2] Department of Entomology, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand

Understanding evolutionary forces that effect malaria parasite transmission in mosquito vectors is important for the vector-borne diseases control. In this study, Toll-like receptor gene family (*Tolls*) involving in signal transduction pathways of insect immune system was selected to estimate the pattern of genetic variability in *Anopheles dirus* and *An. baimaii*, the main malaria vectors in Southeast Asia. Five partial *Toll* immune genes were analyzed using single nucleotide polymorphism data to estimate levels of genetic variation in the mosquito populations. The results showed low levels of nucleotide variation in *Toll6* and *Toll8* sequences, whereas the other *Toll* loci showed intermediate level of sequence variation. Statistical Neutrality tests were used to measure deviation from neutral equilibrium expectation. No significant neutrality tests were detected, implying that mutation and random drift interaction is most likely the cause of genetic variation of the studied *Toll* genes rather than the effect of natural selection. Moreover, Ka/Ks ratio test supported that selection force may not evolve in the variation of the *Toll* immune genes in the mosquito vectors.

578A

Molecular insights into the complex polymorphic inversion system of the E chromosome of *D. subobscura*

Eva M. Puerma, Dorcas J. Orengo, Montserrat Papaceit, Carmen Segarra, Montserrat Aguadé
Departament de Genètica, Facultat de Biologia, i Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona., Barcelona, Spain

Classical cytological studies of chromosomal inversion polymorphism in the Drosophila genus had revealed that inversions are not evenly distributed either across species or among their chromosomal elements. The cytological evidence for heterogeneity extends to the location of breakpoints as revealed by the differential clustering exhibited by some breakpoints, with some being shared by two or more inversions. In order to extend our knowledge about inversion originating mechanisms as well as to assess whether cytologically shared breakpoints are actually reused at the molecular level, we have identified and sequenced the breakpoints of four polymorphic inversions of the E chromosome of *Drosophila subobscura* (E_1, E_2, E_9 and E_3) that occurred sequentially and share some of their breakpoints at the cytological level. Sequence comparison of breakpoint regions has allowed establishing that three of the four studied inversions originated through the staggered-breaks mechanism. They have also revealed the multiple reuse at the molecular level of the most proximal breakpoint at section 58D, which has been used at least by three of the four inversions. In contrast, the breakpoint at section 64C might have been used by inversions E1 and E2.

579B

An Examination of the Effectiveness of Several Ancient DNA Extraction Methods

Austin Reynolds, Lauren Springs, Jaime Mata-Míguez, Deborah Bolnick
University of Texas at Austin, Austin, Texas, USA

Over the past ten years, several methods of extracting endogenous DNA from ancient bones and teeth have been reported in the literature and widely adopted. However, a direct comparison of these methods has not yet been published. In this study, we compared four of these extraction methods and evaluated how much endogenous DNA each recovered from a set of forensic and ancient skeletal samples using qPCR. We also applied a recently published pre-digestion method to each of the extraction methods. This technique has been shown to greatly increase the amount of endogenous DNA recovered from ancient bone and teeth samples. Here we explored the effects of the different extraction methods on the pre-digestion's effectiveness. We performed each of the extraction methods on the same bone or tooth from an individual to reduce the effects of differential preservation on our results. The extractions were performed using both ancient (200-4000 years old) and forensic bone samples from several geographic regions in the Americas. The results of this study will help determine which of the recently developed and commonly used extraction methods in the field of ancient DNA are most effective at recovering endogenous DNA in these temperate and subtropical burial contexts. This will help our lab and others choose the best extraction method and may help reduce inter-laboratory variability in this area.

580C

Evolutionary significance of stochasticity in gene expression

Katsuhiko Mineta [1], Tomotaka Matsumoto[2], Naoki Osada[3], Hitoshi Araki[3]
[1] King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2] National Institute of Genetics, Shizuoka, Japan, [3] Hokkaido University, Hokkaido, Japan

The role of stochasticity in evolutionary genetics has long been debated. To date, however, the potential roles of non-genetic traits in evolutionary processes have been largely neglected. In molecular biology, growing evidence suggests that stochasticity in gene expression (SGE) is common and that SGE has major impacts on phenotypes and fitness. Here, we provide a general overview of the potential effects of SGE on population genetic parameters, arguing that SGE can indeed have a profound effect on evolutionary processes. Our analyses suggest that SGE potentially alters the fate of mutations by influencing effective population size and fixation probability. In addition, a genetic control of SGE magnitude could evolve under certain conditions, if the fitness of the less-fit individual increases due to SGE and environmental fluctuation. Although empirical evidence for our arguments is yet to come, methodological developments for precisely measuring SGE in living organisms will further advance our understanding of SGE-driven evolution.

581D

A mitogenomic view on ancient intercontinental dispersal in gray wolves (*Canis lupus*)

Stephan Koblmüller [1,2], Carles Vilà[2], Maria B. Lorente Galdos[3], Marc Dabad[3], Oscar Ramirez[3], Tomas Marques-Bonet[3,4], Robert K. Wayne[5], Jennifer A. Leonard[2]
[1] Institute of Zoology, Karl-Franzens-University Graz, 8010 Graz, Austria, [2] Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), 41092 Sevilla, Spain, [3] Institut de Biologia Evolutiva (CSIC-UPF), 08003 Barcelona, Spain, [4] Centro Nacional de Analisis Genomico (CNAG, PCB), 08024 Barcelona, Spain, [5] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-1606, USA

Gray wolves (*Canis lupus*) have a very wide native distribution across the entire Holarctic. The fossil record indicates that the species evolved in Eurasia in the early Pleistocene, and then colonized North America in the mid Pleistocene. Previous phylogeographic studies found polyphyly of North American wolves within the diversity of Eurasian wolves with mitochondrial markers, but the support on deep branches was low and genomic data has suggested monophyly of the North American wolves. Here we analyze 105 whole mitochondrial genomes from the main clade of gray wolves within an approximate Bayesian computation framework to test for the number of times wolves colonized North America from Eurasia, and date colonization(s). We find that the mitogenomes of all living wolves in North America, including Mexican wolves, derive from a single colonization event from Eurasia that expanded its range into southern North America before the Cordillerian and Laurentide ice sheets fused in the Last Glacial Maximum, approximately 23KYA. This is more recent than expected based on the fossil record, suggesting that there were earlier colonizations that left no descendents.

582A

Fast and scalable genome-wide SNP genotyping (ddRADseq-ion) on Ion Torrent next-generation sequencing platforms

Kathryn Elmer, Hans Recknagel, Arne Jacobs, Pawel Herzyk
University of Glasgow, Glasgow, UK

Research in evolutionary biology involving non-model organisms is rapidly shifting from using locus-specific molecular markers such as mtDNA and microsatellites to higher throughput and genome-wide SNP genotyping to address questions in population genetics, phylogenetics and genetic mapping. Single-digest and double-digest restriction-site associated DNA sequencing (RADSeq and ddRADSeq) has become an established method for SNP genotyping on Illumina sequencing platforms without requiring existing genomic resources. We developed and optimized a molecular and bioinformatic protocol for conducting double-digest RAD Sequencing on the Ion Torrent (Life Technologies; Ion Proton, Ion PGM) semiconductor sequencing platform (ddRADseq-ion). We sequenced genomic libraries of multiple non-model vertebrate species with representative, complex genomes: common lizards and salmonid fishes. On average, we obtained ~11 000 polymorphic loci per library of six to 30 individuals on Ion Proton with PI chips. Data can be analyzed in existing RADseq pipelines such as STACKS; additionally we developed a script tool to maximize the amount of usable sequence data. We validated our new approach by technical and biological replication, by reconstructing phylogenetic relationships of the salmonid fishes Arctic charr and European whitefish, and by using a hybrid genetic cross of European whitefish to track genomic variants. The ddRADseq-ion method is ideal for optimizing and conducting pilot projects because it is very scalable. We show that our protocol can be used for the rapid, robust and cost effective generation of variable and reproducible genetic markers on Ion semiconductor sequencing platforms.

583B

First Complete Mitochondrial Genome from a Parasitic Plant (*Castilleja paramensis*)

Weishu Fan [1], Kanika Jain[1], Melisa Kozaczek[1], Neethu Shaw[1], Natalia Pabón-Mora[2], Favio González[3], Jeffrey Mower[1]

[1] University of Nebraska-Lincoln, Lincoln, Nebraska, USA, [2] Universidad de Antioquia, Medellín, Antioquia, Colombia, [3] Universidad Nacional de Colombia, Bogotá, Colombia

Parasitic plants obtain part or all of their nutrition from their host plant, resulting in the reduction or loss of photosynthetic activity and genes related to photosynthesis in the plastid genome. *Castilleja paramensis* is a perennial herbaceous hemiparasite in the Northern high-Andean Mountains of Colombia. Here we report the complete mitochondrial and plastid genome of *C. paramensis* to investigate changes in the organelle genomes that may result from a parasitic lifestyle. Comparative plastid genome analyses indicate that *C. paramensis* retains nearly all conserved plastid genes, except that *ndhF* has become a pseudogene. This suggests that the NAD(P)H dehydrogenase complex is in the initial stages of degradation, consistent with observations in other hemiparasitic plants. In contrast, the mitochondrial genome, which is the first completed sequence from a parasitic plant, contains a typical set of genes and no evidence for genome degradation. Parasitic plants can facilitate horizontal transfer, and *C. paramensis* appears to be the donor for a transfer event involving the *atp9* and *ccmFn* genes into another Andean plant. The mitochondrial genome also contains the *cox1* intron, which was previously observed to be over-represented in parasitic plants and horizontally transferred among species; however, phylogenetic analysis of the intron indicates that Orobanchaceae parasites have acquired this intron vertically from a non-parasitic ancestor. Collectively, these results are consistent with a reduction in photosynthetic activity but retention of full mitochondria function in *C. paramensis*.

584C

LD-kNNi - a new, fast and accurate genotype imputation method for non-model organisms

Daniel Money, Kyle Gardner, Sean Myles
Dalhousie University, Truro, Nova Scotia, Canada

Next-generation DNA sequencing (NGS) is increasingly being used to investigate questions of evolutionary importance across diverse organisms, but bioinformatic tools that exploit the full potential of NGS are lacking. Genotype imputation is a powerful method for exploiting the full potential of NGS by filling in missing genotype data, but most imputation tools have been developed for human whole-genome NGS data and benefit from high quality genotype calls from a panel of reference samples. In studies of non-model organisms, methods such as RADSeq and genotyping-by-sequencing (GBS) are often used to generate genome-wide polymorphism data from NGS data and it remains unclear what the most suitable imputation algorithms are for these data. Here we introduce a modified k Nearest Neighbour algorithm that makes use of linkage disequilibrium (LD) information. As this algorithm uses LD information it does not require a reference genome. We compare our method with several genotype imputation methods on GBS data generated from over 700 varieties of apple (Malus domestica), a highly diverse and heterozygous crop without a reference panel of high quality genotype data. We find that our algorithm performs similarly to those that require a reference genome, while out performing those that do not. We report metrics of quality, accuracy and utility for each imputation algorithm and provide recommendations for studies using similar methods in non-model organisms. Our method provides an important step in exploiting the full potential of NGS for crop improvement, especially in species for which genomics resources are underdeveloped or non-existent.

585D

ANGSD-wrapper: scripts to streamline and visualize NGS population genetics analysis

Arun Durvasula, Tyler Kent, Siddharth Bhadra-Lobo, Jeffrey Ross-Ibarra
University of California, Davis, Davis, California, USA

The advent of highly multiplexed sequencing has opened a number of exciting avenues for evolutionary biologists. One of the powerful approaches enabled by inexpensive sequencing is the ability to sequence a large number of individuals, each to relatively low sequencing depth. However, this approach also presents statistical challenges in the analysis of low coverage data. The software ANGSD [1] and related programs [2] were developed to deal with low coverage sequence data. Rather than call genotypes at variable sites, ANGSD performs a number of population genetic analyses on genotype likelihoods, including estimation of the population mutation rate θ, the site frequency spectrum, neutrality tests, inbreeding coefficients, and population structure. ANGSD has already been used in several studies to analyze genome sequence data [3] [4]. However, ANGSD requires considerable familiarity with command line tools and remains inaccessible to many biologists that are not from a computational background. Here we present a software package that aids in the preparation of analy- ses for ANGSD and provides interactive graphing software implemented in R [5] and Shiny [6]. ANGSD-wrapper simplifies multistep analyses such as calculating Tajima's D into a single step. Users supply all the needed in- formation in a single configuration file (Figure 1), and after ANGSD has fin- ished calculations, ANGSD-wrapper provides interactive graphing of the results (Figure 2). ANGSD-wrapper is available on github: https://github.com/ arundurvasula/angsd-wrapper.

586A

Simulating genome evolution in the house mouse: understanding the contribution of Hill-Robertson interference to patterns of genetic diversity

Tom Booker, Peter Keightley
University of Edinburgh, Edinburgh, UK

Understanding the contribution of positive and negative selection to patterns of genetic diversity remains an area of active research. Classically the correlation between genetic diversity and recombination, observed in many taxa, has been interpreted in the context of positive selection exerting its effects on neutral diversity across the genome through Hill-Robertson interference. Increasingly, however, studies suggest a substantial role for Hill-Robertson interference acting through negative selection (background selection) in generating patterns of neutral diversity across the genome. This pattern is apparent in the genome of the house mouse, where there is an observable reduction in nucleotide diversity near exons. Using parameters extracted from population genomic data for *Mus musculus castaneus* we simulated the evolution of the mouse genome using forward-time population genetic simulations with a realistic chromosome structure. Using estimates of the distribution of fitness effects for deleterious mutations (DFE), we ask whether we can explain the patterns of genetic diversity around exons by background selection alone. Furthermore, we ask whether the DFE estimated from the population data matches that of the simulated populations. If the DFE and patterns of genetic diversity estimated from our simulated populations are consistent with those obtained empirically we may be able to infer the mode of selection that has affected the evolution of exons in the mouse genome. The questions these analyses seek to answer are of fundamental importance for our understanding of the evolution of patterns of genetic diversity.

587B

Genomic architecture of flight capacity in gypsy moths

Nathan Havill [1], Melody Keena[1], Andrea Gloria Soria[2], Adalgisa Caccone[2]
[1] *USDA Forest Service, Hamden, CT, USA,* [2] *Yale University, New Haven, CT, USA*

The gypsy moth, *Lymantria dispar* (Insecta: Lepidoptera: Erebidae), is a forest pest that was introduced to the United States from France over 100 years ago. Most females of the European subspecies of gypsy moth (*L. dispar dispar*) are not capable of flight, but females of the two Asian subspecies (*L. dispar asiatica* and *japonica*) can fly. We report progress towards identifying genomic regions associated with flight capacity in this species. We conducted reciprocal crosses between a flightless European strain and six different flight-capable strains from Russia, China, and Japan. Female flight propensity, flight capability, and flight muscle strength were scored for parental, F1, and F2 individuals. Whole genome SNP discovery is in progress using double digestion Restriction-site Associated DNA (ddRAD) sequencing for parental and F2 individuals. A linkage map will be produced and QTL analysis will be performed to establish the genomic regions associated with flight capacity. The results will provide insight into the genomic architecture of complex behavioral traits in insects and allow us to develop cost effective genotyping methods to detect introduction of flight-capable gypsy moths in North America and Europe.

588C

Tissue-specific gene repression ("disallowed genes") in birds

Stefanie De Coster, Katleen Lemaire, Leentje Van Lommel, Mikaela Granvik, Lieven Thorrez, Katrijn Van Deun, Frans Schuit
KU Leuven, Leuven, Belgium

Housekeeping genes are thought to contribute to maintenance of integrity of different cell types in an organism. Some housekeeping genes have a constant expression level so that they can be used as reference genes for normalization. But other housekeeping genes are strongly repressed in one tissue only. Using a genome-wide mRNA expression analysis, we detected such tissue-specifically repressed genes in mice and proposed that this repression contributes to a differentiated tissue phenotype. In the present study we aimed to examine this novel phenomenon in a non-mammalian vertebrate. We performed genome-wide mRNA expression analysis in 21 tissues from the ISA-Brown chicken (Gallus gallus) and used the intersection-union test to identify mRNA transcripts that were significantly repressed in one tissue only. The largest number of tissue-specifically repressed genes was found in chicken liver (63 genes). Interestingly, one of these repressed genes encodes 2-oxoacid CoA transferase (OXCT1, an enzyme which degrades ketone bodies and which is also liver-specifically repressed in mice). By performing chromatin-immunoprecipitation experiments a similar signature of H3K27-trimethylation in liver DNA of both species was observed. In chickens, the gene encoding 3-hydroxybutyrate dehydrogenase, type 1 (BDH1) was also liver-specifically repressed, whereas in mice this gene was liver-specific up-regulated. Our data indicate that the phenomenon of "disallowed genes" is present in two vertebrate groups (mammals and birds) which has implications for evolutionary origin. Moreover, when comparing mouse and chicken, conserved repression of OXCT1 but different gene regulation of BDH1 seems interesting to further compare the functional genomics of ketogenesis in both species.

589D

Wolbachia interference on reproductive fitness of Culex quinquefasciatus by apoptosis and follicular atresia

Fábio Almeida [1], Lincoln Suesdek[1,2]
[1] *Instituto Butantan, São Paulo, Brazil,* [2] *Instituto de Medicina Tropical, São Paulo, Brazil*

Wolbachia pipientis are bacteria capable to infect insects and nematodes and they are one of the most abundant intracellular bacteria on Earth. These bacteria are transmitted maternally and are able to manipulate reproductive traits of their hosts, causing male-killing, feminization of males, induction of parthenogenesis, alterations on reproductive fitness, and cytoplasmic incompatibility (crossing incompatibility) and all of these alterations confers advantages to infected females and reflect on evolution of their hosts. One of their hosts is Culex quinquefasciatus, a house-common mosquito, which is able to transmit some human pathogens such as West Nile Virus and Wulchereria bancrofti. The Wolbachia infection causes, on Culex, cytoplasmic incompatibility and some reproductive fitness disorder, among which we can mention reduction on eggs quantity. This work is an investigation about how these bacteria decrease the number of eggs from Culex and we had some indications that Wolbachia can induce apoptosis and follicular atresia on cells of the ovarian follicle. To test that possibility we investigate, by counting the number of apoptotic cells by confocal microscopy, the occurrence of apoptosis on several times during the pre-vitellogenic phase and vitellogenic phase of adult infected females mosquitoes, compared to artificially uninfected ones. So far we have observed that occurs statistically more apoptosis in infected mosquitoes ovarian cells and follicular atresia was observed only in infected mosquitoes. So far we concluded that the follicular atresia and apoptosis are factors that contribute to eggs quantity reduction and influence the evolution of this species

590A

Shrimps of the genus *Hippolyte* (Decapoda: Caridea) from Indo-Pacific: multigene analysis reveal new findings on taxonomy.

Mariana Terossi, Fernando Luis Mantelatto
Laboratory of Bioecology and Crustacean Systematics (LBSC), Faculty of Philosophy, Sciences and Letters of Ribeirão Preto (FFCLRP), University of São Paulo (USP), Brazil. FAPESP PD 2011/11901-3; BIOTA, Riberirão Preto, São Paulo, Brazil

The genus *Hippolyte* comprises marine shrimp species with worldwide distribution, except in extremely cold waters. There is little differentiation between the *Hippolyte* species and large variation within species. Particularly to Indo-Pacific, the systematic of the genus is chaotic. One example is *Hippolyte ventricosa*, which has been previously recorded in various parts of the Indo-Pacific Ocean, but after the redescription made by Udekem d´Acoz (1999), *H. ventricosa sensu strictu* is known only from India and Pakistan. Currently, the species from Indo-Pacific that have one distal outer tooth in the first segment of the antennular peduncle and do not match with the descriptions of the known species (*H. australiensis*, *H. edmondsoni*, *H. jarvinensis* and *H. ventricosa*) are put together in the *H. ventricosa* group. Here, we analyzed morphological and molecular multi data of four specimens of this group in order to check the taxonomic validity. The molecular data were obtained through protocols of DNA extraction, amplification (nuclear genes 18S and Histone 3, mitochondrial genes 16S and Cytochrome Oxidase I), purification and sequencing. The genetic divergence analyses and phylogenetic tree corroborated that these specimens not belong to the known species previously cited. We hypothesized the existence of three new species among these group, two from Indonesia, and one from Singapore and Vietnam, showing the necessity of a taxonomic rearrangement in. the genus *Hippolyte*.

591B

BAYESIAN INFERENCES OF THE DEMOGRAPHIC HISTORY OF PERUVIAN NATIVE POPULATIONS

Marilia Scliar [1], Rennan Moreira[1], Wagner Magalhaes[1], Lilia Cabrera[2], Douglas Berg[3], Robert Gilman[4,5], Eduardo Tarazona-Santos[1]

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, [2] Asociación Benéfica PRISMA, Lima, Peru, [3] University of California San Diego, San Diego, USA, [4] Johns Hopkins University, Baltimore, USA, [5] Universidade Peruana Cayetano Heredia, Lima, Peru

Archeology reports millenary cultural contacts between Peruvian Coast-Andes and the Amazon Yunga, a rainforest transitional region between Andes and Lower Amazonia. In a previous study, we used Bayesian inferences using the Isolation with Migration model and resequencing data to clarify the demographic relation between two Native Peruvian populations from these regions: Quechuas from Central-Andes and Shimaas from Lower Amazon Yunga. We inferred that the Shimaa ancestors were a small subgroup that split <5300 years BP (after the development of complex societies) from an ancestral Andean population, moved toward the Amazon Yunga, and incorporated the culture and language from their neighbors, but not their genes. While Andean populations are highly homogeneous, further studies on other populations of the Amazon Yunga are necessary to show if the predominant Andean biological origin of the Shimaas is the rule, and not the exception. To this, we have generated ~45 kb of resequencing data (more than twice of our previous study) for 16 Quechuas, 16 Shimaas and 16 Ashaninkas (another population from the Peruvian Amazon Yunga). We are using the Approximate Bayesian Computation methodology to infer demographic parameters, such population effective sizes, time of divergence, and migration rates, using models including the three populations.

592C

Investigation of mtDNA heteroplasmy in *mtCR* and *mtCOI* of *Portunus pelagicus*

Chidchanok Intakham, Pavitchaya Koolkarnkhai, Pradit Sangthong, Wunrada Surat, Passorn Wonnapinij
Department of Genetics, Faculty of Science, Kasetsart University, Bangkok, Thailand

Based on the assumption that an individual carries only one type of mitochondrial DNA (mtDNA), sequences of mitochondrial control region (*mtCR*) and mitochondrial cytochrome oxidase I (*mtCOI*) have been widely used as markers for molecular ecological, evolutionary and population genetic analyses. However a mixture of two or more types of mtDNA in an individual organism, the mtDNA heteroplasmy, has been reported in many species. This feature could potentially affect the analyses based on mtDNA sequence and basic knowledge regarding this feature would be required. This study aimed to reveal mtDNA heteroplasmy in *mtCR* and *mtCOI* of *Portunus pelagicus* and investigate their inheritance pattern. By screening through 14 mother-offspring pairs using PCR and Sanger sequencing, we observed that two pairs potentially carried mtDNA heteroplasmy. 82 polymorphic sites including 63 heteroplasmic sites were observed in *mtCR*, while only14 polymorphic sites with no heteroplasmic sites were observed in *mtCOI*. Taking into account only the mothers, nucleotide diversity of *mtCR* and *mtCOI* were 0.0332 and 0.0033, respectively. Interestingly, the comparison of *mtCR* sequences between mother and her pooled offspring showed percent identity in the range of 94% to 100 %. Hence, these results supported that sequence variation in *mtCR* is greater than in *mtCOI.* Furthermore, the presence of mtDNA heteroplasmy in *mtCR* would reduce the efficiency of using it as a marker for identifying maternal relatives. Due to the detection limits of Sanger sequencing, next-generation sequencing technology will be applied to confirm these results and disclose detail of mtDNA heteroplasmy inheritance.

593D

Photoperiodic adaptation and genomewide patterns of population genetic variation in North European *Arabidopsis lyrata* populations

Tiina Mattila [1], Jaakko Tyrmi[1,2], Outi Savolainen[1,2]
[1] *University of Oulu, Oulu, Finland,* [2] *Biocenter Oulu, Oulu, Finland*

Arabidopsis lyrata is a small perennial plant species inhabiting diverse environments and an important model system for ecological adaptation. The different populations of the species are phenotypically and ecologically diverged and locally adapted. Laboratory studies have shown differences in photoperiodic reactions between populations. Thus we expect that photoperiod pathway genes may have been under directional selection as the population has colonized northern areas. We found that 19 flowering time pathway genes had low diversity, only 20 % of the variation found at a reference gene set. The flowering time pathway genes also had high population differentiation in Northern European *A. lyrata* populations. The strongest evidence for selection was detected in photoperiodic pathway genes. This may indicate selection for photoperiodically mediated flowering time differences during the colonization of Northern populations. However, during the postglacial colonization, the populations have also undergone complex demographic changes possibly causing similar patterns of variation as selection, complicating population genetics selection inference. Hence we aim to describe the detailed demographic history of five *A. lyrata* populations using genomewide resequencing data from six individuals from each population.

594A

Ancient mtDNA analysis of the Eneolithic Tripolye-Cucuteni Culture from Verteba Cave, Ukraine

Ryan Schmidt [1,2], Takashi Gakuhari[1,2], Hiromi Matsumae[3], Jordan Karsten[4], Mykhailo Sokhatsky[5], Hiroki Oota[1]

[1] Kitasato University School of Medicine, Sagamihara, Japan, [2] Japan Society for the Promotion of Science, Tokyo, Japan, [3] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [4] Department of Religious Studies and Anthropology, University of Wisconsin-Oshkosh, Oshkosh, WI, USA, [5] Borschiv Regional Museum, Bilche Zolote, Ukraine

Individuals associated with the Eneolithic Tripolye-Cucuteni Culture (4900 – 2750 calBC) practiced an agropastoral lifestyle. The oldest archaeological sites associated with this culture are found in the Carpathians, but spread to the plains and steppes of Moldova, Romania, and Ukraine. Archaeological evidence suggests the culture arose through the interaction of several Neolithic groups, though their origins are still unknown. Therefore, the Tripolye Culture existed at the intersection of indigenous European hunter-gatherer tribes and socially stratified Neolithic agricultural communities with possible origins in the Near East. Ancient DNA (aDNA) has demonstrated to be an important area of inquiry for investigating ancient groups, especially for peoples who contributed to modern Europe. In this study, we examine aDNA from Verteba Cave (3951 – 2620 calBC) in Ukraine to reconstruct the origins of this Eneolithic culture. Verteba Cave consists of commingled secondary burials. In our preliminary analysis, we sampled a number of left second metacarpal bones as a means of avoiding extracting DNA from the same individual. DNA was extracted from bone based on a modified silica method. Traditional PCR on the mtDNA HVRI was carried out using three sets of overlapping primers and sequenced on an ABI 3130 Genetic Analyzer. Next-generation libraries were then prepared for a subset of individuals and shotgun sequenced on an Illumina MiSeq platform. Our preliminary results indicate haplogroups that are common among modern Eurasian peoples. Future investigations will examine higher resolution mitogenome sequences and genome-wide SNP variation through target enrichment methods to characterize Tripolye population history.

595B

Genome Sequence Analysis of the Chimpanzee RH Blood Group Gene Cluster

Takashi Kitano [1], Antoine Blancher[2], Naruya Saitou[3]

[1] Ibaraki University, Hitachi, Japan, [2] Universite Paul Sabatier, Toulouse, France, [3] National Institute of Genetics, Mishima, Japan

On human chromosome 1, there are two duplicated Rh blood group loci (Hosa_RHD and Hosa_RHCE) that are closely linked but have opposite orientations. This duplication occurred in the common ancestor of humans, chimpanzees, and gorillas. Although several studies have been conducted on ape Rh blood group genes, clear genome structures of the genes remain unknown. Here, we determined the genome structure for the cluster of chimpanzee Rh blood group genes by sequencing five chimpanzee BAC clones. We found three full length loci (Patr_RHα, Patr_RHβ, and Patr_RHγ). In the Patr_RHβ locus, a short version of the gene (Patr_RHβS), which lacked the middle part containing exons 4-8, was observed. When we sequenced the region surrounding exon 7 of 12 chimpanzee individuals and performed a phylogenetic analysis, we also observed four major chimpanzee clusters (Patr_RHα, Patr_RHβ, Patr_RHγ, and Patr_RHδ). These results suggest that chimpanzees mainly have haplotypes with two Rh loci (Patr_RHα and Patr_RHγ) and four Rh loci (Patr_RHα, Patr_RHβ, Patr_RHγ, and Patr_RHδ). The two Rh loci haplotype contained a Patr_RHβS. Patr_RHα and Patr_RHβ were located on the corresponding locations of Hosa_RHD and Hosa_RHCE, respectively, and Patr_RHγ was in the immediate vicinity of 5' end of Patr_RHβ. Although phylogenetic networks implied a cluster of Patr_RHβ and Hosa_RHCE, Patr_RHα and Hosa_RHD did not form a cluster. Instead, Patr_RHγ formed a cluster with Hosa_RHD. The results suggest that rearrangements and gene conversions frequently occurred between these genes and that the classic orthology/paralogy dichotomy no longer holds between human and chimpanzee Rh blood group genes.

596C

The broken respiratory chain in Chromera velia - a phototrophic relative of apicomplexans.

Pavel Felgontov[1,2], Jan Michalek[1,3], Jan Janouskovec[4,5], De-Hua Lai[1], Milan Jirku[1,3], Eva Hajduskova[1], Ales Tomcala[1], Thomas D. Otto[6], Patrick J. Keeling[4,5], Arnab Pain[7], Miroslav Obornik[1,3], Julius Lukes[1,3]

[1] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Ceske Budejovice, Czech Republic, [2] Life Science Research Centre, Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic, [3] Faculty of Sciences, University of South Bohemia, Ceske Budejovice, Czech Republic, [4] Department of Botany, University of BC, Vancouver, Canada, [5] Canadian Institute for Advanced Research, Toronto, ON, Canada, [6] Wellcome Trust Sanger Institute, Hinxton, UK, [7] Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [8] Institute of Microbiology, Czech Academy of Sciences, Trebon, Czech Republic

The respiratory chain is the main metabolic feature of most mitochondria. In several cases of mitochondria and derived organelles, the sets of respiratory complexes were reduced or lost. Using experimental and bioinformatic approaches we have described a uniquely reduced mitochondrion of Chromera velia, a closest known phototrophic relative of apicomplexan parasites.
Mitochondrial genome of Chromera encodes only cox1 nad cox3 on linear molecules and respiratory complexes I and III are completely absent from the respiratory chain which is hereby split in two parts. The function of missing complex III appears to be supplemented by an alternative mechanisms engaging putative lactate:cytochrome c oxidoreductases. In contrast, the second known chromerid alga Vitrella brassicaformis posess respiratory chain similar to apicomplexans as well as mitochondrial-encoded cox1, cox3 and cob.

597D

Whole genome analyses of threespine stickleback populations from the Mediterranean to Northern Europe

Sanne Boessenkool [1], Bastiaan Star[1], Juha Merilå[2], Asbjørn Vøllestad[1]
[1] *University of Oslo, Oslo, Norway,* [2] *University of Helsinki, Helsinki, Finland*

Understanding how genetic variation in wild populations underlies evolutionary adaptation to different environments is a main goal of population genetics. The assessment of this variation has been revolutionized by the application of whole genome sequencing efforts and an increased availability of reference genomes. For instance, the genomic basis for several traits associated with transition from the marine to the freshwater environment has been determined for the threespine stickleback (*Gasterosteus aculeatus*), an ecological model system. Nevertheless, for European stickleback populations the genomic research has largely focused on northern and central populations, neglecting the more divergent southern populations, which possibly represent an older colonisation of the freshwater environment. We have sequenced the genomes of 40 threespine sticklebacks from a range of freshwater populations, covering a geographical range from the Caspian Sea to Northern Norway. Adding previously published genomes we analyse a dataset comprising 70 individuals, including one marine and several North American populations. Biogeographic patterns separate North American stickleback from two lineages distinguishing Atlantic European and Mediterranean populations. We further find clear evidence of founder events or population bottlenecks that affect levels of genetic diversity within and among populations. Most importantly, a diverse set of genomic regions may be associated with freshwater adaptation throughout the stickleback range, indicating that this phenomenon is more complex than previously thought.

598A

The effects of pharmacokinetics and -dynamics on the stochastic emergence of antibiotic resistance

Helen Alexander
ETH Zurich, Zurich, Switzerland

Antibiotic treatments of bacterial infections face the challenge of avoiding the emergence of resistant bacterial strains that can compromise treatment success. An important question is thus how one can best design an antibiotic regimen, given constraints due for instance to patient side effects. Ideally, this regimen should both (1) prevent a newly appearing resistant mutant from establishing in the population, and (2) minimize the growth rate of resistant strains that may already be present. An initially rare antibiotic-resistant subpopulation is subject to significant demographic stochasticity, which has been taken into only limited account by previous studies. I propose a fully stochastic model of bacterial population growth that incorporates pharmacokinetics and pharmacodynamics of the antibiotic; mathematically, this is formulated as a time-inhomogeneous birth-death process. I derive analytical results for the probability of avoiding stochastic extinction when rare (relating to the aforementioned goal 1) and the long-term average net growth rate (relating to goal 2) under regular dosing. I use this model to investigate the optimal dosing interval under a total dose constraint and find that, while long-term growth rate is minimized by an intermediate dosing interval, establishment probability can be better reduced by large infrequent doses. Finally, I illustrate how resistance mutations that alter different pharmacodynamic parameters would be predicted to show different emergence patterns.

599B

The complement system as an epitome of host-pathogen genetic conflicts

Diego Forni [1], Rachele Cagliani[1], Giulia Filippi[2], Alessandra Mozzi[1], Luca De Gioia[2], Chiara Pontremoli[1], Uberto Pozzoli[1], Nereo Bresolin[1,3], Mario Clerici[4,5], Manuela Sironi[1]
[1] Scientific Institute IRCCS E.MEDEA, Bosisio Parini, Italy, [2] University of Milan-Bicocca,Department of Biotechnology and Biosciences, Milan, Italy, [3] Dino Ferrari Centre, Department of Physiopathology and Transplantation, University of Milan, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, Milan, Italy, [4] University of Milan, Department of Physiopathology and Transplantation, University of Milan, Milan, Italy, [5] Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

The complement system is an integral arm of innate immunity and plays essential roles in the clearance of pathogenic invaders, in the maintenance of tissue integrity, and in the elicitation of inflammatory reactions. This system is antagonized by an extremely wide array of pathogens and complement evasion underlies the virulence and/or outcome of several infections. We investigated the evolutionary history of the complement system in primates and of bacterial-encoded complement-interacting proteins. Complement components targeted by several pathogens evolved under the strongest selective pressure in primates and selection acted on residues at the contact interface with microbial/viral proteins. Bacterial interactors also evolved adaptively, with positively selected sites located at the interaction surfaces with primate complement proteins. These results reflect the expectations under a genetic conflict scenario whereby the host's and the pathogen's genes evolve within binding avoidance-binding seeking dynamics. We tested this prediction through in silico mutagenesis and protein-protein docking analyses. Positively selected sites, both in the host's and in the pathogen's interacting partner, were found to modulate binding affinity

600C

OASes and STING: adaptive evolution in concert

Alessandra Mozzi [1], Chiara Pontremoli[1], Diego Forni[1], Mario Clerici[2,3], Uberto Pozzoli[1], Nereo Bresolin[1,4], Rachele Cagliani[1], Manuela Sironi[1]

[1] Bioinformatics, Scientific Institute IRCCS E. MEDEA, Bosisio Parini, LC, Italy, [2] Department of Physiopathology and Transplantation, University of Milan, Milan, Italy, [3] Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy, [4] Dino Ferrari Centre, Department of Physiopathology and Transplantation, University of Milan, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, Milan, Italy

OAS (2'-5'-oligoadenylate synthases) proteins and cyclic GMP-AMP synthase (cGAS, gene symbol: MB21D1) patrol the cytoplasm for the presence of foreign nucleic acids. Upon binding to dsRNA or dsDNA, OAS proteins and cGAS produce nucleotide second messengers to activate RNase L and STING (stimulator of interferon genes, gene symbol: TMEM173), respectively; this leads to the initiation of antiviral responses. We analyzed the evolutionary history of the MB21D1-TMEM173 and OAS-RNASEL axes in primates and bats and found evidence of widespread positive selection in both orders. In TMEM173, residue 230, a major determinant of response to natural ligands and to mimetic drugs (e.g. DMXAA), was positively selected in Primates and Chiroptera. In both orders selection also targeted an α-helix/loop element in RNase L that modulates the enzyme preference for ssRNA vs. stem loops. Analysis of positively selected sites in OAS1, OAS2, and MB21D1 revealed parallel evolution, with the corresponding residues being selected in different genes. As this cannot result from gene conversion, these data suggest that selective pressure acting on OAS and MB21D1 genes is related to nucleic acid recognition and to the specific mechanism of enzyme activation, which requires a conformational change. Finally, a population genetics-phylogenetics analysis in humans, chimpanzees, and gorillas detected several positively selected sites in most genes. Data herein shed light into species-specific differences in infection susceptibility and in response to synthetic compounds, with relevance for the design of synthetic compounds as vaccine adjuvants.

601D

Tetrahymena genome architecture provides the benefits of sex in the absence of sex

Rebecca Zufall [1], Joe West[1], Paul Doerder[2], Ricardo Azevedo[1]
[1] *University of Houston, Houston, TX, USA,* [2] *Cleveland State University, Cleveland, OH, USA*

The vast majority of eukaryotes alive today have experienced some form of genetic exchange, or sex, in their recent evolutionary history. While a complete explanation for this observation remains elusive, many evolutionary benefits of sex have been identified. Thus, taxa that have lost the ability to have sex, that is, asexuals, are expected to be evolutionary dead-ends. Nevertheless, some asexual lineages appear to be quite successful. These "exceptions to the rule" have the potential to provide novel insights into the evolutionary costs and benefits of sex. Here, we address the hypothesis that these asexual lineages are successful because they receive benefits normally provided by sex by non-sexual means. Asexual lineages are common in the ciliate genus *Tetrahymena*. We relate the genomic features of these lineages to their ability to reproduce sexually. We also present the results of a model that support our hypothesis that *Tetrahymena* genome architecture can provide the benefits of sex in the absence of sex.

602A

Web spider phylogeography shed lights on the complex history of Neotropical biomes

Fernanda Fontes, Vera Solferini
University of Campinas, Campinas/SP, Brazil

There are many evidences that Tertiary global events of cooling and dryness contributed for the formation of an extensive drier area between the two neotropical rainforests (Atlantic – AF and Amazon - AM) which probably were continuous in the past. To add knowledge about the historical processes of this biogeographical scenario we studied a neotropical web spider (*Aglaoctenus lagotis*) using phylogeographic tools. One mitochondrial and three nuclear markers were used for the analysis of 104 individuals from 26 populations. The bayesian tree indicated that TMRCA is around 4.5 Mya (95% HPD= 2.4 - 7.3), coincident with the advance of grasslands in South America. The BAPS analysis showed the existence of four genetic groups corroborated by the high F_{ST} values and the presence of exclusive haplotypes in each group. AMOVA indicated that 67% and 73% (mt and nu, respectively) of the variation were between groups. These genetic groups are distributed among almost all biomes, with exception of one group that is restricted to South AF. In addition, AF spider lineages presented Northern and Southern structure. These patterns indicate a common history of the biomes in the tropical region and provide new insights for the described differences between tropical and subtropical portions of AF. The drier area seems to connect AM and Northern AF, a pattern that has also been observed for other organisms. Our results indicate that some details of the complex history of Neotropical biomes can only be detected by the study of a widely distributed species..

603B

How many fish species are hidden under the taxon *G. barreimiae* ?

Sandra Kirchner [1,2], Helmut Sattmann[3], Eisabeth Haring[1,2], Arthur Pichler[1], Luise Kruckenhauser[1]

[1] Central Research Laboratories , Museum of Natural History Vienna, Vienna, Austria, [2] Department for Integrative Zoology, University of Vienna, Vienna, Austria, [3] 3rd Zoological Department, Museum of Natural History Vienna, Vienna, Austria

Garra barreimiae is an endemic species in the southeastern Arabian Peninsula (Hajar Mountains) where it is abundant in the south of the United Arab Emirates and northern Oman. So far, three subspecies have been described: *Garra barreimiae barreimiae* Fowler & Steinitz (1956) can be found throughout the northern Oman, *G. b. gallagheri* Krupp (1988) only at one location (Wadi Bani Khalid, northern Oman) and *G. b. shawkahensis* Banister & Clarke (1977) in the north of the United Arab Emirates. Populations of the species *G. barreimiae* inhabit rivulets or surface water bodies in wadis as well as the artificially constructed falaj water pipeline systems. One hypogean population displaying a troglomorphic phenotype lives in the Al Hoota cave. First phylogeographic analyses of *G. barreimiae* were carried out by Kruckenhauser et al. (2011). The genetic differentiation between different epigean populations and the hypogean population was investigated. They revealed that certain populations belong to a very distant lineage. In the present study we present a comprehensive phylogeographic data set, based on a 750 bp fragment of the mitochondrial control region and 350 individuals, on the genus Garra in Oman. At least four different mitochondrial lineages, with a clear geographic pattern following the water drainage systems, can be identified. This pattern is not always congruent with the described subspecies. Moreover, comparison with GenBank data questions the monophyly of the species.

604C

The architecture of human chromosomes and its implications for genome evolution

Giorgio Bernardi
Department of Science, Roma Tre University, Rome, Italy

Genome organization, the key to understand genome evolution, still is an open problem because of its complexity. The strategy used in this investigation to solve this problem was focused on the compositional patterns of DNA sequences spanning in size from a few hundreds to a few thousands Kilobases, a critical range which corresponds to isochores, interphase chromatin domains/borders and chromosomal bands. This compositional approach showed that, in the three-dimensional organization of interphase chromosomes, chromatin domains and chromatin boundaries correspond to GC-poor and GC-rich isochores (and to their higher order structures), respectively. Early-prophase chromosomes, in contrast, have a fully extended configuration, in which chromosomal bands correspond to alternating GC-poor and GC-rich isochores. At early and/or mid prophase, the single-isochore bands of early prophase fold into loops, that progressively cluster with contiguous loops to form prometaphase bands, half of which then coalesce further into metaphase bands, the other half only undergoing a compaction. Both prometaphase and metaphase bands correspond to higher order structures of isochores (the macroisochores and the megaisochores, respectively) that alternate between GC-poor and GC-rich levels. These results 1) solve the "mystery" of chromosomal bands and explain how the same chromosomal sequences allow configurations as different as those seen at interphase and mitosis, so leading to a unifying view of chromosome architecture; and 2) represent a breakthrough in our understanding of genome organization. The implications of the order hidden by complexity of chromosome structure for genome evolution will be discussed mainly in connection with the role of chance in evolution.

605D

The evolutionary benefits of recombination probed with 1,000,000 double barcodes

Jamie R. Blundell[1,2], Daniel S. Fisher[2], Sasha F. Levy[1,2]
[1] Stony Brook University, Stony Brook, NY, USA, [2] Stanford University, Stanford, CA, USA

Sexual reproduction is common in nature, yet the fitness costs / benefits of sex have never been systematically quantified. What is the balance between breaking up good versus bad combinations of mutations? How often are new beneficial combinations discovered? How important is epistasis in this process? Here we describe an experiment that uses a novel DNA barcoding system in S. cerevisiae to uniquely label ~1000 distinct “paternal" strains (MATalpha) and ~1000 distinct “maternal” strains (MATa) and to precisely measure the fitness of each. Mating the ~1000 paternal with the ~1000 maternal strains en masse, we recover ~1,000,000 offspring strains, uniquely labelled by a double-barcode (inherited from the paternal and maternal strains respectively). The abundance of the double barcode can then be used to measure the fitness of the offspring and relate it to the fitness of the respective parents. Using this data we investigate the statistical benefits of sex.

606A

The Phylogenetics and Systematics of *Callicebus* revisited

Jeferson Carneiro[1], Jean Boubli[2], Iracilda Sampaio[1], Izeni Farias[3], José de Sousa e Silva Jr.[4], Alcides Pissinatti[5], Horacio Schneider [1]

[1] Universidade Federal do Para, Belem, Brazil, [2] University of Manchester, Manchester, UK, [3] Universidade do Amazonas, Amazonas, Brazil, [4] Museu Paraense Emilio Goeldi, Belem, Brazil, [5] Centro de Primatologia do Rio de Janeiro, Rio de Janeiro, Brazil

Callicebus is a genus of Neotropical primate of the family Pitheciidae. Thirty-two species are currently recognized. Based on morphology and geographic distributions they are arranged in groups. However, there are discrepancies in relation to the number and composition of these groups. Hershkovitz (1988, 1990) suggested that *Callicebus* should be divided into four groups of species: *C. torquatus*, *C. moloch* (with *C. personatus* and *C. cupreus*), *C. donacophilus,* and *C. modestus*. However, Kobayashi (1995) organized them into five groups: *C. torquatus*, *C. moloch*, *C. donacophilus* (including *C. modestus*), *C. cupreus*, and *C. personatus*. In the present work the phylogenetic relationships among these groups were revisited using eight previously described *Alu* markers (*Alu* insertions, plus up and downstream regions) and three mtDNA genes (16S, CytB. and COI). Seven of the eight *Alu* were found in all groups. Nevertheless one *Alu* insertion, named Callicebus4, was not found in species of *C. torquatus* and *C. personatus* (P) groups suggesting that species from these groups are basal in the genus. Using a multilocus approach we investigated the phylogenetic relationships of the *Callicebus* genus and the time of diversification between the groups. The phylogenetic results indicate four groups: *C. torquatus*, *C. personatus, C. moloch* (with *C. cupreus),* and perhaps also *C. donacophilis* (with *C. modestus*). The results also suggest three splits: at ~12 Ma originating species of *C. torquatus* group and the remaining species; at ~10 Ma originating species of *C. personatus* group, and at ~5 Ma originating *C. moloch* and *C. donacophilus* groups.

607B

High-coverage sequencing of the Human Genome Diversity Project (HGDP-CEPH) Panel

Anders Bergström [1], Qasim Ayub[1], Yasin Memari[1], Shane McCarthy[1], Swapan Mallick[3], Carlos Bustamante[2], David Reich[3], Manjinder Sandhu[1], Richard Durbin[1], Yali Xue[1], Chris Tyler-Smith[1]
[1] Wellcome Trust Sanger Institute, Hinxton, UK, [2] Department of Genetics, Stanford University, Stanford, CA, USA, [3] Department of Genetics, Harvard Medical School, Boston, MA, USA, [4] Broad Institute of Harvard and MIT, Cambridge, MA, USA

Studies of human population genetics have relied mainly on information from genotyping arrays, and large-scale sequencing efforts carried out to date have not focused on sampling the breadth of human genetic diversity. To meet the need for high-quality genome sequences from diverse human populations, we are sequencing the 952 unrelated individuals from the Human Genome Diversity Project (HGDP-CEPH) panel to high coverage (>30X). This panel contains samples from 52 populations spanning Africa, the Middle East, Europe, Asia, Oceania and the Americas. It constitutes one of the most widely used reference sets in human population genetics research and past studies have employed a range of technologies to generate a wealth of results on these samples, which are available for purposes of evaluation and comparison. We have sequenced a third of the panel so far, achieving high accuracy (genotype error rates <0.1%) through single-sample variant calling, and are in the process of sequencing the remaining samples. This data set will enable analyses to be undertaken that exploit the rich and unbiased information contained in whole genome sequences, such as rare variant sharing between populations, the distribution of deleterious variants and detailed maps of archaic admixture. The data will be made freely available and we believe it will constitute a valuable resource for the human genetics community.

608C

Ancient DNA reveals patterns of residential continuity and mobility at the onset of the Central European Bronze Age

Alissa Mittnik [1], Corina Knipper[2], Ken Massy[3], Philipp Stockhammer[3], Johannes Krause[1,4]
[1] Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany, [2] Curt-Engelhorn-Zentrum Archäometrie, Mannheim, Germany, [3] Institute for Prehistory and Early History and Near Eastern Archaeology, Heidelberg University, Heidelberg, Germany, [4] Max Planck Institute for the Science of Human History, Jena, Germany

At the transition from the third to the second millennium BC, the introduction of bronze for the manufacture of tools, weapons and personal ornaments marked a major step in European prehistory. Trade of the metal raw materials and manufactured goods required regular and organized contacts among communities. On the other hand, local population continuity was a prerequisite for the accumulation of wealth, the establishment of enduring social differentiation, and the formation of regional elites.
The archeological record in the Lech Valley in southern Bavaria, Germany, shows a rapid and gapless transition from the Late Neolithic Bell Beaker Phenomenon to the Early Bronze Age. To investigate social and demographic changes associated with the appropriation of the bronze technology, we studied nearly eighty individuals from six burial sites in the region with respect to their maternally inherited mitochondrial DNA (mtDNA).
The results indicate both local genetic continuity spanning the cultural transition, and, following the onset of the Early Bronze Age, a major influx of mtDNA types previously not found in this region. Integrating stable isotope data with the genetic data reveals a picture of a patrilocal society with remarkable mobility in women. While crucial for understanding the change of local demographics, these findings also have implications for the spread of major technological and societal changes across Europe at the beginning of the Bronze Age.

609D

Origin of heredity by spontaneous symmetry breaking: Evolutionary consequences of complementary replication in a minimum model of protocells

Nobuto Takeuchi [1,2], Paulien Hogeweg[2], Kunihiko Kaneko[1]
[1] *University of Tokyo, Tokyo, Japan,* [2] *Utrecht University, Utrecht, The Netherlands*

Complementary replication provides the molecular basis of heredity in all life forms, accomplishing the fundamental prerequisite for evolution. Evolution, however, does not strictly require complementary replication, as attested by numerous evolutionary models that abstract away from complementarity. Here, we use individual-based mathematical modeling to investigate the consequences of complementarity to the evolution of prebiotic replicating molecules enclosed in protocells. In the model, molecules can serve both as catalysts and templates for replication. A catalyst and template first form a complex and subsequently dissociate, synthesizing the complementary strand of the template. This time lag in replication results in a trade-off between molecules spending time as templates and as catalysts. Consequently, the catalytic activity of molecules is driven toward evolutionary deterioration by selection between molecules within protocells. This tendency, however, is counteracted by selection between protocells, whose growth requires internal molecules to replicate. Our simulations indicate that this conflict between multilevel levels of selection leads to spontaneous symmetry breaking, whereby an initially symmetric replication cycle, where the two complementary strands perform an exactly identical role, evolves into an asymmetric cycle, where one strand becomes the majority and serves the dual function of templates and catalysts, whereas the other strand becomes the minority and functions only as templates (like a genome). The presence of this genome-like minority strand alleviates the within-protocell selection pressure for catalytic deterioration by reducing the effective population size of intracellular molecules. Thereby, it enables protocells to survive under wider conditions than possible without complementary replication and symmetry breaking.

610A

To see and to be seen - A closer look into the evolution of a blind fish

Sandra Kirchner [1,2], Helmut Sattmann[3], Elisabeth Haring[2], Luise Kruckenhauser[2]
[1] University of Vienna, Department of Integrative Zoology, Vienna, 1090, Austria, [2] Natural History Museum Vienna, Central Research Laboratories, Vienna, 1010, Austria, [3] Natural History Museum Vienna, Third Zoological Department, Vienna 1010, Austria

The cyprinid Garra barreimiae is endemic to the southeastern Arabian Peninsula, where it inhabits regularly desiccating water bodies. Since a hypogean population with troglomorphic traits (loss of eyes/pigmentation) is conspecific with the "normally" developed surface population, G. barreimiae seems to be a perfect organism to study the effects of cave life and the origin of troglomorphic characters.
Previous genetic analysis (Kruckenhauser et al. 2011) showed that specimens from the cave population are genetically differentiated from the adjacent surface populations. Few surface individuals were found, which exhibit haplotypes characteristic for the cave population and most of these display an intermediate phenotype. This indicates a quite recent separation of the cave population from the surface populations. For a better understanding of the evolution and the maintenance of the troglomorphic phenotype it is crucial to know if and to what extent gene flow between the cave and surface populations occurs. For this purpose, we will investigate individuals from three different cave localities, including individuals with an intermediate phenotype, as well as several surrounding surface populations with 19 variable microsatellite markers, which were designed and established especially for G. barreimiae in a previous study (Kirchner et al. 2014). The results of the present study will deliver the basis for interpreting the recent population genetic history of the species, detect hybridisation and evaluate genetic variability within the species in general.

611B

tba

612C

Notable Decrease in Transcriptome Conservation During Mammalian Brain Aging

Zeliha Gözde Turan, Mehmet Somel
Middle East Technical University, Ankara, Turkey

Aging is nearly universal across metazoa, but its mechanisms have remained elusive. "Mutation accumulation" is among the keystone aging theories postulated to date. It states that mutational load of genes expressed late in life is higher than of genes expressed earlier in life, due to reduced negative selection with age. The idea was previously tested using phenotypic data, with mixed results. Mutation accumulation could also be tested using molecular data. We had previously indentified a trend toward age-related decrease in genes expressed at later ages in a human brain transcriptome dataset, a trend consistent with the mutation accumulation theory. In this study, we investigated the generality of this phenomenon in humans, chimpanzees, macaques and in different brain regions. Across all tested brain regions and species, we detected the same trend of decreasing transcriptome conservation. The signal was not explainable by changes in total mRNA expression during aging. However, analysis of human skeletal muscle and human skin datasets did not display such a trend, indicating limitations of this approach and/or differences among tissues in aging mechanisms. We discuss the possible factors that could explain the observed trend of lower conservation of the aged brain transcriptome.

613D

Evolution of thylakoid membrane complexes in eukaryotes and functional implications of gene loss in *Chromera velia*

Heather Esson [1], Ales Horak[1,3], Petra Dufkova[1], Roman Sobotka[2], Peter Konik[2], Miroslav Obornik[1,3]

[1] Institute of Parasitology, BC ASCR, Ceske Budejovice, Czech Republic, [2] Institute of Microbiology, Centre Algatech, ASCR, Trebon, Czech Republic, [3] Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

The photosynthetic apparatus includes four multi-protein complexes (Photosystems I and II, ATP synthase, and cytochrome b_6/f) that are located in the thylakoid membranes of cyanobacteria and eukaryotic plastids. These complexes are conserved in all oxygenic phototrophs; however, since plastid genomes have lost genes through nuclear transfer or gene deletion over evolutionary time, it is unclear how these processes have affected the protein content and function of the photosynthetic apparatus in different eukaryotic lineages. In order to address this question, we used BLAST to retrieve the sequences for 61 proteins (representing 29 species of cyanobacteria and eukaryotes) associated with photosynthetic complexes from NCBI and other sequence databases. Each protein was categorized as plastid-encoded, nuclear-encoded, or absent, and results were compared amongst the sampled taxa. We found that photosynthetic gene loss is more extensive in *Chromera velia*, a unicellular alga related to apicomplexan parasites, than in other taxa, with 21 absent proteins (34% of those sampled). In order to explore the structural effects of protein absences on photosystem I (PSI) in *C. velia*, we purified this complex and identified individual components by two-dimensional electrophoresis and mass spectrometry. *Chromera's* PSI is atypical. The complex is tightly associated with two different superoxide dismutases and the conserved PsaD, PsaE and PsaF subunits contain an extra C-terminal region. Moreover, we detected an unusually large spectrum of light-harvesting antennae attached to PSI. These data indicate that *C. velia* rebuilt the PSI complex and uses a different strategy to deal with environmental stresses than do cyanobacteria or plants.

614A

The impact of genetic interactions and linkage disequilibrium on estimates of the heritability of complex traits

Emily Glassberg [1], Warren Ewens[2], Marc Feldman[1]
[1] Stanford University, Stanford, CA, USA, [2] University of Pennsylvania, Philadelphia, PA, USA

An important aim of statistical genetics is to determine the genetic architecture of complex traits. Evaluating progress towards this goal requires an estimate of the proportion of a trait of interest that is expected to be attributable to genetics. To this end, we use the statistic of 'narrow-sense heritability.' Traditionally, narrow-sense heritability is estimated by comparing the phenotypic covariance between parents and offspring to that between random pairs of individuals from the population. However, these estimates assume linkage equilibrium and no genetic interactions between causal loci. Several recent studies disagree regarding the expected effects of genetic interactions on this heritability statistic, perhaps suggesting that the impact may be different for traits with different types of genetic architectures.[1],[2] This study aims to characterize the behavior of the heritability statistic using simulated relationships between genotype and phenotype in a two-locus, two-allele system, allowing for both genetic interactions and linkage disequilibrium. Our results suggest (i) that the heritability statistic is very sensitive to changes in chromosome frequencies, (ii) that genetic interactions are likely to make the largest contributions to heritability when all alleles are common, and (iii) that failing to account for linkage disequilibrium may lead to overestimates of heritability.

[1] Mäki-Tanila, A, and Hill, WG (**2014**). Influence of gene interaction on complex trait variation with multilocus models. *Genetics*, *198*(1), 355–67.

[2]Zuk, O, Hechter, E, Sunyaev, SR, and Lander, ES (**2012**). The mystery of missing heritability: Genetic interactions create phantom heritability. *PNAS*, 109 (4), 1193–8.

615B

De novo assembly of a *Vibrio campbelli* isolate causing shrimp deaths and comparative genomics of *Vibrio* species

Huei-Mien Ke [1,2], Anuphap Prachumwat[3,4], Mei-Chin Lai[5], Isheng J Tsai[2], Wen-Hsiung Li[2,6]
[1] Ph.D. Program in Microbial Genomics, National Chung Hsing University and Academia Sinica, Taichung and Taipei, Taiwan, [2] Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, [3] Centex Shrimp, Faculty of Science, Mahidol University, Bangkok, Thailand, [4] Shrimp-Virus Interaction Laboratory, Agricultural Biotechnology Research Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency, Klong Luang, Thailand, [5] Department of Life Sciences, National Chung Hsing University, Taichung, Taiwan, [6] Department of Ecology and Evolution, University of Chicago, Chicago, USA

Shrimp aquaculture is an important industry, but the production is often drastically reduced by pathogens, including *Vibrio* species. A *Vibrio* isolate from a black tiger shrimp (*Penaeus monodon*)-culturing pond in Thailand was identified and named as *Vibrio campbellii* 1114GL by multi-locus sequence analysis (MLSA). In this study, the virulence and whole genome sequencing were analyzed. We observed a 60 % cumulative mortality of *P. monodon* within 12 hours after injection of 10^6 CFU of *V. campbellii* 1114GL. We sequenced this isolate using Illumina MiSeq and Roche 454 platforms, and assembled the reads into 2 chromosomes with a total length of 5,634,012 bps and mean G+C content of 45.5%. In total, 3,180 and 1,906 proteins were annotated in ChrI and ChrII, respectively. A comparison of this genome with four published genomes revealed an expansion of 22 bacterial Ig-like domains group2 (Big_2) in *V. campbellii* 1114GL, which is related to host cell-adhesion and enhances the interactions between bacteria and hosts. Chromosome-wide synteny analysis of this genome against the other finished *V. campbelli* ATCC BAA-116 genome revealed four large inversions (>25 kbs) as well as many minor rearrangements. Our finished bacterial assembly provides a first step for a better understanding of *V. campbellii* evolution.

616C

Neocortical expansion is associated with size variations in gene families with chemotaxis, cell signalling, cell proliferation and inflammatory response-related functions in mammals

Atahualpa Castillo-Morales [1], Alexandra de Souza[1], Jimena Monzón-Sandoval[1], Araxi O. Urrutia[1], Humberto Gutierrez[2]

[1] University of Bath, Bath, UK, [2] University of Lincoln, Lincoln, UK

The species-specific enlargement of the cerebral neocortex is one of the most characteristic traits of human evolution. An expanded neocortex allows species to store and develop the experience, foraging strategies, cognitive processes and overall behavioural plasticity necessary to survive in complex environmental conditions; however, the genomic footprints of this evolutionary process across mammalian lineages are unknown. Whole genome comparisons have revealed large and frequent changes in the size of gene families, and it has been proposed that these variations could play a major role in shaping morphological and physiological differences among species. Using a genome-wide comparative approach, we examined changes in gene family size (GFS) and neocortex volume ratio (NR) in 29 fully sequenced mammalian species and found a significant over-representation of GFS variations in line with increased NR in mammals. We found that this relationship is not accounted for by maximum lifespan and is not explained by phylogenetic relatedness. Genes involved in chemotaxis, inflammatory response, cell-cell signalling and cell proliferation-related functions are significantly over-represented among those gene families most highly correlated with NR. Genes within this families have a more prominent expression in the human neocortex across development before maximal cortical thickness is reach, and decrease in expression afterwards. Our results suggest that changes in GFS associated with NR represent an evolutionary response to the specific functional requirements underlying increased neocortex volume in mammals.

617D

Genetic architecture of complex human traits: A population-genetic model and empirical evidence

Ge Zhang, Patrick Putnam
Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Although genome-wide association studies (GWAs) have revealed thousands of common variants associated with many complex human traits, the genetic architecture of complex human traits remains poorly understood. In this study we develop a population-genetic model to evaluate the distribution of allele frequency and effect size of genetic variants that influence a complex trait. Our approach assumes that a complex trait is influenced by a large number of additive loci under mutation, selection and drift equilibrium. Using forward simulation, we show that our model is valid over a broader range of demographic and selection scenarios. Calibrated based on nearly 700 SNPs known to be associated with human height, our model predicts that complex human traits are likely to be influenced by tens of thousands variants with an average effect size of 0.01~0.02 s.d. and most (~90%) of the phenotype variance is explained by common variants. In addition, our analysis also suggests that very large GWAs (for example, >500,000 unrelated samples) should be able to identify ~2,000 variants that together could explain more than half of the heritability underlying body height and similar complex human traits.

618A

Regressive evolution in Somalian Cavefish *Phreatichthys andruzzii* : loss of selective constraint on opsin genes

Silvia Fuselli [1], Luca Calderoni[1], Elena Frigato[1], Alex Panziera[1,2], Omar Rota-Stabelli[3], Sandra Kirchner[4], Nicholas S. Foulkes[5], Luise Kruckenhauser[4], Cristiano Bertolucci[1]

[1] Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy, [2] Department of Biodiversity and Molecular Ecology, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy, [3] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy, [4] Central Research Laboratories , Museum of Natural History Vienna, Vienna, Austria, [5] Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Karlsruhe, Germany

Phreatichthys andruzzii is a Somalian cavefish that evolved in complete isolation and absence of light beneath the desert for about two million years. Constant darkness caused extreme degenerative phenotypes, such as complete depigmentation, reduced metabolic rate and complete eye degeneration. The circadian clock is partially degenerated in this cavefish, and mutations in several photoreceptors and clock-related genes play roles in its lack of response to light. However, a detailed description of the molecular mechanisms underlying the regression of this important mechanism is still missing. We investigated the molecular evolution of the non-visual photoreceptor melanopsin Opn4m2, whose premature stop codon accounts for the inability of the peripheral clock to respond to light. To test the hypothesis that other light-related mechanisms are undergoing degeneration, we studied the molecular evolution of the visual pigment rhodopsin, expressed in the brain of *P. andruzzii* and probably involved in its photophobic behavior. The same genes were studied in another blind cavefish, *Garra barreimiae* from Oman, a close relative to *P. andruzzii* that independently colonized subterranean waters and evolved troglomorphic traits. Our results based on within and between species analyses, and on time estimation of pseudogenization, show that both genes lost signature of selective constraints in *P. andruzzii*. Conversely, *in G. barreimiae*, that colonized the cave environment more recently and lacks complete isolation from the surface, the genes are still conserved. Our observations indicate that the long and extreme isolation of *P. andruzzii* in darkness led to a general relaxation of natural selection on light-responsive physiological mechanisms.

619B

Fast and accurate alignment-free phylogenetic inference

Marcin Bogusz, Simon Whelan
Uppsala University, Uppsala, Sweden

Statistical methods for estimating trees from molecular data rely on a two-step process of multiple sequence alignment (MSA) followed by tree inference. There is substantial evidence that MSA quality affects the reliability of phylogenetic inference, since flaws in the alignment tend to cascade downstream. This awkward interaction between MSA and tree inference has lead to several alternatives to the two-step process. Alignment-free methods infer pairwise distances from k-mer occurence and use clustering to infer trees. Full statistical alignment can jointly estimate both the phylogeny and MSA, but is computationally intensive.

We propose an efficient approach that uses the principles of statistical alignment to estimate pairwise distances from unaligned sequences, and then uses bioNJ to infer a tree. We apply pair-hidden Markov models (pair-HMMs), which describe insertion, deletion and substitution events. Pair-HMMs allow us to account for alignment uncertainty by integrating over all possible homologies and have the potential to avoid many of the biases associated with MSA. We find that our method can provide more accurate estimates of pairwise distances than estimates taken from inferred fixed alignments and introduce only a modest increase in the variance of that estimate. We provide a fast Maximum Likelihood approach for estimating the parameters of the pair-HMM evolutionary model and pairwise distance estimates conditional on that model. Simulations show that the bioNJ trees estimated from our pair-HMM-derived distances tend to be more accurate than those obtained using conventional pairwise distance approaches and comparable with those obtained using state-of-the-art methods in the standard phylogenetic pipeline.

620C

Evolutionary mechanisms of algebraic distributions of segmental duplication lengths observed in pair-wise comparison of long DNA sequences

Maxim Koroteev, Pavel Baranov
University College Cork, Cork, Ireland

We introduce a family of models for evolution of long DNA sequences (chromosomes) incorporating random segmental duplications and point mutations which are capable to reproduce algebraic length distributions of exact matches (identical duplicates) with the slope -4 observed earlier in pair-wise comparisons of DNA of distantly related species. The models incorporate selective pressure in the form that mutation rates in some parts of a chromosome differ by orders from those in other parts of the chromosome. We also demonstrate that this scale-free evolutionary dynamics produces sequences which by applying the iterated maps to them demonstrate qualitative agreement with patterns observed for natural DNA sequences. While previous studies stressed that iterated map patterns of natural DNA can be reproduced by a Markov model of low order, our models provide a natural way to obtain both power-law length distributions and those patterns observed in natural chromosomes, as Markov models of reasonable order are not able to reproduce heavy tails of algebraic length distributions.

621D

Structural variation at the glycophorin locus from genome sequencing of 125 Gambian trios

Ellen Leffler [1], Si Quang Le[1], Katja Kivinen[2], Muminatou Jallow[3], Fatou Sisay-Joof[3], Jim Stalker[2], Kirk Rockett[1,2], Gavin Band[1], Dominic Kwiatkowski[1,2], Chris Spencer[1], Gambie Genome Diversity Project[1,3]

[1] Wellcome Trust Centre for Human Genetics, Oxford, UK, [2] Wellcome Trust Sanger Institute, Cambridge, UK, [3] MRC Unit, Fajara, Gambia

Recent population-scale sequencing projects are greatly informative about patterns of polymorphism, such as haplotype structure and rare genetic variation. These studies also provide high coverage sequence across complex regions of the genome for which variation has been difficult to describe but which are of medical or evolutionary interest. One such region is the glycophorin gene cluster on chromosome 4, which encodes GYPA, GYPB and the non-translated GYPE, with the first two underlying the MNS blood group system and serving as red blood cell surface receptors for *Plasmodium falciparum* and other pathogens. Several observed patterns of variation within the locus are suggestive of positive or balancing selection, and may reflect a continued evolutionary battle between host and parasites. However, further interpretation and study is complicated by the high similarity among the three homologous copies, each about 100 kb, which makes assembly, mapping and variant calling in this region problematic. Here, we use whole-genome sequencing data of trios from four ethnic groups in the Gambia to infer large structural variants that are robust to the homology of the region and to estimate their breakpoints and the number of mutational events. We also infer structural variants in worldwide 1000 Genomes populations, describing the distribution of deletions and duplications that alter gene copy number. The identification of this variation can then be integrated into analyses of association with infectious disease and signatures of selection.

622A

PhylomeDB: where every protein has its tree

Salvador Capella-Gutíerrez[1], Marina Marcet-Houben[1], Leszek Pryszcz[1], Miguel A. Naranjo[1], Laia Carreté[1], Ernst Thuer[1], Toni Gabaldón [1,2]

[1] Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra (UPF), Barcelona, Spain,

[2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Phylomes are defined as complete sets of evolutionary histories of genes encoded in a given given organism. By providing a look of the genome past through the perspective of all of its genes, phylomes are powerful tools to study genome dynamics and how the relate to important phenotypic innovation. PhylomeDB is the largest public database for pre-computed gene family trees, currently storing over 5 million phylogenetic trees and alignments including genes from over 1000 fully-sequenced genomes. Trees are built using a sophisticated automated pipeline that includes, homology searches, alignment trimming, evolutionary model testing and maximum-likelihood inference. PhylomeDB allows to visualize phylogenetic trees, multiple alignments and orthology and paralogy predictions. In addition, data can also be downloaded through web API. Each gene tree displays PFAM domains, taxonomy-information panels, functional information, visualization of speciation and duplication events and links to other databases. Finally, the web API provides direct links to trees and phylomes, as well as to custom nodes within a tree topology. PhylomeDB orthology predictions are expanded via MetaPhOrs, which integrates phylogenetic information from other databases (Ensembl, EggNOG, TreeFam, OrthoMCL, among others) to produce a consensus prediction and a consistency-based realibility score.
PhylomeDB is available at http://phylomedb.org/.
MetaPhOrs is available at http://orthology.phylomedb.org

623B

Intramolecular phenotypic capacitance in a modular RNA molecule

Eric Hayden
Boise State University, Boise, USA

Phenotypic capacitance refers to the ability of a genome to accumulate multiple mutations that are conditionally hidden, and only reveal phenotype-altering effects after certain environmental or genetic changes. Capacitance has important implications for the evolution of novel forms and functions, but known mechanisms behind capacitance are limited to the disruption of complex, multi-component systems involving protein interactions. Here I present evidence of capacitance within an individual ribozyme. This naturally occurring RNA molecule has a modular structure, where a scaffold module acts as an intramolecular chaperone that facilitates folding of a catalytic module. I identify conditional mutations that alter the wild-type ribozyme phenotype under a stressful condition (low magnesium) but preserve the phenotype under more relaxed conditions. The scaffold can buffer the deleterious effects of a modest number of random mutations before additional mutations cause a synergistic decline in activity, but this modest buffering is lost in a stressful environment. This conditional buffering is confined to the scaffold module, but controls the catalytic phenotype, demonstrating how modularity enables capacitance. These results show that capacitance does not require the evolution of dedicated chaperone machinery and can be achieved by individual biomolecules with robust scaffold modules. This suggests that capacitance may have contributed to the evolution of novel and diverse RNA functions, including those required for life to begin.

624C

Comparative pigment cell transcriptome analysis provides molecular clues to the labyrinthine pigment pattern of marble trout skin

Ida Djurdjevič [1], Tomasz Furmanek[2], Simona Sušnik Bajec[1]

[1] Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia, [2] Institute of Marine Research, Bergen, Norway

Skin pigmentation is an important phenotypic trait that plays a critical role in social communication, camouflage, and protection against harmful effects of solar radiation. In comparison to most other trout species marble trout (*Salmo marmoratus*) exhibits a striking labyrinthine skin pattern. To identify candidate genes and molecular pathways that may play important roles in trout skin pigmentation, we performed Illumina transcriptome sequencing to catalog global gene expression profiles of pigment cells from the skin of marble trout compared to closely related brown trout (*S. trutta*), that exhibits a spot skin pattern. A total of 812,144,586 pair-end reads of 100 bp were generated (414,612,488 in marble and 397,532,098 in brown trout) and mapped to the Atlantic salmon (*S. salar*) genome. More than 83,000 transcripts were annotated. Comparisons of pigment cell transcriptomes revealed numerous differently expressed genes. KEGG analysis of these genes indicated that specificities in melanogenesis, tyrosine metabolism and MAPK signaling pathways most likely affect the difference in skin pigmentation between the two species. Several key genes involved in the skin pigmentation (eg. *dct, tyr, tyrp1*) showed significant differences in expression patterns of two species. Furthermore, differential expression patterns were also observed when comparing differently colored regions of skin from the same individual fish, most probably associated with specific expression profiles of individual pigment cell types predominating in particularly colored skin region.

625D

Mitochondrial and nuclear markers show different phylogeographic patterns in the Red Sea collector urchin *Tripneustes gratilla*

Elisabeth Haring [1], Omri Bronstein[2], Andreas Kroh[1]
[1] Naturhistorisches Museum Wien, Vienna, Austria, [2] 2 Steinhardt National Natural History Museum and Research Center, Tel Aviv, Israel

Tripneustes gratilla is one of the most widespread Indopacific shallow-water echinoids. *T. gratilla* representatives from the Red Sea are morphologically well differentiated from populations in other parts of the species' range. Consequently, a separate subspecies was proposed for the Red Sea population: *T. gratilla elatensis*.

Phylogeographic studies based on mtDNA demonstrated a lack of phylogenetic structure for *T. gratilla* and haplotype sharing was observed between widely distant regions. Here we present first genetic data for *T. g. elatensis* from the Red Sea, which so far has not been included in phylogeographic studies on *Tripneustes*. Phylogeographic reconstruction was based on two marker sequences (mitochondrial *COI* gene, nuclear *Bindin* gene) with samples from widely distant populations. In contrast to the expectations based on the apparent morphological differences, Red Sea *Tripneustes* are not differentiated from Indian Ocean populations and share haplotypes with specimens from both Philippines and East Africa.

Bindin sequences, however, tell a different story: three clades were found, comprising 1) Red Sea, 2) Indomalayan (Philippines, Guam, Papua New Guinea), and 3) mixed Indo-West-Pacific samples. High levels of differentiation in *Bindin* sequences among these clades suggest ancient radiation, greatly contrasting the COI data. This apparent mismatch of mitochondrial and nuclear DNA data implies a mitochondrial capture event in the recent past. Independent morphological evidence lends support to this, since in-depth investigations indicate that the differentiation of *T. g. elatensis* is not restricted solely to minor difference in corona proportions and coloration, but is also observed in lantern morphology and tubefeet spicules.

626A

Association between diabetic nephropathy and *cndp1* polymorphisms in Puerto Ricans

Priscila Marie Rodriguez-Garcia, Juan Carlos Martinez-Cruzado, Taras K. Oleksyk
University of Puerto Rico at Mayaguez, Mayaguez, Puerto Rico

In 2005, a polymorphism in a gene called *cndp1* was found to be related to diabetic nephropathy (DN) in Europeans. This polymorphism is a tri-nucleotide repeat, in which ther shortest allele (5 repeats or Mannheim allele) was related to protection against DN. After this study, many more where conducted in different populations, but the relationship didn't hold true in all of them. This leads us to think that there are other polymorphisms contributing to this protective factor, which could be localized in chromosomal ancerstry blocks. We will analyze genetic and protein data to compare extreme cases in the sample pool. This data will help us elucidate whether this polymorphisms are related to the development of DN. This study will be done in a very convinient population because of the admixed struture of the DNA of today's Puerto Ricans. This study has the potential to: resolve the current contradictions about *cndp1* polymorphisms and DN, broaden ther spectrum of scientific knowledge about genetic factors influencing the development of DN, and demonstrate the importance of studying Puerto Ricans as a multi-ethnic group.

627B

Positive selection scan in *Glossina morsitans* and five others true flies.

Lucas Freitas [1,2], Jacob Crawford[2], Carlos Schrago[1]

[1] Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil, [2] University of California, Berkeley, Berkeley, California, USA

Flies in the genus *Glossina* are vectors of trypanosomes in both animal and humans, causing the neglected tropical disease known as sleeping sickness. To identify genes under positive selection putatively involved in blood-feeding and viviparity in this genus, we used a positive selection scan over 1498 orthologous genes in six species of Diptera. We selected orthologous proteins (1:1) present in eight species: *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus*, *Drosophila melanogaster*, *Glossina morsitans* and *Lutzomyia longipalpis* (Diptera) plus *Heliconius melpomene* and *Bombyx mori* (Lepidoptera) available in the OrthoDB database. We used the branch-site test in PAML 4.8 to identify genes under positive selection in each of the Diptera species and, despite the elevated divergence time between these species, we found genes with a signal of positive selection in all species tested including 14 in *G. morsitans*. Furthermore, we used PantherDB to assign Gene Ontologies (GO) and GOrilla to test for possible enrichment of genes under positive selection in any GO term. In the gene enrichment analysis, we found enriched GO terms in *A. gambiae*, *D. melanogaster*, *C. quinquefasciatus* and *G. morsitans*, although no good candidates involved in blood-feeding and viviparity. As a next step, we will relax the orthology restrictions and expand the test to include more orthologs.

628C

The origin and genetic diversity of horses in Puerto Rico

Nikole Ayala-Agosto, Juan Martinez-Cruzado, Taras Oleksyk
University of Puerto Rico Mayagüez, Mayagüez, Puerto Rico

Since their domestication, horses demonstrated their usefulness for transportation, warfare, agriculture, sports and even companionship, earning a central role in human history. Over centuries horse breeds have been established by selective breeding for desirable traits such as endurance and strength, one example is the Pure Puerto Rican Paso Fino. Horses were introduced to Puerto Rico during the Spanish colonization and subsequent Spanish settlers that brought different Iberian horses including Barb, Andalusian and Sapnish Jennets. The Pure Puerto Rico Paso Fino was selected for its natural ambling gate characterized by its smooth isochronal four beat lateral gait. Recently it was discovered that footfall patterns (gaits) in horses is controlled by a gene known as the DMRT3 and that a premature nonsense mutation is required on breeds with altered gait characteristics. Here we study a ongoing growing population of unrelated non-purebred horses in Puerto Rico regarded by locals as non-Paso Fino in order to identify the genetic variation of horses in the island and the presence of the DMR3T mutation required for “gaitedness”. We describe the horse population as it admixed and diversified horse from its Iberian ancestors.

629D

Wasabi: a web-based platform for evolutionary sequence analysis

Andres Veidenberg, Alan Medlar, Ari Löytynoja
Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Wasabi is a modern, web–based software environment for evolutionary sequence analysis. The application allows to work with multiple sequence alignments in the context of the corresponding phylogenetic trees and visualizes connected data, like ancestral sequences, within the phylogeny. Wasabi leverages recent web technologies for cross-platform access and fluid graphical interface -- context sensitive menus and drag-and-drop editing provide clutter-free interaction with both trees and sequences. Wasabi streamlines the analysis process by tight integration of external resources and tools, including Ensembl database and phylogeny–aware alignment methods PRANK and PAGAN, and provides a plugin system to extend the environment with additional tools. The central data storage system and user accounts improve reproducibility of analyses by automatically saving intermediate results, and aids collaboration and publishing through web link-based data sharing. Wasabi is open-source, runs inside a web browser and can be launched from http://wasabiapp.org.

630A

The "MetaCopepod" project: Designing an integrated DNA metabarcoding and image analysis approach to study and monitor biodiversity of zooplanktonic copepods

Panagiotis Kasapidis [1], Constantin Frangoulis[1], Ioanna Siokou[2], Efstratios Batziakas[1], Maria Grazia Mazzocchi[3], Maria Luz Fernandez de Puelles[4], Meriem Khelifi Touhami[5], Yesim Ak Orek[6], Aleksandra Gubanova[7], Rana Abu Alhaija[8], Epaminondas Christou[2]
[1] Hellenic Centre for Marine Research, Heraklion, Crete, Greece, [2] Hellenic Centre for Marine Research, Anavissos, Greece, [3] Stazione Zoologica Anton Dohrn, Napoli, Italy, [4] Centro Oceanografico de Baleares, Palma de Mallorca, Spain, [5] Badji Mokhtar University, Annaba, Algeria, [6] Middle East Technical University, Mersin, Turkey, [7] Institute of Biology of the Southern Seas, Sevastopol, Ukraine, [8] The Cyprus Institute, Nicosia, Cyprus

Timely and accurate analysis of zooplankton biodiversity poses an ongoing challenge in ecological and biomonitoring programs, since morphology-based identification of taxa is time consuming and rarely supports species-level resolution. Copepods, the main constituent of zooplankton, play a major role in marine ecosystems functioning, however their study is still limited because of such methodological reasons. The recent advances in genetic and image analysis systems offer an opportunity to overcome these limitations. The 'MetaCopepod' project, based on the combination of next-generation sequencing technologies and image analysis, aims to develop a fast, high-throughput, cost effective and accurate methodology, to assess and monitor the biodiversity of planktonic copepods in the Mediterranean and the Black Sea, in terms of species composition, abundance, biomass and size-distribution, without the need of a taxonomy expert. To achieve this, bulk copepod samples are first analysed using an image analysis software "trained" to automatically recognize, count and size-measure images of copepods. Subsequently, the same samples are massively sequenced for a selected DNA fragment (barcode), and through a bioinformatic pipeline, sequences are compared to a reference genetic database and identified at species-level. The combination of these approaches will allow to analyze copepod communities both qualitatively and quantitatively with accuracy. 'MetaCopepod' will have an immediate impact on copepod studies by unmasking hidden copepod diversity, and facilitating the identification of the numerous small-size species and early copepod life stages. In addition, it will create an efficient biomonitoring tool for detecting ecosystem changes due to global warming, bioinvasions and other human activities.

631B

Identifying and investigating obsessive-compulsive disorder genes using comparative genomics

Hyun Ji Noh [1], Ruqi Tang[2], Elinor Karlsson[1], Ross Swofford[1], Jason Flannick[1], Colm O'Dushlaine[1], Jeremy Johnson[1], Michele Koltookian[1], S Evelyn Stewart[3], James Knowles[4], Carol Mathews[5], Jeremiah M Scharf[6], Guoping Feng[2], Kerstin Lindblad-Toh[1,7]
[1] Broad Institute, Cambridge, MA, USA, [2] MIT, Cambridge, MA, USA, [3] UBC, Vancouver, Canada, [4] USC, LA, CA, USA, [5] UCSF, San Francisco, CA, USA, [6] MGH, Boston, MA, USA, [7] Uppsala University, Uppsala, Sweden

Obsessive-compulsive disorder (OCD), a debilitating mental disorder manifested in time-consuming repetition of behaviors, affects 1-3% of the human population. Genome-wide association studies (GWAS) in human OCD have not yet identified significant genetic associations. To identify genes and candidate mutations, we performed targeted resequencing of coding and non-coding mammalian constrained sequences of 608 genes from the canine OCD GWAS and the orbitofronto-striatal pathway implicated in mouse, in 592 DSM-IV OCD cases and 560 matched controls. In order to investigate the various mechanisms of disease gene actions and evolutionary aspects, we evaluated the genic burden of variants found in coding regions, regulatory regions, conserved bases (signature of purifying selection), and divergent bases (signature of adaptive selection) separately.

Five genes were enriched with variants among OCD patients after multiple testing corrections. Two had excess coding variants, two had excess regulatory variants, and one had excess of overall variants. One gene that had excess of regulatory variants also showed excess of variants found in divergent bases. Genetic replication in independent samples and functional validation of candidate mutations are in progress. In summary, by integrating comparative genomics systematically into a disease genetic association study, we were able to achieve significant associations using only ~600 human cases, instead of tens of thousands cases required for human-only genetics approach. Furthermore, enrichment of variants on divergent bases among OCD patients may suggest that brain functions that have been adapted throughout evolution may be affected in OCD.

632C

Proteome-wide structural analysis of positive selection in mammals

Greg Slodkowicz and Nick Goldman
European Molecular Biology Laboratory; European Bioinformatics Institute; Wellcome Trust Genome Campus Hinxton; Cambridge, UK

Selective pressure acting on a protein site can be estimated by comparing the rates of non-synonymous and synonymous substitutions (dN/dS) across a evolutionary range. This measure can be used to distinguish purifying (dN/dS < 1), neutral (dN/dS ~= 1) and positive selection (dN/dS > 1). Identifying sites under positive selection is of particular interest as it can help pinpoint specific adaptations responsible for differences between species. Here, we investigate how protein structure influences the selective regime experienced by a site. Using a comprehensive set of mammalian alignments and available crystal structures, we estimate the distribution of selective constraint (dN/dS) separately for regions with different secondary structure and solvent exposure. We also evaluate the evidence for linear and spatial clustering of sites under positive selection. As expected, we find dependence on secondary structure and solvent accessibility as well linear dependencies between sites under positive selection as but surprisingly little statistically significant spatial clustering. We also investigate how information about selective constraint can be integrated into wider biological investigations e.g. with the aim of comparing inter-species selective constraint with evolutionary dynamics of the human population.

633D

Computationally efficient inference of detailed population history using the allele frequency spectrum from large sample genomic sequences

Hua Chen [1], Kun Chen[2]
[1] Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, [2] Dana Farber Cancer Institute, Harvard Medical School, Boston, USA

The allele frequency spectrum is an important summary of genomic polymorphism, and extensive theory and methodology of the allele frequency spectrum for population genetic inference have been developed via diffusion approximation and coalescent theory. The coalescent-based JAFS analytically estimates the coalescent likelihood of the sample, and is useful for inferring ancient demographic history. However, the method becomes computationally intractable when the sample size is large. Recent large sample genomic sequencing projects provide unprecedented opportunities for learning population histories, however, the sample size is much beyond the limit of existing methods. Thus these studies have to rely on computationally intensive simulations, or hypergeometric summation, which is difficult to be extended to complicated scenarios.

By appealing to asymptotic distributions of coalescent times for populations with time-varying size, we develop accurate approximation for the allele frequency spectrum for large samples. The approximation is in simple analytical form, and computationally very efficient comparing the simulation-based methods. More importantly, the result is accurate and flexible for various complex demographic scenarios. We demonstrate the precision and accuracy of inferring complex demographic history using simulations. The method is applied to two recently available large-sample sequencing data to infer the population history of the Northern European populations. We expect our method be a useful tool for population genetic studies in the large-scale genomic sequencing era.

634A

Genetic difference of male and female cycads using AFLP markers

Chatuporn Kuleung
Kasetsart University, Bangkok, Thailand

Cycads are dioecious that male and female plants can be identified only by their reproductive organs or cones; however, these organs take several years to develop. To identify sex of cycads at early age, DNA markers could be an effective tool for identifying gender of these plants. AFLP is a technique that can be applied to define many polymorphic regions in a whole genome at a single run. Another advantage of this technique is that it does not require a prior knowledge about the sequence of the organism of interest. These advantages make it an efficient tool for studying genetic variation of novel organisms. Therefore the objective of this research is to investigate genetic difference between male and female cycads using AFLP marker and to develop a specific marker that can be used to differentiate gender of plants in cycads.

635B

Transcriptomic and phenotypic response in female *Drosophila melanogaster* to male ejaculate transfer after long-term nutrition selection.

Wayne Rostant, Irina Mohorianu
University of East Anglia, Norwich, Norfolk, UK

Ejaculate chemicals transferred from males to females during mating can cause significant changes in female behaviour and physiology. These effects can favour the evolutionary interests of males while generating costs in females in an environment-dependent manner. This has been demonstrated for the nutrition-dependent response of female *Drosophila melanogaster* to a specific ejaculate component, sex peptide (SP), which affects female egg laying, sexual receptivity, feeding rate and immune responses. Previous work in our lab has shown that the food environment can significantly shape the strength and direction of selection on mating responses to SP. Another strain of research has shown that SP is an upstream master-regulator, able to induce these diverse phenotypes through efficient induction of widespread transcriptional changes in females. The present study seeks to bridge these two threads using laboratory natural selection on adult diet quality to examine whether phenotypic and transcriptional response in females to SP receipt is altered by selection. After 50 generations of selection on high and low yeast diets, egg laying and female receptivity were examined by mating adult females to males that did or did not transfer SP. Total RNA was extracted from a subset of assayed females and we used RNAseq and recently developed bioinformatics methods to analyse female transcriptomic responses. We present here the results of both the phenotypic and gene expression analyses.

636C

Designing molecular diagnostics from shotgun sequencing data: a case study using Leishmania

Kelly M Harkins [1,2], Rachel Schwartz[3], Lars Fehren-Schmitz[1], Reed Cartwright[3,4], Anne C Stone[2]
[1] Department of Anthropology, University of California, Santa Cruz, USA, [2] School of Human Evolution and Social Change, Arizona State University, Tempe, USA, [3] Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, USA, [4] School of Life Sciences, Arizona State University, Tempe, USA

The software SISRS can be used to find phylogenetically informative loci from next-generation sequencing data without prior knowledge of informative markers or a reference genome. In a novel application of the software, we use these loci to develop diagnostic sets of probes for DNA enrichment via in-solution hybridization; SISRS automatically determines the the loci for these probes. We apply this method to the kinetoplastid parasites *Leishmania* and *Crithidia*. *Leishmania* species can be difficult to differentiate in patients, but infections from distinct species must be treated differently. Following hybridization capture and Illumina sequencing, our results show correct phylogenetic placement of each isolate enriched with the SISRS-derived probes. Thus, SISRS has provided useful diagnostic loci for differentiating *Leishmania*. This software can be applied to other genera in the future, providing benefits in a clinical setting.

L 1A

Accuracy of phasing of *Drosophila melanogaster* genotypes

Marek Bukowicki, Susanne U. Franssen, Christian Schlötterer
Institut für Populationsgenetik, Vetmeduni Vienna, Austria

The number of applications, which build on haplotype information (e.g. GWAS, imputation or ancestry determination) is continuously increasing. Since typically, diploid individuals are being sequenced, the reconstruction of haplotypes (phasing) is becoming an essential part of genomic analyses. Various software packages for phasing exist, but were primarily targeted for and benchmarked on human data. With the availability of high throughput data for other organisms, it is important to evaluate the accuracy of the existing phasing approaches for species, which differ in the amount of linkage disequilibrium, SNP density and available sample sizes.

In this study we investigate phasing accuracy for *Drosophila melanogaster* genotypes using genome-wide Next Generation Sequencing data. *D. melanogaster* has a higher SNP density and less linkage disequilibrium than humans. Publicly available datasets of known phase (DGRP, DPGP) are used for phasing and serve as reference haplotypes with the SHAPEIT2 software. Phasing performance is measured by the switch error rate (SER) and the mean distance (in bp) between switch errors. We examine the influence of several parameters including the number of reference haplotypes and their population origin, completeness of marker information and addition of sequencing read information.

We obtained a 2% SER in *D. melanogaster*, which is similar to humans. With a 4-10-fold higher SNP density, mean switch distance is greatly reduced-from 200 kb in humans to about 10 kb in fruit flies. This suggests that the higher SNP density cannot compensate for the higher recombination rate in *D. melanogaster*. We caution that genome-wide haplotype based analysis which are based on computationally phased data need to be treated carefully. Rather experimental haplotyping may be preferable in Drosophila.

L 5A

High rates of large-scale mutations in Daphnia mutation accumulation lines revealed by whole genome sequencing

Jullien M. Flynn, Frederic J.J. Chain and Melania E. Cristescu
McGill University, Quebec, Canada

Spontaneous mutations provide the raw substrate for evolution on which selection and genetic drift act, and are therefore fundamental to many aspects of evolutionary biology. However, spontaneous mutations have been very difficult to investigate in a direct fashion and in full spectrum. In this study we directly estimated the mutation rates for point mutations as well as large-scale duplications and deletions in *Daphnia* mutation accumulation (MA) lines. *Daphnia* has a compact genome of ~200Mb characterized by a rich gene count and an unusually high occurrence of gene duplications. Using whole genome sequencing of 24 MA lines that have been propagated asexually for an average of 82 generations, we measured a nucleotide mutation rate of 6.72 x 10-9 per site per generation and an indel mutation rate of 9.54 x 10-9 per site per generation, while the rate of loss of heterozygosity (LOH) was orders of magnitude higher at 1.09 x 10-4. The exceptionally high rate of LOH is likely due to a combination of ameiotic recombination (homozigosity) and deletions (hemizygosity). LOH resulted in long stretches of homozygosity, with one of the MA lines experiencing a large-scale LOH event (or events) encompassing over 3.6Mb on a single chromosome. Preliminary results indicate that numerous LOH events are due to deletions. We conclude that large-scale duplication and deletion events occur at a high rate in *Daphnia*, allowing rapid genomic reconfiguration and offering high potential for diversity and adaptation.

6 Population Genomics of Rapid Adaptation

6.1

Rapid evolutionary response to climate change in an ecologically important invertebrate

Thomas Bataillon [1], Galtier Nicolas[2], Aurelien Bernard2[2], Nicolai Cryer[1], Nicolas Faivre[2], Sylvain Santoni[3], Dany Severac[4], Teis Mikkelsen[5], Klaus Larsen[5], Claus Beier[5], Jesper Sørensen[6], Martin Holmstrup[7], Bodil Ehlers[7]

[1] BiRC, AArhus, Denmark, [2] ISEM, Montpellier, France, [3] INRA – UMR AGAP, Montpellier, France, [4] MGX-Montpellier GenomiX, Montpellier, France, [5] Technical University of Denmark, Lyngby, Denmark, [6] Department Biosciences, Aarhus University, Aarhus, Denmark, [7] Department Biosciences, Aarhus University, Silkeborg, Denmark

Although evolutionary response of species is central to our understanding of climate change impacts on biodiversity, few studies have documented genetic adaptations to rapid and recent climate change and most rely on natural cline studies rather than experimental manipulations of the climatic environment. Here we study how the soil living annelid worm, *Cognettia sphagnetorum* (Oligochaeta; Enchytraeidae) responded genetically to a moderate change in climate conditions over a 7-year period. We use a controlled climate change experiment where forecasted future climate conditions have been applied in a replicated natural setting since 2005. Analyzing the transcribed genome of 15 local populations of this worm by RNA-seq, we found that up to 10% of the genetic polymorphism we surveyed exhibit differences in allele frequencies among populations that are associated to measurable changes in soil temperature and soil moisture. We document pervasive changes among populations subject to mean soil temperatures differences by only 0.5 °C over the 7-year period in which soil temperature was manipulated. Our study shows that evolution can happen in response to realistic climate change scenario even over a short time scale. This calls for more widely incorporating evolutionary processes into models predicting future distribution and response of species to climate change

6.2

Predicting Carriers of Ongoing Selective Sweeps with the Haplotype Allele Frequency (HAF) score

Roy Ronen[1], Glenn Tesler[1], Ali Akbari[1], Shay Zakov[1], Noah Rosenberg[2], Vineet Bafna [1]
[1] U C San Diego, La Jolla, CA, USA, [2] Stanford Univ., Palo Alto, USA

Methods for detecting the genomic signatures of natural selection are heavily studied, and have been successful in identifying many selective sweeps. For the vast majority of these sweeps, the adaptive allele remains unknown, making it difficult to distinguish carriers of the sweep from non-carriers. Because carriers of ongoing selective sweeps are likely to contain a future most recent common ancestor, identifying them may prove useful in predicting the evolutionary trajectory — for example, in contexts involving drug-resistant pathogen strains or cancer sub-clones. The main contribution of this paper is the development and analysis of a new statistic, the Haplotype Allele Frequency (HAF) score, assigned to individual haplotypes in a sample. The HAF score naturally captures many of the properties shared by haplotypes carrying an adaptive allele. We provide a theoretical model for the behavior of the HAF score under different evolutionary scenarios, including neutral Wright-Fischer evolution, exponential growth, and the trajectory of HAF-scores during a selective sweep.

We validate the theoretical analysis using extensive simulations, and demonstrate how the HAF-scores change dynamically with the progression of selective sweep, and are different for carriers and non-carriers of a favorable allele. We use this observation to design an algorithm, PreCIOSS (Predicting Carriers of Ongoing Selective Sweeps) to identify carriers of the adaptive allele in selective sweeps, and we demonstrate its power on simulations of both hard and soft selective sweeps, as well as on data from well-known sweeps in human populations.

6.3

Characterizing adaptive evolution in *Drosophila melanogaster* through the patterns of incomplete selective sweeps detected by composite likelihood ratio, extended haplotype, and other summary statistic-based methods

Ha My Vy Thi, Yuseob Kim
Ewha Womans University, Seoul, Republic of Korea

Adaptive evolution occurs as beneficial mutations arise and then increase in frequency by positive natural selection. How, when, and where in the genome such evolutionary events occur is a fundamental question in evolutionary biology. When a beneficial mutation is on the way to fixation, homologous chromosomes in the population are divided into two groups: one carrying the beneficial allele with very low polymorphism at nearby linked loci, and the other carrying the ancestral allele with normal pattern of sequence variation. We recently developed a composite likelihood ratio (CLR) test for detecting incomplete selective sweeps based on the joint sampling probabilities for allele frequencies of such two groups. Tested against simulated data, this method yielded statistical power and accuracy in parameter estimation that are higher than the *iHS* test and comparable to more recently developed nS_L test. Our evaluations suggest that different methods capture different aspects of genetic information regarding incomplete sweeps and thus are partially complementary to each other. We further explored other patterns of variation, captured by summary statistics including w and F_{ST}, that are predicted to exist around the putative beneficial allele. This procedure was also applied to African and North American *Drosophila melanogaster* population genomic data to detect candidate genes under ongoing positive selection. Upon visual inspection of sequence polymorphism, candidates detected by our CLR method exhibited clear haplotype structures predicted under incomplete selective sweeps. From these results we inferred the general rate and mode of positive directional selection operating in *Drosophila melanogaster* population.

6.4

The evolution of dispersal-related traits in invasive cane toads

Lee Rollins [1], Mark Richardson[1], Jack Reid[1], Rick Shine[2]
[1] *Deakin University, Geelong, VIC, Australia,* [2] *University of Sydney, Sydney, NSW, Australia*

The process of biological invasion exposes a species to novel pressures, both in terms of the environments it encounters and the evolutionary consequences of range expansion. Several invaders have been shown to exhibit rapid evolutionary changes in response to those pressures, thus providing robust opportunities to clarify the processes at work during rapid phenotypic transitions. The accelerating pace of invasion of cane toads (*Rhinella marina*) in tropical Australia during its 80-year history has been well-characterised at the phenotypic level, including common-garden experiments that demonstrate heritability of several dispersal-relevant traits. Individuals from the invasion front (and their progeny) show distinctive changes in morphology, physiology and behaviour that (in combination) result in far more rapid dispersal than is true of conspecifics from long-colonised areas. Remarkably, studies of neutral markers show extremely limited genetic diversity in this population. Here we discuss our analyses of differential gene expression in toads from both ends of this invasion-history transect; we have found substantial up-regulation of many genes on the invasion front, notably those involved in metabolism and cellular repair. Gene ontology enrichment analysis confirms the importance of genes underlying energy production and responses to environmental stressors in individuals from the range edge. Such pronounced phenotypic and gene expression differences in a population with extremely limited genetic diversity suggest a possible role for epigenetic control of these traits. We discuss these possibilities and our approach to teasing apart mechanisms underlying rapid evolution in this population.

6.5

Molecular-level response to selection in a mammalian laboratory model of a physiological and behavioral adaptation

Mateusz Konczal [1], Wieslaw Babik[1], Jacek Radwan[2], Edyta T. Sadowska[1], Pawel Koteja[1]
[1] Jagiellonian University, Krakow, Poland, [2] Adam Mickiewicz University, Poznan, Poland

Experimental evolution combined with genome or transcriptome resequencing is a promising approach for advancing our understanding of the genetic basis of adaptation. We applied this strategy to investigate the effect of selection on ecologically important traits in lines derived from a natural population of a small mammal. We analyzed the liver, heart and hippocampus transcriptomes of bank voles that had been selected in two directions: for increased aerobic metabolism (A) and for predatory behavior (P). The organs were sampled at 13th generation of selection; at that point, in each direction the voles from four replicate selected lines differed in the selected trait from four control lines by more than 2 standard deviations. Using RNA-seq we assessed gene expression and allele frequency differences between selected and control lines. We found that initial molecular-level response to selection in A lines occurs primarily via changes in gene expression. In the P lines changes of similar magnitude were observed in expression and allele frequencies. This suggests differences in genetic architecture of phenotypes selected in the A and P lines. Among genes which showed highest differentiation between selected and control lines we identified, using information about gene function and the biology of phenotypes under selection, plausible targets of selection. Because our selection lines were derived from a natural population, the amount and spectrum of variation available for selection probably closely approximated that typically found in populations of small mammals and thus the selection lines constitute an excellent model to study molecular basis of rapid adaptation.

6.6

Predicting evolution using high resolution lineage tracking

Sasha Levy [1,2], Jamie Blundell[1,2], Sandeep Venkataram[2], Dmitri Petrov[2], Daniel Fisher[2], Gavin Sherlock[2]

[1] Stony Brook University, Stony Brook, NY, USA, [2] Stanford University, Stanford, CA, USA

Evolution of large asexual cell populations underlies ~30% of deaths worldwide, including those caused by bacteria, fungi, parasites, and cancer. However, the dynamics underlying these evolutionary processes remain poorly understood because they involve many competing beneficial lineages, most of which never rise above extremely low frequencies in the population. Because of technical or cost limitations, clonal, single cell and population sequencing methods generally fail to characterize these low frequency mutations, and only provide a view of the "tip of the iceberg" of the beneficial mutations that impact the evolutionary dynamics. To observe these normally hidden beneficial mutations, we used random DNA barcodes to construct a sequencing-based ultra high-resolution lineage tracking system in *Saccharomyces cerevisiae* that allows us to monitor the relative frequencies of ~500,000 lineages simultaneously. I will present results from our first experimental evolutions with this system. In populations of 10^8 cells, we identify greater than ~20,000 independent beneficial mutations that occur within the first 100 generations, some of which never reach population frequencies above ~10^{-5}. In contrast with theoretical expectations, we find that the distribution of fitness effects of beneficial mutations is neither exponential nor monotonic. Early adaptation is a predictable consequence of this distribution and is strikingly reproducible. However, as numerous small-effect mutations are outcompeted by rarer large-effect mutations, variability between replicates begins. These results suggest that early evolutionary dynamics of infection, cancer, and drug resistance may be deterministic for a period of time before stochastic effects become important.

6.7

Extreme local adaptation in Drosophila chemosensory perception

J. Roman Arguello [1,2], Margarida Cardoso-Moreira[1,2], Jaaved Mohammed[3], Jennifer K. Grenier[2], Srikanth Gottipati[2,4], Andrew G. Clark[2], Richard Benton[1]

[1] Center for Integrative Genomics, UNIL, Lausanne, Switzerland, [2] Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA, [3] of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, USA, [4] Translational Medicine and Think Team, Otsuka Pharmaceutical Development and Commercialization, Inc., Princeton, NJ, USA

How organisms adapt to new environments is of fundamental biological interest but is poorly understood at the level of genes and neurons. A particularly interesting question regarding local adaptation is how sensory systems and perception are altered during this process. To investigate the extent to which selection has shaped the chemosensory perception within *D. melanogaster* - a preeminent model for studying sensory neurobiology – we have analyzed genome-wide polymorphism and divergence data from the Global Diversity Lines (84 *D. melanogaster* genomes derived from 5 geographically diverse populations). By couching population genomic analyses of chemosensory protein families within parallel analyses of all other large protein families we demonstrate that, surprisingly, chemosensory proteins do not stand out as outliers with respect to signals of adaptive differences between species. However, chemosensory proteins do experience local adaption at extremely high rates, often displaying the strongest signals of selection among large protein families. We show that adaptation has operated almost exclusively on standing variation, and that positively selected genes often harbor unanticipated levels of diversity. Our curated set of selected chemosensory proteins is currently guiding functional studies aimed at understanding the phenotypic impacts of the variants under selection.

6.8

Distinguishing modes of convergent adaptation in genomic data

Kristin Lee, Graham Coop
University of California, Davis, Davis, California, USA

Species often adapt to similar selection pressures across their range leading to convergent evolution. We want to distinguish between the distinct ways this can occur. Selected alleles present in multiple subpopulations at the same loci can have multiple independent mutational origins. Alternatively, adaptation in different populations could proceed by means of selection on the same standing variation, or a single allele spread throughout the subpopulations via gene flow. Positive selection impacts neutral diversity at linked loci due to hitchhiking, which can be modeled as an increase in the variance in allele frequencies within a population. We develop theory to show how shared hitchhiking events between subpopulations act to increase covariance in allele frequencies at loci near the selected site in contrast to independent selective sweeps. We incorporate this hitchhiking effect into a simple multivariate model of allele frequencies under population structure. Based on this theory, we present a composite likelihood-based method utilizing genomic SNP data to identify and distinguish between different modes of convergent evolution across multiple subpopulations. We illustrate our method applying it to various genome-wide datasets.

6.9

Phenotypic adaptation due to recent environmental change

Susanne P. Pfeifer [1], Stefan Laurent[2], Matthieu Foll[2], Hopi E. Hoekstra[3], Erica B. Rosenblum[4], Jeffrey D. Jensen[2]

[1] School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [2] School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [3] Department of Organismic and Evolution Biology, Harvard University, Cambridge, USA, [4] Deptartment of Environmental Science, Policy and Management, University of California, Berkeley, USA

Within the past 7,000 years, a drastic environmental change caused the creation of the world's largest white gypsum dune field, the White Sands National Monument in New Mexico (USA) - a geologically young formation and a rare place to study recent adaptation in natural populations. Two species of lizards, the Eastern Fence Lizard (*Sceloporus undulatus*) and the Little Striped Whiptail (*Aspidoscelis inornata*), have independently and rapidly evolved cryptic, blanched, substrate-matched phenotypes at White Sands (compared to the brown, ancestral phenotype found on the dark adobe soils in the surrounding Chihuahuan Desert) in order to escape predation. Previous studies have linked this adaptive pigmentation to mutations in the MC1R gene, which plays a critical role in the production of vertebrate melanin pigments. We have sequenced the genome of 24 individuals (9 individuals from the light and 15 individuals from the dark populations) for each of the two species to investigate the adaptive history both at the MC1R locus as well as at a large number of neutral loci across the genome, enabling us to address fundamental questions about the extent and nature of adaptive evolution. Observed patterns of variation are consistent with a strong selective sweep at the MC1R locus in *A. inornata* but not in *S. undulatus*, indicating that different adaptive histories are responsible for their similar adaptive phenotype. Furthermore, we inferred the neutral demographic history using a recently developed maximum likelihood framework and used these results to compare the evolutionary process between light and dark populations in both species.

6.10

Adaptive convergence and quasi-heterozygous advantage after experimental evolution of *Escherichia coli*

Christian Sievert [1,2], Taylor Loeffler[2], Larry Panyon[2], Chandler Morris[2], David Winter[1], Kael Dai[1], Xuan Wang[2], Reed A. Cartwright[1,2]

[1] The Biodesign Institute, Arizona State University, Tempe, AZ, USA, [2] School of Life Sciences, Arizona State University, Tempe, AZ, USA

A laboratory-developed *Escherichia coli* strain was adapted to improve the conversion of xylose to lactate using experimental evolution. Three populations were evolved independently, and after approximately 20 generations a significant increase in substrate consumption was detectable. To determine the genetic mechanism for this quick adaptation, the genome of a representative clone for each adapted population was sequenced, and mutations were identified by comparison to the ancestral strain. In one population, a point mutation is present in one copy of a duplicated xylose regulator (*xylR*). Notably, *xylR* is part of a 32 kb genomic region that is duplicated in the ancestral strain but not in wild-type *E. coli* strains. Preliminary results indicate that the evolved strain with one mutated *xylR* allele has a higher fitness than strains with either two wild-type or two mutated alleles. This is analogous to heterozygote advantage in diploid organisms and is an example of duplication being a prerequisite for evolution of novel molecular phenotypes. Interestingly, in a second population, *xylR* was independently mutated at a different location, and again it happened in only one of the alleles. By doing competition assays, we are able to determine an empirical fitness landscape of the *xylR* duplication and mutations. In the third population, a point mutation occurred in cAMP receptor protein (*crp*). Both *xylR* and *crp* are transcription factors in the same regulatory network, and transcript levels indicate that all three populations have converged to upregulate the expression of genes involved in xylose utilization.

6.11

Within-host adaptation of Mycobacterium tuberculosis

Mary O'Neill, Tatum Mortimer, Caitlin Pepperell
University of Wisconsin-Madison, Madison, WI, USA

Tuberculosis is a fascinating, complex system in which to study adaptation of a fast evolving clonal parasite subsisting in an adversarial relationship with its host. Unlike the vast majority of bacterial pathogens, *Mycobacterium tuberculosis* is restricted to its niche as an agent of disease. During tuberculosis infections, bacterial populations expand from a tiny inoculum to populations numbering in the billions, in an environment that is harsh and dynamic. *M. tuberculosis* evolves clonally, to an extent that is exceptional relative to other bacteria. Published work suggests that *M. tuberculosis* drug resistance mutations arise *de novo*, and complex patterns of clonal replacements have been observed during within-host evolution. Our research suggests that positive selection is an important driver of *M. tuberculosis* evolution within hosts. Beyond drug resistance loci, we find evidence of adaptation in genes with roles in regulation, synthesis and transportation of M. tuberculosis cell envelope lipids. We speculate that rapid adaptation of cell envelope lipids is facilitated by functional redundancy, flexibility in their metabolism, and their roles mediating interactions with the host.

6.12

The impact of gene flow, standing genetic variations and strong artificial selection during pig domestication

Laurent Frantz [1,2], Joshua Schraiber[3], Vincent Battista[4], Martien Groenen[2], Greger Larson[1,4]
[1] The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK, [2] Animal Breeding and Genomics Group, Department of Animal Sciences, Wageningen University, Wageningen, The Netherlands, [3] Department of Genome Sciences, University of Washington, Seattle, USA, [4] Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham, UK

Artificial selection, during domestication, has resulted in drastic phenotypic homogenisation (*i.e.* all domestic pigs have extra vertebrae compare to wild boars) as well as phenotypic diversification (*i.e.* coat colour) within a single species. Moreover, recent archaeological and genomic studies have proposed that farmers did not imposed a strong reproductive isolation (allowed for gene flow) between wild and domestic forms during the domestication of multiple species, including pigs. This suggests that genetic variations, standing in diverse wild populations, may have played a role in these dramatic phenotypic alterations. Indeed, the introgression of standing genetic variations, from diverse wild populations into different domestics breeds must have provided farmers with an influx of genetic resource, increasing the potential for phenotypic diversification. However, while the persistence of gene flow during domestication may have been beneficial for phenotypic diversification it may also have had a negative impact on phenotypic homogenisation among breeds at traits that differentiate domestic and wild forms.

Here, I will assess these questions using a novel data set of over 100 pig genomes. I will first demonstrate, using approximate Bayesian computation, how domestication in pigs resulted from a complex interplay between gene flow from diverse populations of wild boars and strong artificial selection. Secondly, I will show how artificial selection has counteracted the effect of gene flow and resulted in the homogenisation of productively important traits such as body length. Lastly, I will demonstrate how farmers also utilised gene flow to diversify phenotypes, taking the example of coat colours in pigs.

6.13

The role of recombination in evolutionary rescue

Hildegard Uecker [1], Joachim Hermisson[2]
[1] *IST Austria, Klosterneuburg, Austria,* [2] *University of Vienna, Vienna, Austria*

Environmental change, if severe, can drive a population extinct unless the population succeeds in adapting to the new conditions. How likely is a population to win the race between population decline and adaptive evolution? We consider the probability of evolutionary rescue after a sudden deterioration of the environment when survival of the population is contingent on mutations at multiple (two) loci. The rescue genotype can hence arise by two-step mutation or by recombination from single mutants. Such a situation arises, for example, in combination drug therapy or the application of herbicide mixtures.

Recombination generically has two opposing effects: it generates and it breaks up favorable gene combinations. We investigate how the fitness effects of mutations as well as epistasis before and after the switch in the environment affect the interplay of these two effects. We furthemore consider how these factors influence the most likely pathway to rescue (two-step mutation or recombination, standing genetic variation or de-novo generation of mutant genotypes). A special emphasis is on the effect of the population dynamics on the rate of adaptation. We find that, even in the absence of density-dependent fitness, a fast eradication of the wildtype population can enhance rescue in the presence of recombination. On the other hand, recombination can prevent rescue when the wildtype disappears slowly.

6.14

The genomic signature of stabilizing selection during temperature adaptation in experimental *Drosophila melanogaster* populations

Susanne Franssen, Robert Kofler, Viola Nolte, Christian Schlötterer
Vetmeduni Vienna, Vienna, Austria

The genetic architecture of local adaptation in natural populations is not yet understood: adaptation is either achieved by a few loci with locally favored alleles or by many quantitative trait loci responding to a changed fitness optimum. In time series data both adaptive processes leave their characteristic signature in the genome.

To evaluate the relative importance of both processes, we performed whole-genome sequencing of replicated *D. melanogaster* populations subjected to a new hot temperature environment. With data from a high temporal resolution for almost 90 generations in the hot environment and a reversed selection for 42 generations we did not find evidence for locally adapted alleles sweeping through the population towards fixation. Rather the genomic response matched expectations for a quantitative trait under stabilizing selection with a new fitness optimum: selected haplotypes started from a low frequency and rapidly increased in frequency in all replicates, however, in later generations, they either plateaued, decreased or diverged between replicates. None of the selected haplotypes rose to a frequency higher than 0.6. In the reversed selection, the frequency these haplotypes decreased again, but did not reach the starting frequency.

Our results suggest that temperature related fitness in laboratory natural selection experiments is a polygenic quantitative trait under stabilizing selection.

6.15

The 1002 *Saccharomyces cerevisiae* genomes project: a framework for genome-wide association studies

Jackson Peter[1], Anne Friedrich[1], Matteo De Chiaria[2], David Pflieger[1], Anders Bergstrom[2], Anastasie Sigwalt[1], Kelle Freel[1], Gianni Liti[2], Joseph Schacherer [1]

[1] *University of Strasbourg / CNRS, Strasbourg, France,* [2] *University of Nice / CNRS, Nice, France*

Genome-wide investigation of the patterns of polymorphism in a large sample of individuals is the first step to assess the relationship between genotype and phenotype within a species. To date, yeast population genomics only focused on a limited number of isolates. Because of their small and compact genome, yeasts and more precisely *S. cerevisiae* represent a powerful model for population genomics. Here, we will present our project: "The 1002 yeast genomes project: a framework for genome-wide association studies". We sequenced more than 1000 genomes of *S. cerevisiae* using the Illumina HiSeq technology, with a mean coverage of 250X. Due to the broad diversity of strains selected for sequencing, this population genomics dataset reveals an accurate picture of the genomic variation within *S. cerevisiae*. Indeed, the next generation sequencing data are suitable to reveal the entire repertoire of single nucleotide polymorphisms (SNPs) as well as the degree of copy number variation (CNVs). We therefore expand the current catalogue of SNPs described in *S. cerevisiae* so far. Furthermore, our strategies to sequence diploid strains allow us to fully characterize the spectrum of CNVs (such as aneuploidies) and the degree of heterozygosity. Furthermore, we performed extensive phenotyping experiments with our strain collection. The high SNPs density allowed us to perform genome-wide association studies. This dataset leads to the identification of a large set of functional polymorphisms that underlie phenotypic variation.

6.16

Genetics of Jaw Divergence in a Trophically Polymorphic Cichlid Fish

C. Darrin Hulsey
University of Konstanz, Konstanz, Germany

Trophically polymorphic species could represent lineages that are rapidly diverging along an ecological axis or could phenotypically mark the collapse of species through introgressive hybridization. We investigated genomic patterns of introgression between the trophically polymorphic cichlid fish *Herichthys minckleyi* and its relative *H. cyanoguttatus* using a combination of population genomics and species tree analyses. Using rad-tag sequencing, we also investigated whether hybridization could explain the variation in the jaws of this phenotypically variable cichlid. We examine the molecular evolution of several critical protein coding loci to determine if they exhibit the signature of adaptive divergence.

117A

Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum

Sebastian Matuszewski [1,3], Joachim Hermisson[1,4], Michael Kopp[2]
[1] University of Vienna, Vienna, Austria, [2] Aix Marseille Université, Marseille, France, [3] EPFL, Lausanne, Switzerland, [4] Max F. Perutz Laboratories, Vienna, Austria

Numerous studies have tried to provide a formal framework for the description of the adaptive process. Out of these, two complementary modelling approaches have emerged: While so-called adaptive-walk models consider adaptation from the successive fixation of de-novo mutations only, quantitative genetic models assume that adaptation proceeds exclusively from pre-existing standing genetic variation. The latter approach, however, has focused on short-term evolution of population means and variances rather than on the statistical properties of adaptive substitutions.

Our aim is to combine these two approaches by describing the ecological and genetic factors that determine the genetic basis of adaptation from standing genetic variation in terms of the effect-size distribution of individual alleles. Specifically, we consider the evolution of a quantitative trait to a gradually changing environment.
By means of analytical approximations, we derive the distribution of adaptive substitutions from standing genetic variation, that is, the distribution of the phenotypic effects of those alleles from the standing variation that become fixed during adaptation.

We find that, compared to adaptation from de-novo mutations, (i) adaptation from standing variation proceeds by the fixation of more alleles of small effect; (ii) populations that adapt from standing genetic variation remain better adapted and can traverse larger distances in phenotype space if the rate of environmental change is fast rather than slow.

118B

Genome sequencing reveals microgeographic adaptation in the widespread Amazonian tree species Eperua falcata (Aublet.): A Bayesian framework.

Louise Brousseau [1,2], Giovanni Giuseppe Vendramin[3], Ivan Scotti[1,2]
[1] INRA, UMR URFM, Avignon, France, [2] INRA, UMR EcoFoG, Kourou, French Guiana, [3] CNR, IBBR, Florence, Italy

In wild plant populations, genetic divergence within continuous stands is common, sometimes at very short geographical scales. While genome-wide divergence is driven by neutral factors which affect the genome in a homogeneous manner, natural selection affects a reduced fraction of targeted genomic regions. This provides a conceptual framework for the discrimination of neutral and adaptive sources of molecular divergence in wild populations. However, properly testing the hypothesis of local adaptation requires replicating the environmental gradient tested, and developing modelling tools (i) able to deal with enormous ‘next-generation' datasets and (ii) taking into account hierarchical (neutral) structuring.

Amazonian rainforest is the world's greatest store of biodiversity, but the evolutionary processes responsible for its maintenance are still poorly understood. Despite the apparent floristic homogeneity of forest landscapes when considered at large spatial scales (continent, region), the structure of plant communities is strongly affected by habitat patchiness over very local scales.

A canopy-dominant tree species widespread in Amazonia was chosen to test the hypothesis of local adaptation at intra-specific level. A genome-scan (~67,000 SNPs) was realized through genome sequencing of populations pools, and a Bayesian model was developed to detect footprints of local adaptation at both regional (~300km) and local scales (~300m). It provides strong evidence of local adaptation over very local scales in spite of extensive gene flow between populations. Such evidence of microgeographic adaptive divergence reinforces the idea that local adaptation would be a key evolutionary process in Amazonia, probably involved in the maintenance of diversity.

119C

An improved evolve and resequence design facilitates discovery of adaptive loci in experimental *Drosophila* populations

Ray Tobler [1,2], Viola Nolte[1], Christian Schlötterer[1]
[1] *Institut für Populationsgenetik, Wien, Austria,* [2] *Vienna Graduate School of Population Genetics, Wien, Austria*

Evolve and resequence (E&R) is a promising new population genomic approach to detect loci responsible for quantitative trait variation and to determine the dynamics of adaptation. To date, however, E&R studies using *D. melanogaster* routinely report vastly more selected loci than are expected under population and quantitative genetic theory. Here we report results from a new E&R study where replicated *Drosophila* populations were subjected to opposing thermal habitats, which uses an updated design inspired by findings from previous empirical and simulation-based studies. We demonstrate that these modifications, which include using an inversion-free species (*D. simulans*), larger founder populations and increased replication, result in more distinct signals of adpative change.

120D

The effect of background selection on evolutionary genetic patterns of human influenza virus H3N2

Kangchon Kim, Yuseob Kim
Ewha Womans University, Seoul, Republic of Korea

Evolutionary genetic studies of human influenza H3N2 have mainly focused on the effect of positive selection as explanation for the limited diversity and genealogical structure of HA segment. However, the evolutionary patterns shaped by negative selection, which removes deleterious mutations that are universal over both antigenic and non-antigenic segments, remain largely unknown. Here, we investigated the impact of background selection on the level of variation and the mode of selective sweep in H3N2 viral segments, using computer simulation of viral population undergoing recurrent selective sweeps. First, we demonstrate that low synonymous diversity of HA segment cannot be explained by positive selection alone, but only with negative selection operating on most nonsynonymous sites. Background selection substantially reducing the effective population size also gives parsimonious explanation for lack of soft sweep involving antigenic cluster changing mutants in HA despite large population size of the virus. Next, we found that per-site synonymous diversity of each genomic segment does not decrease with segment length, although longer fragment carries more sites subject to deleterious mutations. This implies that reassortment rate of H3N2 is low such that negative selection operating on one segment reduces genome-wide genetic variation. Finally, we investigated at which reassortment rate the observed between-segment patterns of genomic variation, where polymorphism in HA segment is particularly low, can be explained by the joint effect of background selection and selective sweeps.

121A

Rapid evolution of phenotypic plasticity during experimental evolution of *Drosophila*

Eszter Ari [2,1], Viola Nolte[1], Ray Tobler[1,3], Christian Schlötterer[1]

[1] Institut für Populationsgenetik, Vetmeduni, Vienna, Austria, [2] Department of Genetics, Eötvös Loránd University, Budapest, Hungary, [3] Vienna Graduate School of Population Genetics, Vienna, Austria

Phenotypic plasticity is the ability of a given genotype to express different phenotypes in response to environmental conditions. Hence, it has been proposed that phenotypic plasticity can facilitate the survival of populations in changing environments. Here, we studied the plasticity of gene expression and its evolution in *Drosophila simulans* populations that were exposed to new environments, differing in temperature. Gene expression was highly plastic, with 37% of the expressed genes differing significantly between warm (23°C) and cold (15°C) environments before adapting to the new temperature regimes. After about one year of adaptation the evolved populations were still highly plastic - 36% of the genes were differentially expressed due to various developmental temperature in both warm and cool adapted but 42% of the genes changed their plasticity during the experiment. Most genes in cool adapted flies (20°C / 10°C day and night temperature for 12-12h) kept their plasticity at a similar level or became less plastic. In the warm adapted populations (28°C / 18°C day and night temperature), however, one third of all expressed genes changed the direction of plasticity. These results show that plasticity in gene expression is evolving rapidly in response to novel environments, suggesting that change in plasticity is adaptive rather than plasticity *per se*.

122B

Genetic differentiation, transcriptome variation, and local adaptation of dominant prairie grass *Andropogon gerardii* along the precipitation gradient of the US Midwest: Implications for climate change

Loretta Johnson [1], Miranda Gray[1], Paul St Amand[1], Matthew Galliart[1], Susan Brown[1], Nora Bello[1], Jesse Poland[1], Karen Garrett[1], Eduard Akhunov[1], Nora Bello[1], Ted Morgan[1], Sara Baer[2], Brian Maricle[3]

[1] Kansas State University, Manhattan KS, USA, [2] Southern Illinois University, Carbondale IL, USA, [3] Forth Hays State University, Hays KS, USA

Big bluestem (*Andropogon gerardii*) is an ecologically dominant grass in Central US, comprising 70% biomass of grasslands. With wide distribution across a precipitation gradient (400-1200mm yr-1), we expect ecotypic variation in drought tolerance and local adaptation. We investigated phenotypic variation, genetic diversity and transcriptional variation using reciprocal gardens across the Midwest gradient (wet to dry) in Illinois, Manhattan, Hays and Colby KS to predict response to future climate change. Three ecotypes from four populations sourced from Central KS, Eastern KS, and Illinois were reciprocally planted in polyculture (with other species). A phenotypic cline in drought tolerance of ecotypes was observed indicating local adaptation of xeric and mesic ecotypes. Because genetic diversity may be critical to adjust to climate change, we assessed genetic diversity and differentiation of the source populations. Seven ecotype-specific loci under diversifying selection were identified with Bayescan and Bayenv2 and marker frequency related to seasonal precipitation and temperature using multivariate logistic regression. Based on BEDASSLE analyses, differentiation among populations was due predominantly to different ecological environments of the source populations (IBE) rather than isolation by distance (IBD). We used RNAseq to compare transcriptional variation of Central KS and Illinois ecotypes growing under abiotic stress in xeric Colby KS. The Illinois ecotype mounted a weak transcriptional response, consisting of abiotic stress genes as an inducible, stress strategy. In contrast, the CKS ecotype shows a strong and varied transcriptional response that matches its increased competitiveness under drought. We provide insights into how this dominant grass may withstand climate change.

123C

Trade-offs in immunity in the metal hyperaccumulator *Noccaea caerulescens*

Anja C. Hoerger [1,2], J. Andrew C. Smith[2], Gail M. Preston[2]

[1] Department of Ecology and Evolution, University of Salzburg, Salzburg, Austria, [2] Department of Plant Sciences, University of Oxford, Oxford, UK

Metal hyperaccumulating plants are able to accumulate exceptionally high concentrations of heavy metals in their shoots to levels that would be toxic to most other plant species. This trait has evolved independently multiple times in the plant kingdom. Although our understanding of the molecular mechanisms involved in metal uptake and tolerance has improved, not much is known about the processes that have led to the evolution of metal hyperaccumulation in plants. Recent studies have provided new insight into the ecological and evolutionary significance of this trait by showing that the metal hyperaccumulating plant *Noccaea caerulescens* can use high concentrations of accumulated metals to defend itself against attack by pathogenic microorganisms. Interestingly, infected *N. caerulescens* plants show none of the inducible defence responses that are used by most plants to provide protection against infection, which suggests that it relies on accumulated metal for disease resistance. The fact that these plants have evolved the ability to uptake and store metals in their shoot tissue, but have in turn lost defences common to most plants suggests a trade-off in expressing both traits. We studied the evolutionary, ecological and functional processes involved in the gain of metal hyperaccumulation and loss of other defensive traits in *N. caerulescens*. Genes involved in the trade-off were identified and analysed using a combined phenotyping and population genomic strategy. Our results provide new insights into the evolution and ecology of metal hyperaccumulation and contribute to the understanding of how plant adaptation to biotic and abiotic stress may be connected.

124D

Silkworm evolution during domestication and breeding processes: genetic and epigenetic perspective

Hui Xiang [1], Xin Li[1], Fangyin Dai[2], Jun Wang[3], Wen Wang[1]
[1] Kunming Institue of Zoology, Kunming, China, [2] Southwest University, Chongqing, China, [3] BGI-Shenzhen, Shenzhen, China

Silkworm originated from the wild silkworm about 5000 years ago and was further introduced abroad along the well known “silk road”. Exploration on the mechanisms on silkworm evolution during domestication and breeding is important for both understanding the nature of artificial selection and informing future efforts in improvement of silk products. Based on the silk gland methylomes we previously generated, we have comprehensively compared silk gland methylomes between domesticated and wild silkworms and identified several genes with different methylation and expression level, which possibly function in metabolism and silkgland development, implying epigenetic influences during domestication. Now we are systematically re-sequencing (12×) 136 domesticated silkworm core strains representing local trains in China (including trimoulting and tramoulting types), Japan, Europe and tropics, respectively; improved strains of Chinese and Japanese type, respectively, as well as 7 geographically different wild silkworm individuals. Combined with the released data, we detected 42 M SNPs, accounting for 10% of the genome. Phylogenic tree indicated the ancient Chinese trimoulting local strain, the obviously derived European local stain and genetic mixture of other local strains. Improved strain of Chinese and Japanese type seems independently evolved, implying that overall similar artificial selection might act on different genetic loci. We are analyzing selection signals and, identifying QTLs with cocoon traits and resistance by GAWS. The studies are promising to provide resources and cues for understanding the history of “silk road”, artificial selection and genomics-enabled improvements in silkworm breeding in both genetic and epigenetic perspective.

125A

Genetic evidence for the Tibetan origins and recent population expansion in Sherpa populations in Himalayas

Sushil Bhandari[1,4], Chaoying Cui[2], Xiaoming Zhang[1], Yi Peng[1], Kun Xiang[1,4], Deying Yang[1,4], Yibo He[1,4], Guoquan Dang[1], Ouzhu Luobu[2], Tianyi Wu[3], Xuebin Qi [1], Bing Su[1]
[1] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China, [2] High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, Tibet, China, [3] National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining, Qinghai, China, [4] University of Chinese Academy of Sciences, Beijing, China

Previous studies suggested that Tibetans originated from an admixture population between Sherpa and Han Chinese populations, and the high frequency of those Sherpa-specific mtDNA haplotypes was resulted from their superior adaptation to hypoxic environment. To verify this conclusion and better understand the demographic history of Sherpa population, we collected 582 unrelated Sherpa samples across their geographic ranges, and screened mitochondrial and Y-chromosomal haplogroups profiles. The Y chromosomal haplogroup D-M174, a macrohaplogroup that is dominant in Tibetan populations and dated to 30 thousand years before present, was absent in several different indigenous population in Nepal but highly occurred exclusively in Sherpa populations (42.59%). The mtDNA analysis further revealed that A15, C4a2'3'4 and M9a1a1c1b, three major founding lineages originated from Tibetan population, contributed to form the genetic diversity in Sherpa populations. The phylogeographic analyses of both paternal and maternal lineages showed a closer affinity of Sherpa with Tibetans than other surrounding Nepalese population. The star like pattern seen in dominant haplogroups (A15c1, C4a3b1 and M9a1a1c1b1a) of mtDNA indicates their rapid population expansion recently. The younger age (700 to 1500 years) of Sherpa specific lineages provide evidence that it hasn't gone under sufficient time for selection. Thus we proposed Sherpa as a descendent of Tibetan populations and their ancestors already acquired high altitude adaptive features in Tibetan plateau before they migrated westward to the Himalayan region where they permanently settled at present.

126B

Evolutionary genetic dissection of innate immunity in humans

Matthieu Deschamps [1,2], Guillaume Laval[1,2], Etienne Patin[1,2], Lluis Quintana-Murci[1,2]
[1] Institut Pasteur, Unit of Human Evolutionary Genetics, Paris, France, [2] Centre National de la Recherche Scientifique, CNRS URA3012, Paris, France

Innate immunity represents the first line of host defense against infection, through the recognition of microorganisms and activation of semi-specific responses. Innate immunity genes (IIGs) therefore provide an excellent model to study how pathogens have exerted pressure on the host genome over time. Here, we used population genomics approaches to assess the impact of natural selection and hominin introgression on IIGs in modern humans. We compiled a comprehensive list of ~1500 IIGs and classified them based on the role of the encoded proteins as (i) sensors of microorganisms, (ii) signal transducers, (iii) transcription (co-)factors, (iv) effectors produced following pathogen recognition and (v) secondary receptors. Using the 1000 Genomes dataset, we first show that IIGs are globally more constrained than the remainder of protein coding genes of the genome. Signal transducers and transcription (co-)factors display the strongest evidence of purifying selection. Among microbial sensors, the family of Cytosolic Nucleic Acid Sensors is the most constrained, highlighting the biological importance of this recently defined class of receptors. We also find that microbial sensors are the class of IIGs preferentially targeted by positive selection, given their signals of strong population differentiation. Finally, our study shows that IIGs are enriched in SNPs exhibiting the highest probability of Neanderthal ancestry. Notably, the TLR1-6-10 cluster, previously found to evolve adaptively in Europeans and Asians, is among the genic regions most likely to derive from Neanderthals. These results provide new insights into the recent evolution of the innate immune system in anatomically modern humans.

127C

A substantial number of adaptive mutations are lost to Hill-Robertson Interference in Drosophila.

David Castellano[1], Maite Barron[2], Marta Coronado[1], Jose Campos[3], Antonio Barbadilla[1], Adam Eyre-Walker[4]

[1] Universitat Autònoma de Barcelona, Barcelona, Spain, [2] CSIC-Universitat Pompeu Fabra, Barcelona, Spain, [3] University of Edinburgh, Edinburgh, UK, [4] University of Sussex, Brighton, UK

Hill-Roberston interference is expected to reduce the number of adaptive substitutions in regions of the genome with low rates of recombination. To investigate this penomenon we have estimated the rate of adaptive amino acid substitution using a McDonald-Kreitman type approach that corrects for the influence of slightly deleterious mutations. We find that the rate of adaptive substitution is positively correlated to the rate of recombination but that the relationship is non-linear - the rate of adaptive evolution asymptotes. Using the observed relationship between rate of adaptive evolution and the rate of recombination we estimate that ~25% of advantageous mutations, which would oteherwise be destined for fixation, are lost through Hill-Robertson interference.

128D

Detecting recent selective sweeps while controlling for mutation rate and background selection.

Christian D. Huber [1], Michael DeGiorgio[2,3], Ines Hellmann[4], Rasmus Nielsen[5]
[1] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California, USA, [2] Department of Biology, Pennsylvania State University, University Park, Pennsylvania, USA, [3] Center for Medical Genomics, Pennsylvania State University, University Park, Pennsylvania, USA, [4] Department Biologie II, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany, [5] Departments of Integrative Biology and Statistics, University of California, Berkeley, California, USA

A composite likelihood ratio test implemented in the program SweepFinder is a commonly used method for scanning a genome for recent selective sweeps. SweepFinder uses information on the spatial pattern of the Site Frequency Spectrum (SFS) around the selected locus. To avoid confounding effects of background selection and variation in the mutation process along the genome, the method is typically applied only to sites that are variable within species. However, the power to detect and localize selective sweeps can be greatly improved if invariable sites are also included in the analysis. In the spirit of a Hudson-Kreitman-Aguadé-test, we suggest to add fixed differences relative to an outgroup to account for variation in mutation rate, thereby facilitating more robust and powerful analyses. We also develop a method for including background selection modeled as a local reduction in the effective population size. Using simulations we show that these advances lead to a gain in power while maintaining robustness to mutation rate variation. Furthermore, the new method also provides more precise localization of the causative mutation than methods using the spatial pattern of segregating sites alone.

129A

Evolution and putative adaptation in the invasive Florida Burmese python population inferred using genome-wide RADseq data

Daren Card [1], Drew Schield[1], Kristen Hart[2], Margaret Hunter[3], Todd Castoe[1]
[1] University of Texas at Arlington, Arlington, Texas, USA, [2] United States Geological Survey, Davie, Florida, USA, [3] United States Geological Survey, Gainesville, Florida, USA

Substantial evolutionary changes are thought to happen over long time periods in most species, making analyses of natural selection and its effects on the genome difficult.

Invasive species, however, represent a promising model for analyzing the processes of evolution and adaptation on timescales that are tractable for study, and have been shown to demonstrate rapid evolutionary responses over short, 'ecological' timescales. Such invasive introductions often demonstrate rapid responses to this shift in environmental conditions and habitat, and present ready opportunity to test the genome-wide effects of natural selection and putative adaptation. The Burmese python (*Python molurus bivittatus*) is ideal for this work due to its recent establishment and proliferation in Florida, a location with climatic conditions much different from those in the species' native range of Southeast Asia. A 2010 freeze event in Florida led to a large die-off of snakes in the Florida Burmese python population (FPP), and likely catalyzed selection-driven evolution. We used discrete population-level sampling of the FPP before (2007) and after (2013) the freeze event and genome-wide marker sequencing (RADseq) to test the hypothesis that large fluctuations in allele frequencies (i.e., evolution) have occurred in the FPP as a result of the freeze event. We found multiple regions of the genome that show major shifts in heterozygosity in just a single generational time period, indicating *in situ* evolution in the FPP. Leveraging the Burmese python genome to identify genes and associated functions linked to these apparently selected loci indicates putative adaptation in the invasive population.

130B

Patterns of Genetic and Haplotypic Variation surrounding a Sweep from Standing Variation

Jeremy Berg, Graham Coop
University of California, Davis, Davis, CA, USA

Perhaps the most distinctive population genetic signature of recent positive selection is the selective sweep, in which a single new mutation fixes in the population, carrying its haplotype along with it, and removing genetic variation in a region surrounding the adaptive site.

Recent empirical and theoretical work has emphasized the fact that natural selection may often call upon standing variation to fulfill an adaptive need, in which case the population genetic signals are more subtle.

While some of these signals have been investigated via simulation, an analytical description of the hitchhiking signature expected when a single IBD genetic variant sweeps from standing variation is lacking.

Here, we use coalescent theory to show that patterns of genetic variation surrounding a sweep from standing variation can be predicted from a composition of the Ewens Sampling Formula and the standard approximation for de novo selective sweep. Our approximation allows for a highly accurate prediction of the expected frequency spectrum of a neutral variant linked to an allele which has just swept from standing variation, as well as clarifying the nature of haplotype frequency variation following a standing sweep.

131C

Genomic Convergence in the adaptation of Dogs and Humans

Guo-Dong Wang [1], Weiwei Zhai[2], Chung-I Wu[3], Peter Savolainen[4], Ya-Ping Zhang[1]
[1] Kunming Institute of Zoology, Kunming, China, [2] Genome Institute of Singapore, Singapore, Singapore, [3] University of Chicago, Chicago, USA, [4] KTH Royal Institute of Technology, Solna, Sweden

One of the man's closest friend, dog has followed human beings and travelled across a wide variety of ecological niches. Do dogs and humans adapt by the same genetic mechanism is an interesting question. Here we performed two whole genome studies of dogs, one is for the positive selection during domestication, the other is high-altitude hypoxic adaptation. For the first one, population genetic analysis identifies a list of genes under positive selection during domestication, which overlaps extensively with the corresponding list of positively selected genes in humans. Convergent evolution is most apparent in genes for digestion and metabolism, neurological process and cancer. The study draws together humans and dogs in their recent genomic evolution. The second one, we found that the hemoglobin levels are very similar between the two groups, suggesting that Tibetan dogs might share similar adaptive strategies as the Tibetan people. Through a whole-genome sequencing approach, we have identified EPAS1 and HBB as candidate genes for the hypoxic adaptation, and shows a significant convergence between humans and dogs in Tibet. Convergent evolution happening in two species bestows on us an unprecedented opportunity to understand these traits by studying the evolution and the phenotypes in both species simultaneously. Our best friend in the animal kingdom might provide us with one of the most enchanting systems for illuminating our understandings of human evolution and disease.

132D

Epigenetic divergence and parallel rapid adaptation in *Heliosperma pusillum* (Caryophylaceae)

Emiliano Trucchi [1], Ruth Flatscher[1], Maria Lorenzo Romero[1], Bozo Frajman[2], Peter Schönswetter[2], Ovidiu Paun[1]

[1] *University of Vienna, Vienna, Austria,* [2] *University of Innsbruck, Innsbruck, Austria*

Epigenetic modifications are expected to occur at a faster rate than genetic mutations, and to be a flexible way to respond to stress conditions and to rapidly generate variation that can be visible to natural selection, even in the absence of genetic mutations. We can then suppose epigenetics to be particularly important in the early phases of adaptation and divergence. The alpine plant *Heliosperma pusillum* and its recently and recurrently diverged, low-montane sibling *H. veselsky* constitute a perfect model system to test this hypothesis. They occur in distinct ecological conditions with divergent morphologies (dense trichomes versus glabrous plants) that are stable and heritable in common garden settings. However, genome-wide data (AFLP, RAD sequencing) fail to group them according to observed phenotypes and to unveil the genetic bases of the ecological and morphological switch. Employing a novel approach (bisulfite-converted RAD sequencing - bsRADseq), we screened the DNA methylation patterns across more than 200,000 nucleotide positions in six population pairs of the two ecotypes. Comparing by sequencing the DNA methylation state and context (CpG, CHG, CHH) in 120 individuals, we demonstrate that epigenetic patterns are much more conserved across representatives of the two ecotypes than predictions given the distinct native environments and that several candidate loci for convergent epigenetic modification during parallel adaptation to a similar environment are present.

133A

Trans-specific polymorphisms in Arabidopsis defence response genes

Polina Novikova [1], Nora Hohmann[2], Marcus Koch[2], Magnus Nordborg[1]
[1] Gregor Mendel Institute, Vienna, Austria, [2] Centre for Organismal Studies (COS) Heidelberg, Heidelberg, Germany

Genetic drift and disruptive selection diminish shared variation between sister species, when continuing gene flow and recurrent mutations can partially restore it. Moreover, multiple alleles in a population can be actively maintained when having a rare allele is advantageous or when individuals with heterozygous alleles show higher fitness. Our aim is to describe shared variation within Arabidopsis genus: what is left due to incomplete lineage sorting, acquired due to gene flow or maintained by balancing selection.

Genomic data (Illumina 100 bp paired reads) of all Arabidopsis lineages - 94 individuals from 25 different lineages of A. thaliana relatives - allow us to make a comprehensive survey of trans-specific polymorphisms, diversity and divergence in the genus. We found 1060 non-synonymous polymorphisms segregating in all 4 major Arabidopsis clades: A. thaliana, A. halleri, A. lyrata and A. arenosa. We showed that it's more sites than expected under the neutral model; there is no sign of gene flow between A. thaliana and its relatives; genes with such trans-specific SNPs follow allelic tree, rather than species tree. Defense response (R genes) and its regulators (DNA demethylation genes) are enriched among the genes with trans-specific SNPs. We conclude that strong balancing selection on Arabidopsis immune system persists throughout millions of years and leads to trans-specific shared polymorphisms.

134B

Whole-genome variations trace evolutionary footprints underlying chicken domestication and rapid adaptation based on next-generation sequencing

Yali Hou [1], Furong Qi[1], Xue Bai[1], Xiquan Zhang[2], Xuemei Lu[1]
[1] Beijing Institute of Genomics, Beijing, China, [2] South China Agricultural University, Guangzhou, China

The evolutionary dynamics underlying domestication and rapid adaptation for animal are of significance. Chicken, an important model organism, has been subjected to intensive human-driven selection, leading to remarkable phenotypic diversity. Resolving the genomic footprint provides insight into the mechanisms by which rapid adaptation shapes phenotypes. We sequenced a panel of 68 chickens from six commercial/local breeds with divergent characteristics and an ancient breed with distinct geographical distributions, profiled the whole-genome variations in terms of 21884097 SNPs, 1937065 INDELs, 10445 SDs and 30732 deletions, which exhibit substantial power to distinguish genealogies with distinct originations, selection purposes and geographical distributions. Demographic history reveals the chicken domestication time at genome-wide molecular level. Selective sweep analysis based on self-developed method revealed modern chickens have undertaken extensive changes in multiple systems involving nervous and brain development, endocrine and sensory systems, which were common events in animal domestication. The variants involving domestication and rapid adaptation are overrepresented in intronic regions including transcription factor binding sites, highlighting the importance of regulatory mechanism. Several intriguing genes/regions potentially associated with meat, lipid metabolisms and reproductive functions have been identified for meat-purpose and egg-purpose breeds, respectively. We creatively introduced the conception of entropy and exploited hundreds of breed-divergent islands and breed-specific islands, profiling the dynamics under adaptation. Additionally, INDELs and copy number variations also contribute to the genomic reconstructions during rapid adaption. This study traced the genomic dynamics and functional changings during breed divergence underlying intensive selection, provided promising insight into the putative genetic architectures of complex traits.

135C

Genomics of male-killing suppression in the Blue Moon Butterfly *Hypolimnas bolina*

Louise Reynolds [1], Emily Hornett[2], Greg Hurst[1]
[1] *University of Liverpool, Liverpool, UK,* [2] *University of Cambridge, Cambridge, UK*

The suppression of male-killing in the Blue Moon Butterfly *Hypolimnas bolina* is a compelling case study of rapid evolution. *H. bolin*a is infected with a male-killing strain of the endosymbiont *Wolbachia*. Affected populations have extremely female-biased sex ratios. *H. bolina* has however, evolved the ability to suppress the male-killing effects of *Wolbachia*. Selection for this 'suppressor locus' is extremely strong. The introduction of the suppressor into Samoa saw the population sex ratio change from 100:1 to 1:1, within the space of 5 years.

We have identified the doublesex (*dsx*) gene as a strong candidate for the suppressor locus in *H. bolina*. Doublesex controls whether cells develop male or female characteristics. It resides within the region to which we have mapped the suppressor locus, co-segregates with suppression and was the most strongly selected locus during a selective sweep for the suppressor gene. Analysis of *dsx* in pre and post suppressor populations has revealed an unusual signature of selection, indicative of gene duplication, in suppressed populations.

This leads us to believe that *dsx* is the target of selection in this system and also shows that reproductive parasites may act to drive the evolution of insect sex-determination systems.

136D

The Genetic Architecture of Lightened Skin Pigmentation in the Southern African ‡Khomani San

Alicia Martin [1], Xiaomin Liu[2], Julie Granka[1], Christopher Gignoux[1], Meng Lin[3], Alexandra Adams[1], Catherine Guenther[1], Caitlen Uren[4], Marlo Möller[4], Cedric Werely[4], David Kingsley[1], Marcus Feldman[1], Eileen Hoal[4], Carlos Bustamante[1], Brenna Henn[3]
[1] Stanford University, Stanford, CA, USA, [2] BGI-Shenzhen, Shenzhen, China, [3] SUNY Stony Brook, Stony Brook, NY, USA, [4] Stellenbosch University, Tygerberg, South Africa

Skin pigmentation is one of the most recognizably diverse and strongly selected phenotypes in humans, but its highly genetic basis has been primarily studied in European, Asian, and African American populations. By analyzing baseline pigmentation from >4800 individuals in 41 populations, we find that latitude explains ~49% of the variation in skin pigmentation, and that variability decays with distance from the equator. Light skin pigmentation is also observed in the far southern latitudes of Africa among KhoeSan hunter-gatherers of the Kalahari Desert. The KhoeSan hunter-gatherers are believed to have diverged from other populations ~100,000 years ago, and maintain extraordinary levels of genetic diversity. It is unknown whether their light skin represents convergent evolution or the ancestral human phenotype. We have collected ethnographic information, pigmentation phenotypes, and genotypes from 107 individuals and high coverage exome sequencing from 91 ‡Khomani San individuals from the Kalahari. Skin pigmentation is highly heritable ($h^2 > 0.9$), but previous GWAS genes do not explain significant variation in ‡Khomani individuals. Rather, a collection of candidate pigmentation genes identified through human divergence and model organism studies explain nearly half of the heritability. We identify new skin pigmentation associations near MITF, SMARCA2, TYRP1, and other genes in the KhoeSan that are supported by conservation, selection, previous model organism studies, and/or functional follow up. Our results indicate that genetic predictors of skin pigmentation are likely more globally heterogeneous than previously suggested and highlight the strength of diverse population studies to explain phenotypic variation in the context of human evolutionary history.

137A

Natural selection acting on context-dependent SNPs in coronary heart disease

Srilakshmi Raj [1], Gregory Dyson[2], Charles Sing[3], Andrew Clark[1]
[1] Cornell University, Ithaca, NY, USA, [2] Wayne State University, Detroit, MI, USA, [3] University of Michigan, Ann Arbor, Ann Arbor, MI, USA

We used a modified Patient Rule-Induction Method (PRIM) to test for genetic and environmental context dependence of SNP marker effects in predicting coronary heart disease (CHD) risk, in a genome-wide scan for context-dependent effects in the genome. PRIM performs a sequential binning of risk sets, similar to a regression tree, to detect non-linear and heterogeneous epistatic and GxE interactions that are not captured by traditional methods. PRIM was used on 840,000 common autosomal SNPs genotyped in 8,342 European-American adults from the Atherosclerosis Risk in Communities study. With CHD risk as the outcome and genome-wide SNPs as the genetic predictor variables, age, body mass index (BMI), type 2 diabetes diagnosis, smoking and hypertension were included as measures of context. PRIM analyses were carried out on a SNP-by-SNP basis, separately in males and females, and significance was assessed by extensive permutation testing.

The PRIM approach successfully identified combinations of SNPs and measures of context that significantly increased CHD risk in particular subsets of the sample. Gene ontology analysis of these context-dependent SNP effects, approximately 800 in females and 700 in males, showed evidence of an excess inclusion of metabolism-related terms. There was also evidence for positive selection, as indicated by the CLR and iHS tests. Finally, we found enrichment for eQTLs among the context-dependent signals. We were able to find evidence that SNPs which show a strong influence on complex disease risk in the context of environmental or other predictive risk factors appear more likely to be influenced by natural selection.

138B

Fly genomes reveal evolutionary origins of herbivory

Andrew Gloss, Richard Lapoint, Noah Whiteman
University of Arizona, Tucson, Arizona, USA

Nearly half of all insect species are herbivorous, and herbivorous lineages diversify more rapidly than their nonherbivorous relatives. However, evolutionary transitions to herbivory are rare, suggesting there are barriers to feeding on living, chemically defended plant tissue. To identify loci under positive selection during the evolution of herbivory, we studied herbivorous flies in the genus *Scaptomyza*, which is nested within the *Drosophila* radiation. This lineage transitioned from saprophagy to feeding on *Arabidopsis thaliana* and other mustards, which produce insecticidal compounds known as isothiocyanates (mustard oils), within the past 5-15 million years. We sequenced the genome of *Scaptomyza flava* and generated genome-wide polymorphism data for three herbivorous *Scaptomyza* species. Analyses of polymorphism and divergence revealed that genes known to be expressed in the midgut, a major site of contact with plant toxins during feeding, are evolving more rapidly in herbivorous *Scaptomyza* than in yeast-feeding *Drosophila.* Further, putative targets of positive selection are enriched for genes involved in physiological processes affected by isothiocyanates in humans. Interestingly, these include orthologs of genes underlying human diseases with symptoms similar to those resulting from isothiocyanate toxicity. However, patterns of gene duplication and loss in most large gene families, genome-wide rates of protein sequence evolution, and the proportion of amino acid substitutions fixed by positive selection are similar to yeast-feeding *Drosophila*. Overall, the transition to herbivory in *Scaptomyza* is marked by targeted, rapid evolution of genes involved in key functions such as detoxification, rather than dramatic genome-wide shifts in patterns of protein evolution.

139C

Joint analysis of human polymorphism and mammalian divergence reveals widespread selection on synonymous sites in humans and suggests a major role for epistasis

Arbel Harpak, Anand Bhaskar, Jonathan Pritchard
Stanford University, California, USA

The interplay of sequence evolution and purifying selection has long been a difficult issue to tackle. New large-sample human exome sequencing efforts now give us an unprecedented opportunity to examine in fine detail selection acting in recent millennia. In this study, we tackle this problem by utilizing polymorphism data from over 60,000 individuals from the Exome Aggregation Consortium (ExAC) along with divergence data from several mammalian species. We classify polymorphic sites in humans according to the various observed patterns of mammalian divergence at each site. We then compare the selective constraint on these classes by considering the excess of rare variants in the site frequency spectrum of humans. We find that synonymous sites are under significantly more constraint than nonsynonymous sites for which a parallel mutation occurred in a closely related mammal. Furthermore, this difference in constraint is significantly more pronounced the more closely related to humans the mammal under consideration is, a finding that is consistent with epistasis playing a key role in determining the deleteriousness of a variant. We find that the lion share of this effect can be attributed to the sequence context of a variant, and investigate different mechanistic levels at which this epistasis may act. Finally, our results suggest that synonymous mutations have pervasive functional consequences, putting in question their utility as neutral proxies in demographic inference and selection scans. We therefore consider how extensively used methods such as the McDonald-Kreitman test could be adjusted to account for the observed constraint on synonymous sites.

140D

Screening whole genomes for traces of recent selective sweeps using coalescent tree topology

Martina Rauscher [1], Saurabh Bhandari[2], Johannes Wirtz[1], Thomas Wiehe[1]
[1] University of Cologne, Cologne, Germany, [2] Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

In evolutionary biology, a major topic is the detection of genetic signatures left by the action of natural selection. For instance, recent and strong positive selection leads to the reduction or elimination of variation around the selected site, known as selective sweep.
Several statistical tests have been established to distinguish characteristic features of neutral from non-neutral evolution in the past. Some recently introduced test statistics exploit the fact that sweeps produce highly unbalanced coalescent tree topologies. Here, we apply such a test which focuses on the topology of coalescent trees to detect swept regions. We performed whole genome screens in human and Drosophila melanogaster using microsatellite and SNP data .
Although we find little agreement between the microsatellite and the SNP based test results, many of the significant regions coincide with previously annotated sweeps. In addition, we detect several new candidate sites. To corrobarate significance we subjected the regions found in human to a re-sampling strategy and confirmed several of the population-specific sweep candidates. We present and discuss these newly identified regions.

141A

Parallel evolution of weedy traits in weedy rice populations

Zhongyun Huang [1], Lin-Feng Li[2], Kenneth Olsen[2], Ana Caicedo[1]
[1] University of Massachusetts Amherst, Amherst, MA, USA, [2] Washington University in St. Louis, St. Louis, MO, USA

Parallel evolution is the evolution of similar phenotypes from similar starting phenotype. Different populations evolving independently may gain phenotypic similarity via the same or different genetic mechanisms. Our study system, weedy rice, is a great model for parallel evolution. Previous work from our lab and others has revealed that weedy rice has had multiple origins, including ancestry from varieties of rice cultivars, local wild rice and admixture events. Although weedy rice from different geographic regions has evolved independently, weedy populations reportedly possess similar sets of traits, defined as "weedy traits", which permit persistence in the recently created (<10,000) agricultural settings. Weedy rice populations have been shown to vary in the degree to which they display some typical weedy traits such as plant height, hull color, awn presence and flowering time. On the other hand, red pericarp and seed shattering tend to be more consistent characteristics. To assess the extent of parallel genetic and phenotypic evolution in weedy rice, we characterized a subset of weedy traits in populations of diverse origins from South Asia, and compared these to US weedy rice populations. We performed target sequencing on fifteen candidate genes that possibly underlie weedy traits, and that have had causal mutations involved in domestication phenotypes defined. We will compare sequence data of coding and flanking regions within a panel including cultivated, wild and weedy rice to reveal whether the same genes and same mutations have contributed to similar phenotypes in various rice populations.

142B

The role of epistasis in the dynamics of polygenic adaptation

Katya Kosheleva, Michael Desai
Harvard University, Cambridge, MA, USA

Populations adapt by randomly acquiring and fixing beneficial mutations. When populations are genetically diverse, many selected sites segregate simultaneously and compete with each other for fixation. In such populations, interactions between mutations- either due to physical linkage or epistatic interactions among sites- complicate evolutionary dynamics. To investigate the relative importance of these multi-locus interactions, we evolved populations of haploid budding yeast from variable pools of standing genetic diversity: 34 replicate populations derived from a diploid cell of a single strain background; 12 replicate populations derived from 80 spores picked from intercross lines of two diverged lab strains; and 5 replicate populations of 3 different sets of 6 intercross spores. Each set of founder pools was evolved for 960 generations subject to 3 different recombination rates, imposed by different frequencies of forced mating and sporulation. Pooled populations were then sequenced every 240 generations. Frequencies of standing and novel variants were quantified over each of these timepoints, and the trajectories of putatively selected sites were compared with an additive model to quantify the prevalence of epistatic interactions. The tendency of both linkage and epistasis to help maintain or purge standing genetic variation was quantified, and the interaction of standing epistatic networks with novel variants entering the population examined. By investigating how evolutionary dynamics, and consequently, evolutionary outcomes depend on the amount and distribution of standing variation, our work explores the role of historical contingency and determinism in guiding adaptation.

143C

Inference of the recent demographic history of polar bears using a simulation-based approach

Shakuntala Gopal [1], Shiping Liu[2], Eline Lorenzen[3,4], Francois Balloux[1], Pascale Gerbault[1], Matteo Fumagalli[1]

[1] University College London, London, UK, [2] Beijing Genomics Institute, Shenzhen, China, [3] Natural History Museum, Copenhagen, Denmark, [4] University of California, Berkeley, USA

Polar bears (*Ursus maritimus*) are uniquely adapted to the extreme conditions of life in the high Arctic, and have become a model organism in evolutionary biology. Recent studies have focused on estimating the divergence time and migration rate between polar bears and brown bears (*Ursus arctos*); however, genetic evidence shows population structuring within polar bears. Whole-genome sequencing data from Greenlandic populations show clear differentiation between East and West Greenlandic polar bears. Here, we aim to infer the recent demographic history of polar bears in Greenland by exploiting previously generated large-scale genomic data from 18 samples. Choosing the joint site frequency spectrum as the observed summary statistic, we used a coalescent approach to simulate comparable summary statistics. We applied an Approximate Bayesian Computation technique to estimate several demographic parameters of interest in polar bears, including divergence time and migration rates between East and West Greenlandic populations. Our results will shed light on past demography and can be used to pinpoint the geoclimatic events that have shaped the present-day diversity and structure of the species. Moreover, our estimated evolutionary model between the East and West polar bears will be used as a null model to estimate selective forces that can contribute to polar bear adaptation to the Arctic environment.

144D

tba

145A

Selection signature detection in Chinese Holstein and Simmental cattle

Minhui Chen [1,2], Dunfei Pan[1], Junya Li[3], Guosheng Su[2], Aiguo Wang[1], Li Jiang[1], Qin Zhang[1], Jian-Feng Liu[1]

[1] China Agricultural University, Beijing, China, [2] Aarhus University, Aarhus, Denmark, [3] Chinese Academy of Agricultural Science, Beijing, China

Identification of signals left by recent positive selection opens a feasible way in targeting genomic variants underlying complex traits and fitness. In present study, we simultaneously detected a genome-wide detection of recent positive selection occurring within and between Chinese Holstein and Simmental populations which have been under artificial selection with distinct purposes. We conducted analyses via various complementary analytical approaches, including EHH, XP-EHH and F_{ST} based on the Illumina 770K high density SNP array for pursuing more comprehensive detections. We successfully constructed profiles of genome-wide selective signals in the two cattle populations. Further annotating these regions exploited a set of novel promising functional genes related to growth, reproduction, immune response and milk production, as well as genes previously reported to be under positive selection. Furthermore, we also found that around 20 windows of identified regions overlapping in the two cattle breeds, demonstrating the convergent evolution between the two breeds. Finally, we investigated the distribution of SNPs presenting low values of F_{ST} across five distinct functional regions across the genome; we found a higher proportion of low-F_{ST} SNPs involved in high MAF bins of the cattle genome in comparison with those of human genome. This supports the more common balancing selection in cattle. Findings herein provided insight into the mechanisms of positive selection and also facilitated follow-up functional studies on potential candidate genes related to various economically important traits in cattle.

146B

The *Atpalpha* gene: local adaptation on *D. subobscura*

Cinta Pegueroles [1,2], Gemma Calabria[2], Laura Pineda[2], Carla Esteve[2], Maria Martí-Solano[3], Charles F. Aquadro[4], Francesc Mestres[1], Marta Pascual[1]

[1] Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra (UPF), Barcelona, Spain, [2] Dept. Genètica, Facultat de Biologia and IRBio. Universitat de Barcelona, Barcelona, Spain, [3] Dept. Experimental and Health Sciences, UPF, Hospital del Mar Medical Research Institute, Barcelona, Spain, [4] Department of Molecular Biology and Genetics, Cornell University, Ithaca, USA

Drosophila subobscura presents a rich polymorphism in chromosomal inversions, and most of them have unambiguously proved to be adaptive. Adaptation seems related with reduction of recombination due to physical constraints in inverted regions, which favors the linkage between a given allele and an inversion. We have studied the Atpalpha, a candidate gene for thermal adaptation in this species, which is closely located to the breakpoints of inversions O_7 and O_1, although outside of them. In particular, we have evaluated the genetic variability of this gene in several European populations (Málaga, València, Barcelona and Athens) in three adaptive arrangements (O_{3+4}, O_{3+4+1} and O_{3+4+7}). We found significant genetic differentiation between arrangements but not between populations. We observed almost fixed non-synonymous changes between arrangements, with most of the changes closely located at the nucleotide and protein three-dimensional structure. Most of the substitutions detected in both O_{3+4+1} and O_{3+4+7} arrangements were described in other species to confer resistance to ouabain, a class of plant toxin capable of blocking ATPases. We also evaluated the genetic contents of four additional arrangements: O_{3+4+10} has the same allele than O_{3+4+1}; O_{3+4+2}, O_8 and O_{ST} have the same allele than O_{3+4}. Similarities in Atpalpha at the protein level between O_{3+4+1} and O_{3+4+7} arrangements may be the result of parallel or collateral genetic evolution in response to selective pressures in similar environmental conditions. In contrast, O_{3+4+1} and O_{3+4+10} arrangements share one of the inversion breakpoints, suggesting that their shared genetic content may be due to a common ancestor.

147C

Development of SNP markers for origin discrimination between Korean and Chinese *Lycii Fructus* (Goji berry)

Nam-Kuk Kim, Changhee Lee, Soomin Park, Jin-Ho Lee, Hye-Ri Bae, Seong-Hun Lee
Experiment Research Institute, National Agricultural Product Quality Management Service (NASQ), Gimcheon, Gyeongsangbuk-do, Republic of Korea

At present, both Korean and Chinese agricultural products are commonly found on the Korean market. Therefore, the correct origin labeling is very important for support consumers' right to know. For this purpose, we developed single nucleotide polymorphism (SNP) markers to discriminate between Korean and Chinese Goji berry. To develop the discrimination markers, we analyzed for origin specific alleles based on the next generation sequencing (NGS). Among the 623 SNPs, 6 SNPs were selected as candidate markers. To test the ability of SNP markers for discrimination between Korean and Chinese Goji berry, the genotype was investigated by allele-specific PCR (AS-PCR) methods. The probability of discrimination was found to be 100% in a subsequent validation set consisting of 500 samples. These SNP markers could be useful for discrimination between Korean and Chinese Goji berry and would contribute to the prevention of falsified origin labeling.

148D

Population genomics of European woodland strawberry

Tuomas Toivainen [1], Daniel J Sargent[2], Jon Hallsson[3], Hrannar Hilmarsson[3], Timo Hytönen[1,4]
[1] Department of Agricultural Sciences, University of Helsinki, FIN-00014, Helsinki, Finland, [2] Istituto Agrario San Michele all'Adige, ITA-38010, San Michele all'Adige, Italy, [3] Agricultural University of Iceland, Keldnaholt, 112, Reykjavik, Iceland, [4] Department of Biosciences, University of Helsinki, FIN-00014, Helsinki, Finland

The geographical distribution of the woodland strawberry (*Fragaria vesca ssp. vesca*) covers almost whole Europe. Environmental conditions vary greatly between the margins of its geographical range; it can grow in various climates from the Mediterranean region with hot and dry summer (southern Spain, latitude 37°N) to the cold climate of the most northern parts of Norway (latitude 70°N). In addition to a wide ecological breadth, a perennial lifestyle, efficient clonal reproduction by stolons, sexual reproduction by mixed mating and seed dispersal by animals are interesting features concerning adaptation potential in this species. However, a little is known about phenotypic differences between populations or the amount of genetic variation in general. The aim of the study was to examine population genomics in northern European woodland strawberry populations. Specifically, what kind of population structure do we find, how much variation populations harbor (θ_W), and how differentiated populations are in general (F_{st}). To answer these questions, we genotyped by sequencing 95 plants from Norway, Iceland and Finland covering the latitudes 60-70. Results will be discussed.

149A

Evolution of robustness to protein mistranslation by accelerated protein turnover

Dorottya Kalapis [1], Ana Rita Bezerra[2], Zoltán Farkas[1], Péter Horváth[1], Zoltán Bódi[1], Andreea Daraba[1], Béla Szamecz[1], Ivo Gut[3], Mónica Bayes[3], Manuel Santos[2], Csaba Pál[1]
[1] Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary, [2] Institute for Biomedicine – iBiMED, Health Sciences, University of Aveiro, Aveiro, Portugal, [3] Centro Nacional de Análises Genómico, Parc Científic de Barcelona, Barcelona, Spain

Translational errors occur at high rates, influence organism viability and the onset of genetic diseases. To investigate how organisms mitigate the deleterious effects of protein synthesis errors during evolution, a mutant yeast strain was engineered to translate a codon ambiguously (mistranslation). It thereby overloads the protein quality control pathways and disrupts cellular protein homeostasis. Laboratory evolutionary experiments revealed that fitness loss due to mistranslation can rapidly be mitigated. Genomic analysis demonstrated that adaptation was primarily mediated by large-scale chromosomal duplication and deletion events. By altering the dosages of numerous, functionally related proteins simultaneously, these genetic changes enabled rapid adaptation to mistranslation. Evolution increased the level of tolerance to mistranslation through acceleration of ubiquitin-proteasome mediated protein degradation and protein synthesis. As a consequence of rapid elimination of erroneous protein products, evolution reduced the extent of toxic protein aggregation in mistranslating cells. However, there was a strong evolutionary trade-off between adaptation to mistranslation and survival upon starvation: the evolved lines showed fitness defects and impaired capacity to degrade mature ribosomes upon nutrient limitation. Moreover, as a response to an enhanced energy demand of accelerated protein turnover, the evolved lines exhibited increased glucose uptake by selective duplication of hexose transporter genes. We conclude that adjustment of proteome homeostasis to mistranslation evolves rapidly, but this adaptation has several side-effects on cellular physiology. Our work also indicates that translational fidelity and the ubiquitin-proteasome system are functionally linked to each other and may therefore co-evolve in nature.

150B

Genomic footprint of local climate adaptation

Ann-Marie Oppold [1], Simit Patel[2], Markus Pfenninger[1,2]
[1] Goethe University Frankfurt am Main, Frankfurt am Main, Hessen, Germany, [2] Senckenberg, Biodiversity and Climate Research Centre, Frankfurt am Main, Hessen, Germany

Local adaptation is commonly observed in species with a wide distribution range. However, little is known about the genomic basis of local adaptation or the molecular mechanisms driving this process. We analysed populations of the non-biting midge (*Chironomus riparius*) along a temperature gradient across Europe by integrating common-garden experiments with genome-wide scans. Ambient temperature is a crucial environmental factor for ectothermic organisms, so differences among populations along the temperature gradient would be indicative of local adaptation.

Common garden experiments at three different temperatures (14, 20 and 26 °C) showed that populations sampled from cooler regions were less fit at the highest experimental temperature compared to populations sampled from warmer regions. These temperature-dependent fitness effects indicate local climate adaptation along the temperature gradient.

To identify candidate genes involved in local adaptation along the temperature gradient, we applied the Pool-Seq method to scan the population genomes and found more than 2.5 million single nucleotide polymorphisms (SNPs) among five natural populations. Based on F_{ST}-analyses, more than 2000 highly differentiated and even fixed (F_{ST}=1) SNPs were identified and annotated. We further associated the identified SNPs to environmental variables (temperature, latitude, longitude) and found significant correlations. To also account for patterns of polygenic adaptation and not merely focus on outlier loci, we further aimed to perform gene-set enrichment analyses.

Our results give insight in the complexity of the genome-wide distribution of adaptive genetic variation in natural populations.

151C

Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain

Adrien Rieux [1], Vegard Eldholm[2], Johana Monteserin[3], Beatriz Lopez[3], Benjamin Sobkowiak[1], Viviana Ritacco[3], Francois Balloux[1]

[1] University College London, London, UK, [2] Division of Infectious Disease Control, Norwegian Institute of Public Health, Oslo, Norway, [3] Instituto Nacional de Enfermedades Infecciosas-ANLIS Carlos Malbrán, Buenos Aires, Norway

The rise of drug-resistant strains is a major challenge to containing the tuberculosis (TB) pandemic. Yet, little is known about the extent of resistance in early years of chemotherapy and when transmission of resistant strains on a larger scale became a major public health issue. Here we use tip-dating analyse based on heterochronus sampling of the bacteria to reconstruct the time-line of antimicrobial resistance evolution during a major ongoing outbreak of multidrug-resistant TB in Argentina. We estimate that the progenitor of the outbreak strain acquired resistance to most common drugs by around 1973, indicating continuous circulation of a multidrug-resistant TB strain for four about 40 years, about one decade before the earliest documented transmission of Mtb extensive drug resistant (XDR) strains with such profiles globally. Our results suggest that mutation rate is a poor predictor of the emergence of drug-resistant TB and highlight the importance of other potential determinants.

152D

Host adaption in the plant pathogenic fungus *Mycosphaerella fijiensis*

Jean Carlier [1], Marie-Françoise Zapater[1], Daniel Bieysse[1], Yanetsy Montero[2], Véronique Roussel[1], Rémy Habas[1], Luis Perez-Vicente[2], Catherine Abadie[1], Stephen Wright[3]
[1] CIRAD, UMR BGPI, Montpellier, France, [2] INISAV, Havana, Cuba, [3] University of Toronto, Toronto, Ontario, Canada

Plant pathogenic fungi are able to erode quantitative host resistance through changes in aggressiveness, thereby threatening the durability of host resistance. Such erosions are suspected in some areas in the fungus *Mycosphaerella fijiensis*, responsible for a recent and devastating banana pandemic, Black Leaf Streak Disease (BLSD). This study aims to test for the action of host-specific adaptation and to detect host-selected genes in *M. fijiensis*. We collected six samples in Cuba in three locations distributed throughout the banana production zones where resistant cultivars have been used for about 15 years. For each location, about 40 isolates were collected from two banana plots containing either a resistant variety or a susceptible variety located two to 10 km apart. We also included in the study three samples from Honduras where the disease was first introduced in the Latin America- Caribbean area. A significant host effect was detected in some locations for some aggressiveness traits evaluated under controlled conditions. A genome scan approach was conducted from whole-genome sequencing of pools of individuals (pool-seq). Differentiated genomic regions were detected between pathogen populations from the two cultivars in some locations. Further analyses have been undertaken to characterize these regions.

153A

The genetics of environmental adaptation in house mice across the Americas

Kathleen Ferris, Taichi Suzuki, Andreas Chavez, Megan Phifer-Rixey, Michael Nachman
UC Berkeley, Berkeley, CA, USA

Understanding the genetic basis of adaptation is a central goal in evolutionary biology. The house mouse, *Mus musculus domesticus,* is an excellent system for identifying the loci underlying adaptation to different environments. While native to Western Europe, it has recently expanded its range to a variety of new habitats throughout the Americas. Previous work has shown adaptive phenotypic differentiation of mice along the East Coast of North America; mice from higher latitudes are larger and build bigger nests compared to mice from lower latitudes. We have collected tissue and phenotypic data from latitudinal transects across North and South America including one from Arizona to Alberta. The environment changes dramatically along this transect, particularly in temperature, moisture, and daily temperature fluctuation. Phenotypic clines in body size and coat color are seen along this transect. Whole exome sequence data from 50 mice representing five populations are being used to conduct genomic scans for selection using analytical methods that account for population structure. Live mice were collected from several populations and used to create new wild-derived inbred lines. Reared in a common laboratory environment, these mice reveal genetically-based phenotypic differences among populations, including phenotypes that are likely to be adaptive in different environments.

154B

Evolution and diversity of one of the sensory receptors, free fatty acid receptor GPR120

Risa Iwasaki, Jun Gojobori, Yoko Satta
The Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa, Japan

Sensory receptor genes are believed to have a large diversity in their sequences and the number of genes among different species. This may be because these receptors require various forms of information, such as chemicals, light and temperature, depending on the species' habitats. For example, *TAS1R* and *TAS2R* families which are sweet/umami and bitter taste receptors, respectively, have gene number and sequence variations in vertebrates. The *GPR120* (FFAR4) receptor is a member of the GPCR protein family that recognizes various fatty acids. Its evolutionary characteristics, however, have not been fully examined so far. To investigate the molecular evolution of this receptor, orthologs and paralogs to human *GPR120* have been extrapolated from vertebrate genomes. Compared to the *Tas1r* and *Tas2r* families, *GPR120* does not show gene number variation and retains a conserved mode of molecular evolution. The d_N/d_S ratio in 58 mammalian *GPR120* is 0.18 and significantly more conservative than that in 50 *Tas1r* sequences (0.27, $P < 0.001$). However, the proportion of segregating amino acid sites from the 1000 genome database in *GPR120* is not significantly different from that of *Tas1r*: 0.040 (= 39/975) for *Tas1r* and 0.053 (= 20/374) for *GPR120* ($P = 0.28$). To examine which gene is an outlier, inter- and intra- specific variation of other receptor genes including opsin and thermo-sensory receptor genes, and house-keeping genes including hemoglobin genes, are now examined in the same way. Study of these receptor genes will aid in elucidating the cause of the discrepancy in variation between and within species.

155C

Reversal evolution of plasmid copy number in *Escherichia coli*

Judith Ilhan [1], Itzhak Mizrahi[2], Tal Dagan[1]

[1] Christian-Albrechts-University, Kiel, Germany, [2] Agricultural Research Organization, Volcani Center, Bet-Dagan, Israel

Plasmids are mobile genetic elements that play an important role in lateral gene transfer during microbial evolution. Yet, the evolutionary dynamics shaping the genome of these extra-chromosomal elements remain elusive. Here we study plasmid evolutionary dynamics experimentally by evolving a broad-host range plasmid in *Escherichia coli* MG1655. In addition to the original plasmid, having a low copy number (~10), we modified the plasmid origin of replication to create a new high (~700) copy number plasmid having an identical backbone. The plasmids where transformed into identical hosts and evolved in parallel over 800 generations at 37°C and 42°C using a chemostat system. A comparison of growth rate over time between hosts carrying the two plasmids revealed a fitness decrease in the high-copy plasmid host populations. Furthermore, after about 100 generations we observed a stark copy number decrease of the high-copy plasmid in both culturing temperatures. Over time, the copy number of the modified plasmid stabilized into the original low copy number. Our results demonstrate host-plasmid co-evolutionary dynamics in real time and suggest that plasmid copy number adaptation occurs rapidly. Current efforts are aimed at pinpointing the molecular mechanisms of plasmid copy number evolution.

156D

Population structure of an androdioecious species, *Cardamine amara*

Hiromi Matsumae [1], Jörg Bachmann[1], Andrew Tedder[1], Alejandro Morales Suárez[1], John Pannell[2], Kentaro K. Shimizu[1]

[1] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [2] University of Lausanne, Lausanne, Switzerland

The coexistence of hermaphrodites and female-sterile individuals (males), or androdioecy, is exceptionally rare in Angiosperms. This rarity is linked to the theoretically validated requirement for males to have a minimum two-fold higher siring success (male advantage) than the male component of hermaphroditic individuals in order to be maintained within a population. If hermaphrodites within the population are capable of self-fertilization, then male advantage must dramatically increase. Swiss populations of *Cardamine amara* (Brassicaceae) have recently been documented as androdioecious, in which males are supported via significantly increased clonal propagation, increasing per genotype male advantage to a ratio of approximately 3.7:1. This increase in clonal reproduction is likely to benefit males in two ways; 1) increase pollen number and associated siring success, and 2) genet survival/ maintenance over time. The interplay then, between sexual reproduction and clonal (asexual) reproduction allows one to look in more depth at population structure and architecture at various spatial scales, and disentangle the roles played by both sexual and asexual reproduction, and indeed delimit the roles both of these play in shaping population life-history traits. Here we present data from both a suite of microsatellite markers, and whole genome resequencing data from populations across Switzerland and Germany, at large and small spatial scales. We discuss a population structure and genetic diversity of two sexes in *C. amara* understanding a trajectory how the androdioecy has evolved in natura.

157A

Lake Baikal amphipods: transcriptomes, detailed phylogeny and test for positive selection.

Sergey Naumenko [1,2], Georgii Bazykin[1,2], Anna Etingova[4], Maria Logacheva[1,2], Anna Klepikova[2], Alexey Penin[1,2], Lev Yampolsky[3], Alexey Kondrashov[2,5]
[1] Institute for Information Transmission Problems, Moscow, Russia, [2] Moscow State University, Moscow, Russia, [3] East Tennessee University, Johnson city, USA, [4] Baikal Museum., Irkutsk, Russia, [5] The University of Michigan, Ann Arbor, USA

The endemic species flock of Lake Baikal amphipods is a striking example of endemic adaptive radiation. This species flocks contains over 250 species and demonstrates a high level of morphological and ecological diversity, including miniature and giant species, both conservative amphipod morphology and highly armored forms, a unique plankton species as well as highly specialized active predators, abyssal scavengers and parasites.
We have sequenced whole body RNAs from 64 species, and identified 5000 - 30000 protein coding genes per species, clustered to 1000 orthologous groups.
We present results on the detailed phylogeny of the flock, heterozygosity levels, heterogeneity of speciation rate over time, GO-annotation, rates of synonymous and nonsynonymous substitutions, MK-test for positive selection and test of parallel molecular evolution.
Additionaly we present a method for orthology inference for datasets with numerous species using transcriptome data based on clustering of core species proteins and mapping of reads from remaining species.

158B

On the quest for the causal SNP – Studying Adaptation of replicated *Drosophila simulans* populations under cold stress

Thomas Taus, Reka K. Kelemen, Raymond Tobler, Viola Nolte, Christian Schlötterer
Vetmed Uni, Vienna, Austria

Evolve and resequence (E&R) is a promising approach to identify selected genomic regions and to describe their evolutionary dynamics in populations adapting to a novel environment. However, recent E&R studies with sexually reproducing organisms faced the problem of a large number of false positives, making the identification of the genes under selection difficult. Computer simulations as well as empirical results in yeast suggested that the small number of replicates typically used for previous E&R studies could explain the large number of false positives. We analyzed 10 replicate populations of 1000 *Drosophila simulans* individuals, after 30 generations in a cold environment, and identified clear selection signatures: pronounced chimneys of SNPs deviating from neutral expectations were seen in Manhattan plots and this selection signature became more pronounced with an increasing number of generations. Interestingly, these selection signatures could be also recognized if only 5 replicates were analyzed, suggesting that the number of replicates cannot explain the differences between previous *D. melanogaster* studies and this *D. simulans* experiment. We will explore the experimental factors, which caused these striking differences and discuss how this could translate into improved experimental design for future E&R experiments.

159C

Detecting selection: Pool-Seq vs. haplotypes

Daniel Gómez-Sánchez [1,2], François Mallard[1], Viola Nolte[1], Christian Schlötterer[1]
[1] Institut für Populationsgenetik, Vienna, Austria, [2] Vienna Graduate School of Population Genetics, Vienna, Austria

The analysis of allele frequency spectra in natural populations is a powerful approach for the identification of targets of selection. In particular the comparison of local populations has been suggested to provide insights into the selective forces operating locally. Despite the plunging sequencing costs, large scale population analyses on the genomic scale are still expensive, therefore the analysis of pools of individuals (Pool-Seq) has been proposed as a cost effective alternative. Nevertheless, it is not yet clear to what exent the power to detect selection is being compromised by the absence of linkage information in the Pool-Seq data. To address this question in real populations, we compare the power to identify selection signatures in three *D. simulans* populations by the analysis of at least 30 haplotypes for each population to Pool-Seq data from the same populations.

7 Origins and evolution of molecular innovation

7.1

The link between pervasive transcription and de novo gene evolution

Jorge Ruiz-Orera[1], José Luis Villanueva-Cañas[1], M.Mar Albà [1,2]
[1] Fundacio IMIM (FIMIM), Barcelona, Spain, [2] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

The sequencing of complete genomes from closely related species has revealed that many genes originated from previously non-coding genomic sequences. *De novo* evolved genes, once thought to be very rare, have now been found in many different species. The birth of a new protein represents a radical molecular innovation that can have important functional consequences. One example is the *Arabidopsis*-specific 59 amino acid long protein QQS, which alters carbon and nitrogen allocation and leads to increased seed and leaf protein content.

How can a new functional gene evolve from scratch? The finding that a large portion of the genome is expressed implies that there is abundant raw material available for this process. In addition, ribosome profiling has shown that many of the transcripts produced, even if they only contain very short open reading frames, are likely to be translated. This is expected to generate a large number of peptides that can be further tested by natural selection. We will present the results of our quest for *de novo* genes in mammalian genomes using deep transcriptomics.

7.2

Transcription of primate-specific endogenous retrovirus ERVH creates new genes and defines human naïve-like stem cells.

Jichang Wang[1], Gangcai Xie[3], Manvendra Singh[1], Avazeh Ghanbarian[1], Tamas Rasko[1], Attila Szvetnik[1], Huiqiang Cai[1], Daniel Besser[1], Alessandro Proigione[1], Nina Fuchs[1,4], Gerald Schumann[4], Wei Chen[1], Matthew Lorincz[5], Zoltan Ivivs[4], Zsuzsanna Izsvak[1], Laurence Hurst [1]

[1] Max-Delbrück-Center for Molecular Medicine, Berlin, Germany, [2] University of Bath, Bath, UK, [3] AS-MPG Partner Institute for Computational Biology, Shanghai, China, [4] Paul-Ehrlich-Institute, Langen, Germany, [5] University of British Columbia, Vancouver, Canada

Naïve embryonic stem cells (ESCs) hold great promise for therapeutics as they have broad and robust developmental potential. While such cells are readily derived from mouse blastocysts it has been impossible to readily isolate human equivalents. Here we show that a sub-population within cultures of human ESCs (hESCs) and induced pluripotent stem cells (hiPSCs) manifest key properties of naïve state cells and highlight the role of spontaneously created genes in their definition. These naïve-like cells can be genetically tagged, and are associated with elevated transcription of HERVH, a primate-specific endogenous retrovirus (ERV). HERVH elements provide functional binding sites for a combination of naïve pluripotency transcription factors, including LBP9, recently recognized as relevant to naivety in mice. We find that LTR7 of HERVH initiates chimeric transcripts, functions as an alternative promoter or modulates RNA processing from a distance. 128 and 145 chimeric transcripts were identified in hiPSCs and hESCs, respectively, many with a HERV derived conserved domain implicated in protein binding. The chimeric transcripts between HERVH and a downstream gene generally lack the 5' exon(s) of the canonical version (e.g. SCGB3A2) while part of HERVH/LTR7 is exonized (e.g. RPL39L). A significant fraction of HERVH sequence can be incorporated into novel, lineage-specific genes (e.g. ESRG) or lncRNAs (e.g. RP11-69I8.2). Via knockdowns we show that disruption of LBP9, HERVH and spontaneous new HERVH-derived transcripts (e.g ESRG, linc00458) compromises self-renewal. These observations define HERVH expression and associated new genes as a hallmark of naïve-like hESCs, and establish novel primate-specific transcriptional circuitry regulating pluripotency.

7.3

New domains as actors of molecular innovation: physico-chemical properties, evolutionary analyses and origination

Tristan Bitard-Feildel [1], Steffen Klasberg[1], Magdalena Heberlein[1], Isabelle Callebaut[2], Erich Bornberg-Bauer[1]

[1] Westfälische Wilhelms University, Institute for Evolution and Biodiversity, Evolutionary Bioinformatics, Münster, Germany, [2] IMPMC, Sorbonne Universités - UMR CNRS 7590, UPMC Univ Paris 06, Museum National d'Histoire Naturelle, IRD UMR 206, Paris, France

The phenotypic diversity of species is a consequence of molecular changes at the genomic level. One prominent mechanism for creating innovations is the reuse of functional protein units called domains. Understanding the evolutionary mechanisms behind domains' emergence and functional adaptation is crucial for our comprehension of molecular innovation.

Current domain annotation methods detect domains by Hidden Markov Models which are built from sequences known to be belong into the same domain family. However, by definition, these methods must fail for the detection of recently emerged domains. We propose a pipeline for the detection of such new domains[1]. The method is based on the Hydrophobic Cluster Analysis[3] of proteins and relies on the conservation of the hydrophobic properties of protein sequences. The methodology is used to detect new domains on a set of Drosophila orthologous proteins[1]. A second group of domains is created from Pfam. The second group of domains are selected based on their specificity toward the Arthropod clade[2] and correspond to clade specific orphan domains.

The two groups of domains are compared against domain sequences from Pfam. Analyses of their physico-chemical properties, their organisation at the gene level, their evolutionary pressure and their tissue specific expression reveal striking different signals between domain groups. Based on the analysis of these properties, we discuss different scenari of origination and evolution.

[1]Bitard-Feildel et al, Biochimie, in press
[2]Moore et al, Biochim. Biophys. Acta, 2012
[3]Gaboriaud et al, FEBS Lett.}, 1987

7.4

The recent de novo origin of protein C-termini

Matt Andreatta, Josh Levine, Scott Foy, Lynette Guzman, Matt Cordes, Joanna Masel
University of Arizona, Tucson, AZ, USA

Protein-coding sequences can arise either from duplication and divergence of existing sequences, or de novo from non-coding DNA. Unfortunately, recently evolved de novo genes can be hard to distinguish from false positives, making their study difficult. Here we study a miniature version of the process of conversion of non-coding sequence into coding: the co-option of short segments of non-coding sequence into the C-termini of existing proteins via the loss of a stop codon. Because we study recent additions to potentially old genes, we are able to apply a variety of stringent quality filters to our annotations of what is a true protein coding gene, discarding the putative proteins of unknown function that are typical of recent fully de novo genes. We identify 56 examples of C-terminal extensions in Saccharomyces and 28 in Drosophila, all of them recent enough to still be polymorphic. We find one putative gene fusion that turns out, on close inspection, to be the product of replicated assembly errors, further highlighting the issue of false positives in the study of rare events. Four of the Saccharomyces C-terminal extensions (to ADH1, ARP8, TPM2 and PIS1) that survived our quality filters are predicted to lead to significant modification of a protein domain structure.

7.5

How novel proteins arise in evolution: the invention of protein structure and the flux of genes through genomes

Victor Luria [1], Amir Karger[2], John Wesley Cain[3,4], Bradley J.S.C. Olson[5], Rafik Neme[6], Marc W. Kirschner[1]

[1] Harvard Medical School, Department of Systems Biology, Boston, MA, USA, [2] Harvard Medical School, Research Computing, Boston, MA, USA, [3] Harvard University, Department of Mathematics, Cambridge, MA, USA, [4] University of Richmond, Department of Mathematics, Richmond, VA, USA, [5] Kansas Sate University, Department of Biology, Manhattan, KS, USA, [6] Max Planck Institute for Evolutionary Biology, Plön, Germany

The origin of new proteins poses fundamental questions for evolution. Since Darwin, selection has been demonstrated in experimental and natural settings, yet the quality and quantity of phenotypic variation on which selection acts is poorly understood. Recent analyses in several organisms suggest genes can originate from ancestrally non-genic sequences. How easy is it to evolve new proteins directly from genomic as compared to random DNA? What is the stability, spectrum of interactions and function of short proteins designated as de novo proteins?

First, we identified APCDD1, a novel human protein, and showed it inhibits Wnt (Shimomura*, Luria*, Nature 2010) and BMP signaling in neurons.

Second, to determine the frequency and content of novel genes, we investigated how they may arise as a function of measurable gene, genome and population parameters. Surprisingly, we find novel genes may arise frequently and that gene turnover is fast, suggesting many are made but few are kept. Using proteomic, RNA-Seq and genomic data, we identified thousands of novel genes in 3 related algal genomes at a transition to multicellularity, and 3 chordate genomes. Determining predicted structure content, we found novel proteins are short, disordered and promiscuous, suggesting protein evolutionary maturation involves lowering disorder and partner number. While starting experiments, we analyze 9 other eukaryotic genomes to detect if class-wide properties exist for all novel genes. Since de novo genes can acquire genetic partners and biochemical functions, the next challenge will be finding what fraction are maintained in genomes that continuously generate and destroy new genes.

7.6

Inter-kingdom horizontal gene transfer contributes to genomic innovations in diverse animal phyla

Sandie Degnan, Selene Valverde, Simone Higgie
University of Queensland, Brisbane, Australia

Inter-species horizontal gene transfer (HGT) is one mechanism by which organisms can extend their phenotype and thus their evolvability. The full extent of HGT into metazoan genomes is currently unknown, but the potential for HGT from other kingdoms to have a contribution in shaping animal genomes is obvious given the constant and intimate interactions between animals and microbes through more than 600 million years of animal evolution. We have applied a novel pipeline to conservative identification of HGT events in several invertebrate animals to reveal that HGT has made a consistent contribution to animal evolution, but in a largely lineage-specific manner. We focus especially on HGT events in early branching metazoans for which there is a publicly-available draft genome, and find that the level and diversity of HGT events appears to be largely taxon restricted. Each species studied have their own repertoire of HGT contributed gene families and encoded protein domains. Nonetheless there are number of interesting parallel enrichments that appear to have happened independently. Specifically, all taxa investigated have a disproportionate number of solenoid-repeat domain-containing proteins that are thought to mediate protein-protein interactions. Using the marine sponge *Amphimedon queenslandica* as a case study, we show that multiple apparent HGT loci can be explained by a small number of transfer events followed by duplication, that horizontally acquired genes appear to have important functions, and that at least some HGT events can be traced back to a last common poriferan ancestor.

7.7

Evolutionary pervasive transcription across the genome is a continuous source of material for frequent de novo gene emergence.

Rafik Neme, Diethard Tautz
Max Planck Institute for Evolutionary Biology, Plön, SH, Germany

Systematic genome comparisons have shown that new genes can emerge from previously non-genic regions of the genome. De novo transcripts are usually non-coding, but can acquire functional open reading frames. Despite the pre-conceived notion that de novo gene birth should be rare, several recent studies have shown this to be an active and frequent mechanism. In fact, surveying gene emergence across long phylogenetic divergence times (hundreds of millions of years) invariably shows very high emergence rates for the youngest lineages. This has led to the prediction that high rates of gene emergence are balanced through subsequent high loss rates. This holds true at the protein level, where constraints to generate and maintain stable open reading frames are very stringent, characterized by frequent losses.
We set to determine the dynamics at the transcriptional level, seeking to understand which limitations would occur prior to the evolutionary transition between transcription and translation of de novo genes. For this purpose we gathered genomic and transcriptomic data from closely related mouse species, covering various divergence times between a few thousands to approximately 10 millions years. Our results confirm the long-standing idea that the genome is pervasively transcribed. This includes also untranscribed regions in one genome, which are transcribed in closely related species, and therefore only a few mutational steps away. This type of pervasive transcription provides ample substrate for the emergence of new genes, and in contrast to the dynamics at the protein-coding level, gains of de novo transcripts are not quickly balanced by losses.

7.8

Elucidating the Sources of Species-Specific Genes

Bryan Moyers
University of Michigan, Ann Arbor, MI, USA

It is observed that there are genes in a species for which no homologs can be found in other species. These species-specific genes can drive phenotypic diversity. But how these genes arise is not clear. Several suggestions have been made since the modern synthesis, and this has been an intense area of difficult questioning. From Francois Jacob's remarks that "the probability that a functional protein would appear de novo by random association of amino acid is practically zero" to Susamu Ohno's influential book "Evolution by Gene Duplication" to the modern hypotheses about mechanisms of *de novo* gene birth. We analyze the evidence for one such mechanism presented by Carvunis *et al*. in light of recent findings that homology detection error can contribute significantly to trends associated with gene age. Some of these trends can be predicted *a priori* from the workings of homology detection programs, whereas others are not predictable. It is therefore possible that some of the trends used to support Carvunis *et al*'s mechanism are actually unrelated. We find that the evidence for *de novo* gene birth is weaker than previous studies have indicated. We suggest a model by which species-specific genes are partially accounted for by homology detection error. We provide estimates of how much homology detection error contributes to apparent species-specific genes, and suggest that only the remaining fraction of such genes need to be explained by *de novo* gene birth models.

7.9

GenTree: an integrated resource for gene age dating and annotation refinement in human

Yi Shao[1], Chunyan Chen[1], Hao Shen[2,1], Manyuan Long[3], Yong Zhang [1]
[1] Institute of Zoology, CAS, Beijing, China, [2] Department of Computer Science, Hunan University of Technology, Zhuzhou, China, [3] Department of Ecology and Evolution, The University of Chicago, Chicago, USA

Various mechanisms including duplication and de novo origination create new genes, which in turn drive phenotypic evolution. However, these species- or lineage-specific genes are narrowly transcribed and poorly conserved. Those created by partial duplication and de novo origination usually encode short proteins. Such features cause the lability of new gene annotation in the mainstream practice (e.g. Ensembl). Specifically, for 1,828 primate-specific coding genes annotated in Ensembl v51, 61% were reannotated as noncoding RNAs, pseudogenes or deleted in Ensembl v73. By contrast, the biotype was changed for only 8% of genes predating primate split. The annotation instability of de novo genes is particularly serious for which a mere eight percent (5/60) of previously identified candidates were kept as coding genes in Ensembl v73. In order to alleviate this problem, we are developing a web interface, GenTree, (http://gentree.ioz.ac.cn) to help infer gene age and refine gene models. On the one hand, we took advantage of the vertebrate whole-genome alignments provided by UCSC and inferred genes' origination time together with the origination mechanism. Published age datasets including phylostratigraphy and ProteinHistorian were cross-referenced. On the other hand, sequencing based transcriptome data were imported to refine the exon-intron structure of candidate primate-specific new genes, while mass spectrum based proteome information was integrated to help differentiate coding and noncoding genes. Actually, at least 116 Ensembl annotated pseudogenes are protein-coding as supported by mass spectrum. Currently, GenTree only consists of human data and the future update will cover model organisms such as mouse and fruitfly.

7.10

The place of De Novo Gene Origin in the Evolutionary Dynamics of the Genome-as-Population

Lee Altenberg
The KLI Institute, Klosterneuburg, Austria

Because genetic elements in a genome undergo duplication, loss, transformation, horizontal gene transfer, and de novo origination, at rates degrees depending on the elements, they combine all the dynamical features that comprise a Darwinian system – heritable variation in fitness – but at the level of genome-as-population. Fragments of the evolutionary dynamics at this level have been investigated, but in conceptual isolation from one another. These include the systems-biology determinants of gene duplicability, consequences of small duplication vs. whole gene duplication origins and retention of duplicate genes, concerted gene family evolution, transposable element dynamics, horizontal "gene transferability", and de novo gene origin. In de novo gene origin, non-expressed sequences become altered and gain expression. Their longevity in the genome, however, requires that the evolve function to the point where to lose them would be deleterious. The requirement for the evolution of function therefore filters out from a "House-of-Cards" distribution of de novo genes those sequences that 1) attach to the "interactome" in such a way as to improve or generate functions with selective opportunity, and 2) can improve function through subsequent mutational change. The characteristics of retained de novo genetic sequences therefore will already depart from the raw distribution of new de novo genetic sequences by exhibiting phenotypic modularity with respect to functions under stabilizing selection, and sequence evolvability. They should thus exhibit some of the same properties that are observed in sequences exhibiting high "gene duplicability".

7.11

Retrogenes illuminate dynamics of new gene structure and regulatory evolution in mammals

Francesco Nicola Carelli [1,2], Maria Warnefors[1,2], Henrik Kaessmann[1,2]

[1] University of Lausanne - CIG, Lausanne, Switzerland, [2] Swiss Institute of Bioinformatics - SIB, Lausanne, Switzerland

New genes are thought to have substantially contributed to phenotypic innovation. However, the mechanisms governing their functional evolution are poorly understood. To illuminate the dynamics of new gene evolution, we investigated so-called retrogenes, which originate as intronless copies of their "parental" source genes through an RNA-mediated mechanism. These copies, usually devoid of the parental promoter, need to gain new regulatory elements and, potentially, new exons/introns to evolve into functional retrogenes, making them ideal models to study new gene origination. Here, we explored retrogene evolution in nine representative mammals and one bird using extensive transcriptome and chromatin modification data. We find that regulatory elements of retrogenes were recruited from other genes, inherited from their parental genes, or, most frequently, obtained from genomic elements in their vicinity (e.g., CpG islands and enhancers). Remarkably, we observed that retrogenes may rapidly evolve new multiexonic structures and may even undergo alternative splicing. Finally, we investigated the functional relevance of mammalian retrogenes, based on evolutionary gene expression analyses, and identified retrogenes that may have contributed to the specific organ biology of different mammalian lineages. Notably, we identified a number of "orphan" retrogenes that functionally replaced their parental genes during evolution. Altogether, our work highlights how intronless gene copies, usually initially devoid of regulatory elements, can evolve into actively transcribed, complex multiexonic genes within a short evolutionary time. It thus provides novel insights into the general mechanisms underlying the origins and functional evolution of new genes.

237A

Just a one dimensional view of it all: is there a hard limit to understanding molecular evolution using only frequency-based analytics of DNA sequence?

Gregory Babbitt, Mohammed Alawad
Rochester institute of Technology, Rochester NY, USA

Frequency-based analyses in biology have intellectual precedence all the way back to the earliest genetic experimentalists and theoretical evolutionists who founded the "frequentist" paradigm in modern statistics. In recent decades, we have heavily applied this way of thinking to molecular evolutionary research, defining mutational events as abstract symbolic alterations to one dimensional strings representing biological polymers. Deep down, we acknowledge there is an underlying molecular behavior that ultimately shapes observed frequencies of mutation (e.g. notions of mutational "tolerance" and the existence of varying substitution penalties in alignment algorithms). Yet we seldom ask what sorts of molecular dynamics might be driving spectra of mutation through the processes of selection and drift. Comparisons of PDB structures are often added to evolutionary studies, but it is often difficult to deduce how changes in atom position impact function. While the reduction of molecular biology to this single-dimension has proven powerful in past decades, when technologies were developing, it is now time to embrace a more realistic view of mutation; one that is grounded in molecular dynamic modeling and/or biophysical experimentation. Here, we will review examples of our lab's recent work to quantify mutational impact on DNA polymer biophysics and its relation to nucleosome formation, synonymous codon bias, and organization of the genetic code. We will also present new results demonstrating how mutations variably impact other levels of the central dogma (RNA self-interactions and protein ordering). We propose some novel ways to locally detect natural selection acting singularly or collectively on these molecular dynamic traits.

238B

How do *de novo* genes recruit downstream interacting genes and reshape global gene networks?

Chen Xie, Predrag Kalajdzic, Rafik Neme, Cemalettin Bekpen, Sven Künzel, Diethard Tautz
Max Planck Institute for Evolutionary Biology, Plön, Germany

De novo genes are molecular entities which originated recently in evolutionary history of a given lineage from non-genic ancestral sequences. They can quickly acquire essential functions, probably through recruitment of downstream interactors and by reshaping global gene networks. These aspects have not been extensively addressed in previous studies. We systematically identified *de novo* transcripts from different populations and species of the genus *Mus* based on RNA-Seq data. Transcripts belonging to different age groups cover a range of evolutionary distances between a few thousand to a few million years. We performed gene knock-out experiments of the young *de novo* transcripts expressed in cell lines using the CRISPR/Cas9 technology. We obtained transcriptome data from these knock-out cell lines, and analyzed differential expression between normal and knock-out cell lines. We further investigated gene networks by integrating data retrieved from manually curated pathway databases, protein-protein interactions, co-expression networks, and regulatory networks.

239C

Evolution of structural and functional innovation in a family of deubiquitinating enzymes

Caitlyn Vlasschaert[1,2], Xuhua Xia[2], Douglas Gray [1,2]

[1] Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, [2] University of Ottawa, Ottawa, Ontario, Canada

The human genome encodes more than 50 highly specialized ubiquitin-specific proteases (USPs), a subclass of deubiquitinating enzymes with a characteristic catalytic domain. It is likely that the current panoply of USP genes emerged from a smaller number through a series of poorly understood gene duplication events. We have utilized bioinformatics and laboratory-based approaches in a study of the closely related USP4, USP15, and USP11 enzymes (the first of which was discovered in our laboratory). Based on published information these three enzymes have partially redundant function, but neofunctionalization has occurred during the course of their divergence and each has substrates that are unique. Analysis by a variety of computational methods has revealed that existing phylogenies (based solely on protein sequence identity) are incorrect. USP4 and USP15 arose from whole genome duplication prior to the emergence of jawed vertebrates, whereas USP11 was subsequently generated by a small scale duplication of the USP4 locus after which USP11 rapidly diverged or in some species was lost. Mice have been obtained that have inactivating mutations of USP4 or USP15; such mice are viable but there is significant deviation from expected Mendelian ratios when the strains are crossed. Our data suggest that USP4 and USP15 share functions that are essential , whereas the innovations of associated with USP11 are dispensable. The scope and nature of these functions will be discussed.

240D

TEMPLE: A new tool to analyze genetic diversity at transcription factor binding sites.

Stefan Laurent [1], Maria Litovchenko[2], Jeffrey Jensen[1]
[1] EPFL, Lausanne, Switzerland, [2] LMU, Munich, Germany

Cis-regulatory modules and, more specifically, the Protein-DNA interactions between transcription factors (TF) and TF binding sites (TFBS) play a key role in the determination of gene expression. Mutations within these regions can modify the transcriptional output in very different ways, ranging from subtle tissue-specific modifications to binary on/off responses in the whole organism. Therefore, neutral and selective forces are expected to have shaped the genetic variation at TFBS, and describing the distribution of fitness effects of these mutations will improve our understanding of regulatory evolution. Here we present TEMPLE, new software that allows easy visualization and analysis of the genetic variation occurring at the level of bioinformatically predicted transcription factor binding sites (TFBS). TEMPLE facilitates the functional characterization of genetic variation in regulatory sequences and identifies candidate TFBS that are characterized by strong genetic differentiation among populations. We here use TEMPLE to describe genetic variation found within ChiP-Seq peaks of several TF in Humans and Drosophila melanogaster and evaluate the contribution of indel and SNP polymorphism to TFBS differentiation across populations.

241A

Tandem repeats in coding region constitute another mechanism of adaptive radiation in cichlid fish

Langyu Gu, Walter Salzburger
Basel University, Basel, Switzerland

Convergent evolution provides an ideal scenario for testing the role of natural and sexual selection in adaptive radiations. However, to what extent the same genes and genetic pathways or different ones contribute to convergent phenotypes is still unclear. Cichlid fishes, which feature mutliple convergent phenotypes, provide an ideal model system to answer the questions why certain key traits are important in some lineages but not in others, and how these phenotypes evolved and are maintained developmentally. Egg-dummies are an evolutionary key innovation of East African cichlids that have been suggested to contribute to their evolutionary success. Using comparative transcriptomic and genomic analyses, we identified an egg-dummy candidate gene showing cichlid-specific tandem repeats in its coding region, which represent a a functional domain and are under positive selection. More interestingly, the polymorphism of these repeats is associated with species richness. We hypothesize that the repeats enlarge the available binding properties of this gene, constituing yet another mechanism of adaptive divergence in cichlids.

242B

Genome Architecture Changes in Archaic and Modern Humans

Rebekah Rogers
UC Berkeley, Berkeley, CA, USA

Changes in genome structure are a source of genetic novelty that can generate new genes or alter expression patterns for adjacent genes. Using paired-end read mapping in Illumina sequence data for archaic DNA, we identify 1005 changes in genome architecture between the Altai Neanderthal and the modern human reference genome as well as 1350 changes in genome architecture between Denisovan and the modern human reference. Regions with altered genome structure lie within 10 kb of a total of 234 genes in Denisovan and 180 genes in Neanderthal. These changes in genome structure between Neanderthals and modern humans are associated with testis-specific expression in modern humans, as well as testis-specific changes between chimpanzees and modern humans. Such associations with testes-specific expression are consistent with the "out-of-the-testes" hypothesis that new genes commonly originate in the testes. Using these female archaic samples, we identify 159 variants across the two archaic human genomes that indicate genomic traffic between the autosomes and the Y, raising the possibility of sex-specific changes in the modern and archaic human lineages. Additionally, we clearly identify human-specific cases of ectopic recombination, including one case of genome shuffling surrounding olfactory receptors on chromosomes 14 and 15 that can now be dated to after the divergence of these archaic and modern human reference genomes. These results suggest that changes in genome architecture between archaic and modern humans are one common form of variation that contributes to genomic differentiation of modern and archaic humans.

243C

Protein evolution: dN/dS, radical and conservative substitutions, and mutation-selection models

Claudia C. Weber [1,2], Ashley I. Teufel[1,3], Simon Whelan[2], David A. Liberles[1]
[1] 1. Department of Biology, Temple University, Philadelphia, PA 19122, USA, [2] 2. Evolutionary Biology Centre, Uppsala University, 752 36 Uppsala, Sweden, [3] 3. Department of Molecular Biology, University of Wyoming, Laramie, WY 82072, USA

Elucidating how novel molecular traits arise requires understanding the functional effects of amino acid substitutions. However, many widely used approaches do not fully leverage the available data.

A common method of estimating selection pressure on proteins is dN/dS, the ratio of non-synonymous to synonymous nucleotide substitutions. However, this may be coarse-grained, as some non-synonymous changes might have more severe effects on protein function than others, depending on the physicochemical properties of the amino acids and their locations within proteins.

Information about side-chain properties can be incorporated by comparing the rate ratio of radical and
conservative amino acid substitutions (dR/dC). We find that incorporating dR/dC into a codon model
improves model fit for taxa with high Ne. In contrast to count-based methods, the estimator is robust to varying base composition and Ts/Tv.

Although mutation-selection models that estimate selection coefficients are becoming more widely used, they still disregard key process information such as linkage between sites, variation in Ne, or underlying structural information. To address some of these limitations, we can build codon models that account for linkage and shifting Ne across lineages, as well as protein structural effects, in a mutation-selection model framework. Preliminary evidence shows that considering these factors is important when they are acting. This work is being extended in a phylogenetic context to evaluate inference of selective pressures employing these models.

Considering dN/dS, dR/dC, and mutation-selection models will provide insight into selective forces acting on proteins as greater biological complexity and statistical formalism come together in model improvement.

244D

ORFans with potential viral origin and mitochondrial inheritance

Liliana Milani, Fabrizio Ghiselli, Marco Passamonti
University of Bologna, Bologna, Italy

ORFans are open reading frames having no detectable similarity to known sequences. These putative genes have the shared feature to be lineage-specific, that means they are conserved within a taxonomic group. In the last few years, novel mitochondrial ORFans have been identified and partially characterized in several animals. A case study involves ORFan genes detected in the mitochondrial genomes of animals with doubly uniparental inheritance (DUI) of mitochondria, a system characterized by two mitochondrial lineages, transmitted separately by the two sexes. Several *in silico* analyses on these elements suggest an origin through endogenization of viral sequences, whose insertion at different evolutionary times possibly led to the ORFan sequence variability found in the different analyzed species. The most deeply studied DUI ORFan belongs to the male-transmitted mitochondrial genome of the Manila clam, *Ruditapes philippinarum*. Its germ line-specific transcription and translation during male gametogenesis was well documented, but it is still unclear whether its occurrence in gametes is just a way the parasite sequence evolved to spread through generations or the sequence was co-opted for some other function. What is clear is that the product of this gene is present in germ cells since their first proliferation and is stored in sperm mitochondria. Differently from species with strictly maternal inheritance, in DUI organisms sperm mitochondria are transmitted from father to son, in a process that have been lasting for hundred million years, and whose exceptionality might have been caused by the acquisition of a viral sequence.

245A

Genetic innovation through duplication in humans and Great Apes

Marina Brasó-Vives [1], Diego Hartasánchez[1], Arcadi Navarro[1 ,2]

[1] Institute of Evolutionary Biology (Universitat Pompeu Fabra – CSIC), Barcelona, Spain, [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain, [3] Centre for Genomic Regulation (CRG), Barcelona, Spain, [4] National Institute for Bioinformatics (INB), Barcelona, Spain

Duplications are a very important feature of eukaryotic genomes and an essential source of genetic innovation. Large (1-200 Kb) and highly identical (≥90%) duplications, known as segmental duplications (SDs), conform around 5% of the human genome and can include several functional elements. Despite their fundamental role in the generation of novel genetic material, most of the mechanisms underlying the molecular evolution of nucleotide sequences of duplicated regions are still not well understood. Two outstanding gaps in our knowledge are the way in which these regions undergo concerted evolution and the non-standard signature that natural selection may leave on their patterns of variation. Under neutrality, differences between duplicates are expected to follow a stochastic distribution along the duplicated region. Analyzing Great Ape and human genomes, we have identified some cases of coincident patterns of divergence between SD copies across species, which may be indicative of a common selective pressure acting independently on the same pair of SDs on different branches. Furthermore, SDs suffer a decay of identity around their edges with time. By analyzing the flanking regions of currently annotated SDs, we have been able to identify regions that had once undergone concerted evolution and have now become single-copy regions of new formation and possible novel function. We show that duplication-specific approaches allow for the identification of regions of genetic innovation that may have been under selective pressure during recent human evolution and have contributed to a better understanding of the mechanisms through which new genes and functions arise within SDs.

246B

Characterization of *de novo* gene evolution across the mammalian phylogeny.

José Luis Villanueva-Cañas [1,2], M. Mar Albà[1,3]

[1] Fundació Hospital del Mar Research Institute (FIMIM), Barcelona, Spain, [2] Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

Mammals have managed to conquer almost every ecosystem on earth, developing some very special adaptations along the way such as homeothermy, hair, specialized skin with integumentary glands, flight, and echolocation among others. A number of molecular evolutionary innovations underlie this exuberant adaptive diversification.

Whereas some mammalian evolutionary innovations are related to changes in gene regulation, others involve the birth of new functional genes. A fraction of these new genes have originated by gene duplication, but others have arisen *de novo* from previously non-coding sequences. *De novo* gene evolution is attracting the attention of the scientific community as the number of described lineage- or species-specific genes in different organisms continues to grow. In order to quantify and characterize innovation due to *de novo* gene evolution in mammals we have performed a comprehensive study using 68 mammalian genomes. Performing sequence similarity searches against gene annotations and RNAseq-derived transcripts, we identified 13,063 unique orphan gene birth events which occurred in different mammalian branches over the past 200 My. We present an analysis of the tissue expression patterns of these genes using GTEX and mouse ENCODE data.

247C

Evolution of metabolic pathways in nematodes: integrating genomics and metabolomics

Gabriel Markov [1], Jan Meyer[1], Oishika Panda[2], Hanh Witte[1], Frank Schroeder[2], Ralf Sommer[1]
[1] Max-Planck Institute for Developmental Biology, Tuebingen, Germany, [2] Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA

Metabolic pathways are an important subject for evolutionary questions, because enzymes provide a direct link between genotype and phenotype. To evaluate the conservation of a given pathway between species, we have to test for the conservation of both, the enzymes and the metabolites. In addition, the catalytic activity for each step in the pathway has to be tested at the level of individual enzymes. Our detailed comparative genomic approach in nematodes indicates that there is a very high level of lineage-specific gene duplications among enzymes, with one-to-one orthologs being the exception rather than the rule. Complementary metabolomic approaches also revealed the existence of lineage-specific metabolites. Therefore, it is crucial to determine which of those specific metabolites are involved in a conserved biological process. The induction of a conserved life-history transition (dauer entry) by different small molecules from the same family (ascarosides) in *Caenorhabditis elegans* and *Pristionchus pacificus* offers such a framework. Small-molecule synthesis in *C. elegans* requires *daf-22* encoding for an enzyme involved in beta-oxidation of fatty acids. We demonstrate that those shared precursors are not synthesized by orthologous enzymes in *P. pacificus* and *C. elegans*. Paralogous genes also differ in their protein domain structure. Using differential metabolomic profiling of novel CRISPR-Cas9 generated mutants for the two *P. pacificus* paralogous copies of *Cel-daf-22*, we show that the structurally more divergent paralog is more conserved at the functional level, and that the shift in substrate specificity also correlates with differences in the length of the fatty-acid chain of the dauer-inducing molecules.

248D

Tracing the Evolutionary History of the SLC1 Gene Family

André Lehnherr, Matthias Gesemann, Stephan C.F. Neuhauss
Institute of Molecular Life Sciences, Univerität Zürich, Zürich, Switzerland

Efficient removal of glutamate from the synaptic cleft is essential to end synaptic transmission protecting neurons from excitotoxicity. This is achieved by excitatory amino acid transporters (EAAT), which belong to the Solute Carrier Family 1 (SLC1) gene family. Besides glutamate and associated cation transport, EAATs induce a chloride conductance upon glutamate binding. Phylogenetic analysis of the SLC1 gene family revealed duplication and deletion events during evolution leading to varying number of SLC1 genes in different species. While the zebrafish genome contains thirteen SLC1 genes, amphibians have retained nine and mice and humans only seven SLC1 genes.
In this study, we compare expression and function of selected vertebrate SLC1 genes across species (Zebrafish, Xenopus, anole lizard, chicken and mouse) in the retina. Retinal expression of SLC1 RNA provides an ideal system to compare the abundance of SLC1s in different vertebrate species, identifying potential sub-, non or neofunctionalization events. In fact we found striking differences in RNA expression between different species. For further examination of these differences, we elaborate the biophysical properties of the SLC1 family members by expression of the transporters in Xenopus oocytes and subsequent two electrode voltage clamp recordings. This allows us to correlate expression and function across species and thereby follow the evolutionary history of the SLC1 gene family.

249A

Collapsed duplications? What to expect and what to look for.

Diego A. Hartasánchez [1], Marina Brasó-Vives[1], Arcadi Navarro[1 ,2]

[1] Institute of Evolutionary Biology (Universitat Pompeu Fabra - CSIC), Barcelona, Spain, [2] National Institute for Bioinformatics (INB), Barcelona, Spain, [3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain, [4] Centre for Genomic Regulation (CRG), Barcelona, Spain

Gene duplication has been recognized as being a source of evolutionary novelty for a long time. Theoretical modeling of the concerted evolution of duplicates has been very successful in explaining the variation present within gene families. However, the same cannot be said of segmental duplications (SDs), which encompass large fragments (more than 1 kb) of the genome and comprise coding and non-coding regions. Despite extensive experimental data describing SDs and the recognition of their relevance from an evolutionary perspective, efforts to understand their molecular evolution and to develop tests to determine the evolutionary pressures they have undergone are scarce. Genome-wide scans for selection, for example, typically avoid regions of known SDs, since the models underlying most tests assume regions to be single-copy, and because sequencing repetitive regions is known to be problematic. Given this framework, we have modeled and analyzed the effect of concerted evolution of SDs on common statistical tests. Our aim is two-fold: first, to describe the type of signature imprinted by natural selection on duplicated regions when these are collapsed, which is common for unknown duplications; and second, to study the effect of recombination (in particular, interlocus gene conversion) between duplicates along the fixation process of the SD. Our results show that unidentified duplications can render confounding results if collapsed when building genome assemblies, and on the other hand, that by collapsing duplications, one can actually extract relevant information about their evolution.

250B

Unravelling the rapid molecular rewiring needed for the evolution of parthenogenesis

Philipp Schiffer [1,6], Etienne Danchin[2], Ann Burnell[3], Michael Thorne[4], Simon Wong[5], Thomas Wiehe[1]

[1] Universität zu Köln, Köln, RHEINLAND, Germany, [2] INRA, Sophia-Antipolis, France, [3] Maynooth University, Co. Kildare, Ireland, [4] British Antarctic Survey, Cambridge, UK, [5] ICHEC, Dublin, Ireland, [6] EMBL, Heidelberg, Germany

Sexual reproduction is the main mode of reproduction in animals and requires a intricate system of genetic interactions. The switch from sexual to obligate parthenogenetic reproduction requires a major rewiring of Gene Regulatory Networks (GRNs) within very few generations. Still, in Nematoda parthenogenesis repeatedly evolved as a novel trait. It is known that many parthenogenetic animals are polyploid and evolved through hybridization, but this has not been analysed in detail on the genomic level.

We present data from Panagrolaimus nematodes, where amphimixis and automictic parthenogenesis are present. We sequenced the genomes of several species to explore genomic aspects of the evolution of parthenogenesis and the changes in the associated GRNs of development and reproduction. We also developed computational methods to detect a hybrid polyploid background in the genomes.

The assembled genomes serve as a basis for detailed studies into the molecular machinery allowing for the evolution and maintenance of parthenogenesis as a novel trait. The available diploid and polyploid genomes also allow us to further investigate the genomic consequences of hybrid polyploidisation.

We find large divergence in GRNs compared to the model nematode species C. elegans and gene expression differences between closely related Panagrolaimid species. In the light of the divergence from the model system we use RNA-Seq data to analyse the molecular toolkit in Panagrolaimus and to elucidate changes in parthenogens.

Our data illustrate that rapid turnover of genes in developmental GRNs of Nematoda and inter-species hybridisation appear to be tightly linked to the evolution of parthenogenesis in Panagrolaimus.

251C

Molecular innovations and convergent evolution of structural proteins in hair and feathers

Leopold Eckhart, Bettina Strasser, Veronika Mlitz, Erwin Tschachler
Medical University of Vienna, Vienna, Austria

Hair and feathers are major evolutionary innovations of mammals and birds, respectively. Both types of skin appendages consist of dead epidermal keratinocytes that are stabilized by extensive cross-linking of proteins in a process called cornification. Here, we investigated which molecular innovations were associated with the evolution of feathers and compared them to innovations associated with the evolution of hair. By candidate gene identification in recently sequenced avian genomes and gene expression analyses, we show that feathers contain cysteine-rich intermediate filament proteins and apparently unstructured proteins with even more extreme cysteine contents (>30%). This protein composition is similar to that of mammalian hair. Phylogenetic analysis suggests that the genes encoding cysteine-rich structural proteins of feathers and hair evolved from different ancestral genes, which were located in different clusters of genes with skin epithelium-specific expression. Genome sequence analysis of more than 30 phylogenetically diverse species of birds confirmed conservation of cysteine-rich feather proteins. The results of our study show that two important examples of evolutionary innovations at the level of tissue morphology, i.e. hair and feathers, depended on similar innovations in the molecular architecture of epithelial cells.

252D

Inferring selective constraint from population genomic data reveals recent regulatory turnover in the human brain

Daniel Schrider, Andrew Kern
Rutgers University, Piscataway, NJ, USA

A central problem in evolutionary genetics is to uncover recent adaptive changes to the genome that underlie important phenotypic differences between closely related species. We present a novel approach blending phylogenetics, population genomics, and machine learning in order to identify recent gain and loss of function in the genome. Our method seeks to identify stretches of sequence that have recently acquired selective constraint (i.e. gain of function), or lost constraint (loss of function), within a single species. Briefly, we use a powerful supervised machine learning technique called a support vector machine (SVM) to classify regions of the human genome as functional or nonfunctional based on their levels of diversity-a task our SVM performs with high accuracy on both real and simulated test data. We then contrast this information with phylogenetic evidence of selective constraint in order to reveal functional turnover. This approach can in principle identify the gain or loss of any type of functional DNA element, whether it be a protein-coding gene, a non-coding RNA, or a regulatory region. We applied this method to the human genome, finding that putative gains of function are highly enriched in human-specific regulatory regions. Strikingly, we also found a large excess of functional turnover in regulatory regions of genes essential for central nervous system development, many of which have evolved human-specific patterns of expression. Thus, our approach appears to reveal regions responsible for human-specific changes in brain development, underscoring its power for uncovering the genetic bases of evolutionary novelty.

253A

Resurrecting the Ancestral Structural Dynamics of an Antiviral Innate Immune Receptor: Rapid Changes in Structure and Function

Charles Pugh, Oralia Kolaczkowski, Bryan Korithoski, Bryan Kolaczkowski
University of Florida, Gainesville, FL, USA

The resurrection and functional characterization of ancestral proteins has provided an unprecedented view of how proteins evolve, not only improving our understanding of how particular protein families diversified but also informing the development of molecular-evolutionary theory. However, most mechanistic studies of molecular-functional evolution have focused on proteins that function in relatively stable biological processes such as metabolism and development. It is not clear that the insights gained through studying these systems translate to faster-evolving proteins, such as immune receptors. Here we present the first mechanistic investigation of the functional evolution of the RIG-like receptors (RLRs), a family of proteins that bind viral RNA in the cytoplasm and initiate innate immune responses. Using kinetic binding assays and molecular dynamics simulations, we demonstrate how RLRs altered their RNA-binding preference throughout early metazoan evolution by reorganizing the shape and electrostatic distribution across their RNA binding pocket. In contrast to what has been observed for proteins involved in metabolism and development, we find that RLR-RNA binding shifted multiple times in a relatively short evolutionary period and not coupled to gene duplication events. We demonstrate at least one reversion of RLR molecular function from a derived to an ancestral state through a novel structural mechanism, indicating the presence of multiple structural trajectories implementing similar functions. These data reveal the mechanistic signatures underlying an important evolutionary arms race between viral RNAs and host receptors and indicate that the evolutionary dynamics of immune receptors may be quite different from more ‘well behaved’ protein families.

254B

Whole genome duplication increased the adaptive potential of crop plants

Niv Sabath, Ayelet Salman-Minkov, Itay Mayrose
Tel Aviv University, Tel Aviv, Israel

Whole genome duplication (WGD) has long been hypothesized to alter species adaptation capabilities. To test this hypothesis, we examined the largest evolutionary experiment in history - plant domestication. Over thousands of years, humans have attempted to cultivate numerous plants species through directional selection to fit the requirement of a crop, but only a small fraction was found suitable. We used chromEvol, a probabilistic evolutionary method that describe the different pathways by which the evolution of chromosome number proceeds, to infer the location of WGD events along the phylogenies of 108 plant genera. This further allowed us to classify WGD transitions into ancient and recent events. We found that domesticated plants have gone through more ancient (but not recent) WGD events relative to their wild congeners. This pattern is apparent across both eudicots and monocots, trees and herbs, and is more apparent in perennial than in annual plants, suggesting that the genetic consequences of WGD have conferred many of these plants genetic preconditions that enabled successful domestication. Our results highlight the potential use of directed WGD in crop improvement efforts.

255C

Evidence for the conservation of transit peptide-independent protein import into mitochondria and hydrogenosomes

Sriram Garg [1], Jan Stölting[1], Verena Zimorski[1], Petr Rada[2], Jan Tachezy[2], William Martin[1], Sven Gould[1]

[1] Institut for Molekulare Evolution, Heinrich-Heine-Universitaät, Düsseldorf, Germany, [2] Department of Parasitology, Charles University in Prague, Faculty of Science, Prague, Czech Republic

The origin of the N-terminal targeting sequences (NTSs) of its cognate import substrates were key steps in the evolutionary transition from a proteobacterial endosymbiont to an eukaryotic organelle, yet the events surrounding the early history of mitochondrial import remains obscure. While the main translocon of the outer membrane, TOM40 is ubiquitous among organelles of mitochondrial origin events surrounding the origin of its accessory receptors and the corresponding NTSs are far less clear. To better understand the extent of evolutionary conservation in mitochondrial protein import we investigated the targeting behavior of *T.vaginalis* hydrogenosomal proteins in the yeast *S.cerevisiae* and yeast mitochondrial proteins in Trichomonas with and without their NTSs. Hydrogenosomes import yeast mitochondrial proteins even in the absence of their native NTSs while excluding cytosolic proteins. Conversely, mitochondria import hydrogenosomal proteins even without their short NTSs. The data indicate that only proteins germane to mitochondria posses structural properties that are sufficient to target them to into the matrix of mitochondrial organelles. This surprising conservation — from excavates to opisthokonts — of a mitochondrion specific, NTS–independent recognition and import route indicates they present in the eukaryote common ancestor representing a primordial form of mitochondrial protein targeting. Consistent with that view, proteomic profiling of Trichomonas hydrogenosomes indicates that 90 % of hydrogenosomal matrix proteins lack a recognizable N-terminal leader. The reduction of NTSs in hydrogenosomes corresponds to the loss of the cognate receptors during evolutionary reduction of the translocon in these organelles and a reversion of protein import to its simpler ancestral state.

256D

Evolutionary and functional characterizations of Human specific UV response and cellular aging - the SPATA31A gene family

Cemalettin Bekpen, Diethard Tautz
Max Planck Institute for Evolutionary Biology, Plön, Germany

Clustering analysis of segmentally duplicated regions in the human genome have suggested that some of the rapidly expanding duplication blocks have been formed around “core” duplicons. Interestingly, the core duplicons are highly transcribed and numerous human/great-ape specific gene families have recently been described that map to these core regions of the genome. We propose that core duplicon gene families confer a selective advantage in human evolution and may have a role in the function of the particular human conditions.

We investigate the structure, transcriptional diversity, and protein localization of the *SPATA31A* gene family, one of the core duplicon gene families in humans. *SPATA31A* genes encode one of the most rapidly evolving genes and have expanded to ~10 copies in the human genome from a single ancestral locus in Old World monkeys. Phylogenetic analysis and cDNA sequencing suggest that the *SPATA31A* gene family have two distinct subfamilies or subtypes showing recent adaptive evolution followed by segmental duplication. We showed that the protein product of *SPATA31A* gene families localize primarily to the nucleus and may be involved in the function of DNA repair mechanism upon stimulation by UV response.

We are using the state of the art technique CRISPR/Cas derived RNA guided endonucleases to engineer the *SPATA31A* genes from primary human cell line and mice both at the cellular and organismic level. Our initial analysis indicates that targeted disruption of the member of *SPATA31A* genes leads to faster duplication rate and prolong life span of HFF cells.

257A

Mechanisms of duplicate-gene evolution and preservation

Parul Johri, Jean-Francois Gout, Casey McGrath, Tom Doak, Michael Lynch
Indiana University, Bloomington, IN, USA

Gene duplication is an important source of evolutionary novelty, resulting in new gene functions and contributing to the emergence of new species. Although most new gene duplicates are rapidly lost from a population by pseudogenization, a number of duplicates are preserved either exactly or are retained with a change in function. However, it is not well understood why certain genes are more prone to retention than others and what evolutionary forces are responsible for long-term retention of duplicates. To elucidate the evolutionary forces influencing the fate of gene duplicates, we need to understand the evolution of duplicates in the early post-duplication stages at the level of populations. We propose to do this for duplicates created by whole-genome duplications (WGDs) by comparing the genomes of closely-related species with and without WGDs, and examining variation within these species.
The *Paramecium aurelia* complex consists of 15 sibling species that experienced two rounds of WGDs before the species radiation and continue to retain about 50% of paralogs created by the most recent WGD. We sequenced, assembled and annotated three of the species belonging to the complex, along with two outgroups that did not share the WGD. We have now also sequenced about 10-13 isolates (sampled world-wide) of each of these species and are currently in the process of examining the levels of polymorphism and divergence across the genome. This will help us dissect the evolutionary forces operating on the paralogs and give insights into the mechanisms responsible for their preservation and loss.

258B

Recurrent evolution of tRNA and mRNA editing in animal mitochondria

Dennis Lavrov
Iowa State University, Ames, Iowa, USA

One of the unusual features of organellar DNA in general and mitochondrial DNA in particular is a frequent occurrence of RNA editing. The term 'RNA editing' refers to a variety of mechanistically unrelated biochemical processes that evolved independently in diverse eukaryotic groups and alter RNA sequence during or after transcription. Best-known examples of RNA editing in animal mitochondria are those of transfer RNA. Mt-tRNA editing occurs in a variety of animal lineages and usually involves the 3' end of the tRNA molecule. At its extreme, up to a half of each molecule can be added at the RNA level. In contrast to tRNA editing, mRNA editing is less common in animal mitochondria. However several examples of this process have been reported in representatives of placozoans, nematodes, and tunicates. Here we describe a pan-editing of mitochondrial transcripts in calcaronean sponges that appears to be a synapomorphy for this group. We show that mitochondrial mRNA editing in calcaronean sponges shares has some animal-specific feature, but likely occurs by a novel mechanism. We also demonstrate the presence of mRNA editing not only alters the pattern of expression of mitochondrial genes, but also exerts an extreme pressure on evolution of mitochondrial coding sequences in this group.

259C

Paralog interference shaped the evolution of homospermidine synthase

Elisabeth Kaltenegger, Dietrich Ober
Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Homospermidine synthase (HSS) catalyses the first specific step of pyrrolizidine alkaloids, which are part of the plant's chemical defence against herbivore. HSS evolved by duplication of the conserved deoxyhypusine synthase (DHS). DHS is involved in the posttranslational activation of the eukaryotic initiation factor eIF5a and essential for proliferation of eukaryotic cells. DHS is only active when it assembles into a homotetramer thereby forming the active site in the interface of the interacting subunits. At least 5 independent duplication events led to evolution of HSS within the plant kingdom. In the family of Convolvulaceae surprisingly a period of purifying selection was detected immediately after the duplication. This is contradictory to most duplication models as they predict a release of functional constraints after duplication. An interference of the paralogs at protein level is suggested to mediate this period of purifying selection. When a gene encoding a self-interacting enzyme is duplicated, the gene products initially can cross-interact and form a paralogous heteromer in which the duplicate copies are physically and functionally connected. Deleterious mutations in one copy are predicted to have a dominant-negative effect and to be purged by selection.

We postulate that paralog interference contributes to the stabilization of duplicated copies of homomeric enzymes thereby extending the temporal window during which neofunctionalization can evolve. The effect of paralog interference will be strongest when the activity of the protein depends on the interaction of the subunits like enzymes with active sites in their interface or transcription factors which bind DNA as dimers.

260D

Functional diversification of an ecologically important gene family: Serine proteases in Daphnia

Jeffry L. Dudycha
University of South Carolina, Columbia, SC, USA

The freshwater grazing herbivore Daphnia has long been an important model in the ecology of consumer-resource interactions. In an initial attempt to identify genes involved in exploitation of different resources, serine proteases emerged as a potentially important gene family. Serine proteases are enzymes that cleave peptide bonds and serve multiple functions, including digestion, immune responses, and clotting. We characterized all of the serine proteases in the Daphnia pulex genome, and identified 73 trypsins, 1 elastase, 14 chymotrypsins. These classifications reflect different biochemical substrates that are cleaved. We also identified 18 potentially novel serine proteases whose substrate could not be classified from standard motifs, and a further 108 genes that had lost one or more elements necessary for proteolytic function. Many of these 108 genes continue to be expressed, however. Phylogenetic analysis identified multiple evolutionary origins of chymotrypsins. We used microarrays to determine that the majority of serine proteases with characteristics of digestive enzymes were up-regulated under low food conditions. Interestingly, other serine proteases tended to be up-regulated under high food conditions. We also traced functional diversification of serine proteases through the acquisition of additional functional domains, such as chitin-binding and clip domains. A handful of genes acquired individual sets of multiple functional domains. In a comparative analysis with the 200M yr divergent Daphnia magna genome, we were able to infer the ancestral complement of serine proteases and track the gain and loss of molecular function.

261A

Comparative transcriptomics in caecilian amphibians

María Torres-Sánchez[1], Christopher Creevey[2], David Gower[3], Mark Wilkinson[3], Diego San Mauro[1]

[1] Complutense University of Madrid, Madrid, Spain, [2] Aberystwyth University, Aberystwyth, UK, [3] The Natural History Museum, London, UK

Next-generation sequencing (NGS) technologies are rapidly transforming the study of evolutionary and comparative genetics, offering an unprecedented opportunity to characterize variation and diversity in both model and non-model organisms. One of the recent revolutions is the use of high-throughput RNA sequencing (RNAseq) to analyse transcriptomes at the nucleotide level. In a comparative context, all this knowledge becomes key to unfold the evolution of biological diversity. We are studying tissue-specific transcriptomic landscapes in several species of caecilians (order Gymnophiona), an ancient group of tropical limbless amphibians with over 300 million years of independent evolution from frogs and salamanders. In total, we have generated two billion reads from 40 tissue samples from across five species of caecilians of four different families (Caeciliidae, Rhinatrematidae, Siphonopidae, Typhlonectidae) spanning various degrees of evolutionary divergence. This information is allowing for gene discovery and characterization of the functional elements of the genome underlying particular adaptations. We are particularly interested in the origin and diversification of certain molecular constituents of the skin of caecilians, such as toxins, antimicrobial peptides, and other biologically active components. Furthermore, by comparing a wider range of tissues in a phylogenetic and comparative framework, we are searching for similarities and differences that can be readily interpreted in terms of ecology and evolution of specific adaptations (such as e.g. aquatic lifestyle).

262B

Reconstructing the evolution of gene repertoire in the Lachancea yeast clade

Nikolaos Vakirlis, Ingrid Lafontaine
UPMC, Paris, France

Yeasts from the *Lachancea* clade diverged from *Saccharomyces cerevisiae* before the ancestral whole genome duplication. The availability of complete and well-annotated genome sequences for 10 *Lachancea* species allow estimating the dynamics of the gene repertoire within an entire clade. We clustered the entire set of annotated CDS within the *Lachancea* genomes into homologous gene families and established a robust reference species phylogeny. We then used gene families and species phylogeny to infer the scenario of evolutionary events that occurred along the *Lachancea* tree. We showed that the vast majority of extant *Lachancea* genes were already present in the *Lachancea* common ancestor and that gene content variations occurred mainly by gene duplication and loss events (more than 2,500 duplication/loss events were characterized). However, a significant fraction of the genes (ca. 2%) has been gained since divergence from the *Lachancea* common ancestor, with several dozens of horizontally transferred genes and hundreds of potential gene creations. Our analysis shows that the impact of horizontal gene transfer has been so far underestimated in *Saccharomycetes* and that the study of *Lachancea* genomes is well suited to unravel the mechanisms of gene creation.

263C

Emergence of enzymatic function in a non-catalytic ancestor – the plant chalcone isomerase family

Miriam Kaltenbach [1], Jason R. Burke[2], Joseph P. Noel[2], Dan S. Tawfik[1]
[1] Weizmann Institute of Science, Rehovot, Israel, [2] Salk Institute for Biological Sciences, La Jolla, CA, USA

Chalcone isomerase (CHI), a key enzyme in plant flavonoid biosynthesis, is involved in the generation of thousands of compounds with functions such as protection from UV light or attraction of pollinators. CHI catalyses the conversion of an acyclic chalcone to the flavanone naringenin ($k_{cat}/K_M > 10^6$ $M^{-1}s^{-1}$). For a long time, CHI's evolutionary history remained a mystery, but recent evidence suggests that it evolved from a non-catalytic precursor. This ancestor relates to an older lineage of fatty-acid binding proteins (FAPs) as well as to another protein family of unknown function and with no CHI activity (CHILs for "CHI-like").[1] Therefore, studying CHI's evolution offers the unique opportunity to understand the emergence of catalysis in a non-enzymatic protein fold.

We reconstructed the CHI family ancestor, as well as the older CHI/CHIL common ancestor. Using directed evolution, we retraced the transition from non-catalytic to catalytic ancestor and characterized increasingly active variants obtained along the trajectory by enzymology, protein crystallography, and complementation of CHI knockouts in *Arabidopsis thaliana.*

We found that the CHI/CHIL common ancestor has no detectable CHI activity. However, catalysis emerged within a single point mutation. Intriguingly, this and other mutations along the trajectory were minor exchanges between hydrophobic amino acids, highlighting how subtle reshaping of an inert binding pocket can lay the foundation for an efficient enzyme. Future work is aimed at validating the evolving CHI activity in a living plant and at identifying the evolutionary origin and function of the CHIL family.

[1] Ngaki et al., *Nature* **2012**, *485*, 530-536

264D

Evolutionary plasticity of the *Plasmodium* export element: *Plasmodium* exportomes contain non-canonical PEXEL/HT proteins

Jana Schulze[1], Marcel Kwiatkowski[2], Janus Borner[1], Hartmut Schlüter[2], Iris Bruchhaus[3], Thorsten Burmester[1], Tobias Spielmann[3], Christian Pick [1]

[1] University of Hamburg, Institute of Zoology, D-20146 Hamburg, Germany, [2] Department of Clinical Chemistry, University Medical Center Hamburg-Eppendorf, D-20246 Hamburg, Germany, [3] Bernhard Nocht Institute for Tropical Medicine, D-20359 Hamburg, Germany

Plasmodium species remodel their host cell to promote asexual proliferation. These modifications are facilitated by exported effector proteins, collectively referred to as the exportome. Export of several hundred effector proteins is mediated by a protease cleavage site, known as *Plasmodium* export element (PEXEL) or host targeting (HT) signal. The PEXEL/HT is usually comprised of five amino acids, of which R at position 1, L at position 3 are important for proteolytic cleavage. Non-canonical PEXEL/HTs with K or H at the position 1 and/or I at position 3 are presently considered non-functional. Here, we demonstrate that such non-canonical PEXEL/HTs are cleaved and mediate efficient export of a REX3 and a GBP reporter, but not of a KAHRP reporter, indicating that their functionality depends on the sequence environment. We further demonstrate that proteins with a non-canonical PEXEL/HT are overrepresented in *P. falciparum* and other *Plasmodium* species and that export of some of these proteins in fact is mediated by non-canonical PEXEL/HTs. We conclude that the evolutionary plasticity of the PEXEL/HT is higher than appreciated and that non-canonical PEXEL/HT proteins account for considerably large proportions of *Plasmodium* exportomes.

265A

Sequencing the *Crotalus atrox* genome to identify the evolutionary origins of rattlesnake venom

Noah Dowell [1,2], Matt Giorgianni[1,2], Sean Carroll[1,2]
[1] *University of Wisconsin, Madison, Wisconsin, USA,* [2] *Howard Hughes Medical Institute, Madison, Wisconsin, USA*

The mechanisms underlying the origin of biological novelties are not well understood. Prey-killing venoms have evolved independently multiple times across the animal kingdom. Approximately half of the 6000 snake species are venomous but the evolutionary origins of snake venom proteins are unclear. Within rattlesnakes, there appears to have been a relatively rapid evolution of at least two distinct venom types. *Crotalus atrox*, a large bodied generalist, possesses hemorrhagic venom with a relatively high LD_{50} while *Crotalus scutulatus*, a small-bodied snake possesses a more potent, neurotoxic venom with a low LD_{50}. Using a comparative genomics approach, we sought to identify the genetic mechanisms through which these divergent venom types evolved. Sequence analysis of *C. atrox* and *C. scutulatus* venom loci revealed gene gains through duplication events and gene losses via mutation but the unexpected retention of the genetic capacity to produce a neurotoxic venom in *C. atrox*. Phylogenetic analysis combined with the transcriptome data allowed us to infer the order of molecular events that lead to the evolution of neurotoxic venom. First, gene duplication produced multiple copies of toxin genes (acidic and basic phospholipase A2s) with appropriate transcriptional regulatory elements, and subsequent neofunctionalization of an acidic-Pla2 yielded a neurotoxin in *C. scutulatus.* In contrast, in *C. atrox,* pseudogenization is likely to have erased the acidic-Pla2 while the basic-Pla2s remain. These results indicate that there has been dynamic evolution of rattlesnake venom genes over a relatively short time-scale.

266B

Evolution of substrate specificity in the alkaline phosphatase superfamily.

Magdalena Heberlein [1], Bert van Loo[1], Florian Hollfelder[2], Erich Bornberg-Bauer[1]
[1] Institute for Evolution and Biodiversity, Westfalian Wilhelms University Münster, Münster, Germany, [2] Department of Biochemistry, University of Cambridge, Cambridge, UK

Proteins have been present for billions of years and, during this time, have evolved to a wide range of highly diversified molecules that can perform their functions with great efficiency. The mechanisms that led to creation of all the current proteins from a limited repertoire of ancient genes required selection of certain evolutionary pathways.

Although not all the mutations appearing at a genomic level result in significant and measurable functional changes, many of them are deleterious or permissive and only become beneficial upon occurrence of further mutations (so called epistatic ratchet). Therefore, running into a dead-end when randomly exploring the sequence space is quite likely to happen.

In the presented research we want to learn which of the possible evolutionary paths are favoured by evolution and why. To address this question, we study two subfamilies – arylsulfatases (ASs) and phosphonate monoester hydrolases (PMHs)– the most closely related members of the alkaline phosphatase (AP) superfamily of highly promiscuous enzymes.

Using computational methods we have inferred and reconstructed the common ancestor of the extant ASs and PMHs. Synthesis of the reconstructed gene will enable the study of ancient protein properties in the laboratory. We also apply directed evolution methods to study the limits of specialization shown for ASs in previous studies. The generated mutant libraries are cloned into a special autodisplay system and expressed on the cell surfaceof E.coli. The single cells displaying enzyme molecules are compartmentalized in microdroplets together with a substrate and screened for enhanced activity and specificity.

267C

The nervous system of Xenacoelomorpha: a genomic perspective

Elena Perea-Atienza [1], Brenda Gavilán[1], Josep F. Abril[1,2], Pedro Martínez[1,3]
[1] Departament de Genètica, Universitat de Barcelona, Av. Diagonal, 643, 08028 Barcelona, Catalonia, Spain, [2] Institut de Biomedicina de la Universitat de Barcelona (IBUB), Av. Diagonal, 643, 08028 Barcelona, Catalonia, Spain, [3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08010 Barcelona, Catalonia, Spain

Phylogenetic studies maintain that acoelomorph flatworms (acoels and nemertodermatids) constitute a sister group of Xenoturbellida, forming the monophyletic phylum Xenacoelomorpha. Given the relative phylogenetic positions of its constituent three clades, with Xenoturbella as the earliest branch followed by the split of Nemertodermatida and Acoela, Xenacoelomorpha has becomes an interesting animal group for studying evolutionary trends (whether genomic or morphological).

The major aim of our research group is to understand the structure, evolution and development of nervous systems. Since the members of the clade Xenacoelomorpha present different degrees of "cephalization", analyzing the nervous system development inside this phylum could provide us with insights about the early organization of the bilaterian nervous system and the origin and formation of 'cephalized' nervous systems (brains).

Recently, the involvement of several research groups (included ours) in sequencing several xenacoelomorph genomes has allowed us to initiate molecular evolutionary studies of some specific gene families. We present here the characterization of three gene families involved in several aspects of the nervous system's formation: the basic helix-loop-helix (bHLHs), G protein-coupled receptors (GPCRs) and Wnts. We have focused our analysis in the acoel *Symsagittifera roscoffensis* and the xenoturbellid *Xenoturbella bocki* (the most complete genomes we have). How their evolutionary history is reflected in the progressive degree of 'cephalization' seen in the phylum constitutes the target of our study. In parallel, several new techniques have been developed in the lab to better map the detailed structures of the nervous tissue in these organisms.

268D

POSTER: Identification of recently evolved genes in human and chimpanzee using deep transcriptomics.

Jorge Ruiz-Orera [1], Jéssica Hernández[2], Tomàs Marquès-Bonet[2], Cristina Chiva[3], Eduard Sabidó[3], M Mar Albà[1,4]

[1] Evolutionary Genomics Group, Research Programme on Biomedical Informatics, Hospital del Mar Research Institute, Universitat Pompeu Fabra, Barcelona, Spain, [2] Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Spain, [3] Proteomics Unit, Centre de Regulació Genòmica (CRG), Barcelona, Spain, [4] Catalan Institution for Research and Advanced Studies, Barcelona, Spain

For a very long time, major mechanisms driving the evolution of new genes were thought to be restricted to gene duplication or rearrangements in existing protein-coding material. Nevertheless, recent comparative genomic analyses have shown that some genes are originated de novo from previously non-functional genomic sequences and that may be related with the emergence of species or lineage specific adaptations.

However, identification of recently evolved genes is challenging since such young genes evolve rapidly, are poorly expressed, tissue-restricted and they contain short open reading frames. Therefore, most young proteins are expected not to be annotated or to be misclassified as non-coding since they often exhibit similar properties.

We integrated strand-specific RNA-seq data from four different species, human, chimpanzee, macaque and mouse, generated in our laboratory with other existing RNA-seq datasets to investigate the presence of recently evolved genes in human and chimpanzee. By excluding genes with homologues in other species by a series of thorough searches, we obtain a collection of genes originated in human, chimpanzee or the common human-chimpanzee branch. We find a large number of de novo genes expressed in testis, and some genes also in brain, liver and heart. By combining data from proteogenomics analyses, ribosome profiling and coding potential statistics we identify a set of recently evolved genes that may encode novel proteins. We further find enrichment in certain transposable elements and regulatory motifs in promoters from young genes, which may be related with the birth and the tissue-restricted expression patterns of these new genes.

269A

Investigating the mechanisms underlying the origin of orphan genes in yeast

William Blevins [1,2], Mar Albà[1,2], Lucas Carey[2]
[1] *IMIM, Barcelona, Spain,* [2] *UPF, Barcelona, Spain*

Orphan genes i.e. taxonomically restricted genes or de novo genes, are genes that have no detectable homology with genes from other species. We generally expect to be able to detect homologues for genes that arose from gene duplication, recombination, and horizontal-gene-transfer events, so where did these genes come from? There are several theories which could account for the lack of homologues; the most interesting of which is de novo gene birth. In de novo gene emergence, segments of non-coding DNA undergo a series of mutations that enable transcription and potentially lead to new proteins with novel functions.

As de novo genes are typically short and expressed at low levels, deep RNA sequencing of several closely-related species is necessary to reliably distinguish de novo transcripts from transcriptional noise. We have sequenced the transcriptomes of 12 species of yeast in both rich medium and oxidative stress conditions. This data enables us to answer standing questions surrounding de novo genes such as: Are the mechanisms involved in de novo gene emergence and fixation affected by stress conditions? Do de novo genes appear more frequently near bidirectional promoters? Which genomic features are the best predictors of the future transcription of proto genes? Our objective is to evaluate the existing hypotheses regarding the mechanisms responsible for de novo gene emergence, and to propose new mechanisms where necessary.

270B

Network-level architecture and the evolutionary potential of underground metabolism

Richard Notebaart[1], Balázs Szappanos[2], Bálint Kintses[2], Ferenc Pál [2], Ádám Györkei[2], Balázs Bogos[2], Viktória Lázár[2], Réka Spohn[2], Allon Wagner[3], Eytan Ruppin[3 ,4], Csaba Pál[2], Balázs Papp[2]
[1] Department of Bioinformatics (CMBI), Centre for Molecular Life Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands, [2] Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary, [3] The Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel, [4] The Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

A central unresolved issue of metabolic network evolution is to understand how innovations emerge that produce novel adaptive phenotypes. It is widely stated that weak 'underground' catalytic side activities of enzymes can provide raw material for the evolution of novel functions. Such underground reactions appear to be frequent, however, it remains unknown to what extent this raw material could generate evolutionary novelties in the context of the entire metabolic network. Here, we computationally reconstructed the first underground metabolic network of *E. coli* by compiling the known underground reactions into a comprehensive genome-scale metabolic model. We revealed that most underground reactions are not isolated and nearly half of them are completely connected into the metabolic network. Furthermore, we show that many of these underground reactions can form novel pathways producing key biomass precursors and, typically, these pathways have similar properties to the native ones under standard conditions. We estimate that under specific environments at least ~20% of the connected underground reactions confer a fitness advantage when their activity is enhanced. *In silico* predictions of novel phenotypes showed significant agreement with the *in vivo* evolutionary potential characterized by our genome-wide gene overexpression screen. These findings demonstrate for the first time that the genetic basis of evolutionary adaptations via underground metabolism are predictable.

271C

The players of the "collaborative non-self recognition" *Malus* self-incompatibility system

Jorge Vieira[1,2], Bruno Aguiar[1,2], Înes Pratas[1,2], Vanessa Nunes[1,2], Cristina Vieira [1,2]
[1] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [2] Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal

In flowering plants, a widespread mechanism to prevent inbreeding is gametophytic self-incompatibility (GSI), where the self-incompatible phenotype of the pollen is determined by its haploid genome. In Solanaceae, Plantaginaceae, Rubiaceae, and Rosaceae species pistil specificity is determined by an extracellular ribonuclease, called *S-RNase*. The male determinant, always a F-box gene, can be determined by one gene, as in the self recognition *Prunus* (Rosaceae) system, or multiple genes, as in the non-self recognitions Pyreae (Rosaceae; called *SFBB*s) and Solanaceae (called *SLF*s) systems. In the collaborative non-self recognition system, each of the multiple *S*-pollen genes interacts with only a subset of its non-self *S-RNases*, and none of them can interact with their respective self-*S-RNase*. Therefore, in such systems, a large number of *S*-pollen genes must exist in a single *S*-haplotype. Presently, in Pyreae 16 genes have been identified using both PCR approach using primers for conserved regions from genomic DNA and BAC libraries. To identify the most complete set of *Malus x domestica SFBB* genes, we used a de novo RNA-seq approach to analyze the pollen transcriptomes of 10 *S*-haplotypes, as well as the young leaf, ovary, stigma, style, sepal, petal transcriptomes of Golden delicious (*S*2*S*3). High diversity and/or deletion of *SFBB* genes as predicting targets of non self-*S-RNases* are not a common mechanism of recognition in *M. domestica,* as recently proposed in Solanaceae.

272D

Fabaceae Gametophytic self-incompatibility is not determined by Rosaceae, Solanaceae, and Plantaginaceae *S* - *RNase* lineage genes

Jorge Vieira [1,2], Bruno Aguiar[1,2], Ana Cunha[1,2], Vanessa Teixeira[1,2], Cristina Vieira[1,2]
[1] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [2] Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal

Fabaceae species are important in agronomy and livestock nourishment. They have a long breeding history, and most cultivars have lost self-incompatibly (SI), a genetic barrier to self-fertilization. To improve legume crop breeding, crosses with wild SI relatives of the cultivated varieties are often performed. Therefore, it is fundamental to characterize Fabaceae SI system(s). We address the hypothesis of Fabaceae gametophytic (G)SI being *RNase* based.

We identified *SSK1-like* genes (present only in species presenting S-*RNase* GSI), in the *Trifolium pratense* (SI), *Medicago truncatula* (SC)*, Cicer arietinum* (SC)*, Glycine max* (SC), and *Lupinus angustifolius* (SC) genomes. Therefore, we characterize the *S*-lineage *T2-RNase* genes in these genomes. We identified *S-RNase* lineage genes in *T. pratense*, *M. truncatula* and *C. arietinum,* that in phylogenetic analyses cluster with Pyrinae *S-RNases*. In *M. truncatula* and *C. arietinum* genomes, where large scaffolds are available, these sequences are surrounded by F-box genes that in phylogenetic analyses also cluster with *S*- pollen genes. These genes show, however expression and polymorphism incompatible with determining GSI. All *T. pratense S-RNase* like genes show low levels of polymorphism. To address if other *T2-RNases* could be determining Fabaceae GSI, here we obtained a style with stigma transcriptome of *Cytisus striatus*, a species that shows significant difference on the percentage of pollen growth in self and cross-pollinations. Expression and polymorphism analyses of the *C. striatus T2-RNase* like genes revealed that none of these genes, is the *S*-pistil gene. Thus, we find no evidence for Fabaceae GSI being determined *S-RNase* like genes.

273A

Convergent evolution enabled phytoplasmas to generate 'zombie plants'

Florian Ruempler [1], Lydia Gramzow[1], Guenter Theissen[1], Rainer Melzer[2]
[1] Friedrich Schiller University Jena, Department of Genetics, Jena, Germany, [2] University College Dublin, School of Biology and Environmental Science, Dublin, Ireland

Phytoplasmas are pathogenic bacteria that are obligate parasites of plants and transmitting insects. Remarkably, they reprogram plant development such that leaf-like structures instead of floral organs develop. Due to the induced alterations in floral structure, infected plants are often sterile; they mainly serve to propagate and reproduce phytoplasmas and have thus been termed ‘zombie plants'. The molecular mechanism underlying the developmental reprogramming relies on specific interactions of a secreted phytoplasma protein called SAP54 with a relatively small subset of MIKC-type MADS-domain transcription factors involved in flower development. The secreted part of SAP54 interacts with the keratin-like domain (K-domain) of MADS-domain transcription factors and destines them for degradation. However, it remains largely unclear how this intricate mechanisms of manipulating host development originated. Since a better knowledge about the origin and evolution of SAP54 is key to understand how phytoplasmas can induce the development of ‘zombie plants' we analyzed structural characteristics of SAP54 using in silico methods. We found that the protein most likely forms a structure which is strikingly similar to the recently identified X-ray crystal structure of the K-domain of SEPALLATA3, one of the targets of SAP54. We propose, therefore, that the interaction between SAP54 and the plant MADS-domain proteins is mediated by a mechanism that is similar to that of the interaction of two K-domains. Furthermore, we provide evidence that SAP54 underwent convergent structural and sequence evolution to mimic the K-domain and thus to establish the specific interactions that eventually lead to the degradation of MADS-domain proteins.

274B

Gene duplicability of core angiosperm genes is highly consistent across 37 species

Riet De Smet [1], Zhen Li[1], Jonas Defoort[1], Setareh Tasdighian[1], Kevin Vanneste[3], Steven Maere[1], Yves Van de Peer[1,2]

[1] VIB/Ghent University, Zwijnaarde, Belgium, [2] University of Pretoria, Pretoria, South Africa, [3] WIV-ISP, Brussels, Belgium

Gene duplication is an important mechanism for the production of genomic novelties. Hence, which genes tend to undergo duplication and do get fixed following duplication is an important question. It has been observed that gene duplicability or the ability of genes to get fixed following duplication can be a non-random process, with certain genes being more amenable to survive duplication events than others. Primarily, gene essentiality and the type of duplication (small-scale versus genome duplication) have been shown in different species to influence gene duplicability. However, an overall picture of ‘gene duplicability' is lacking, mainly due to the fact that studies investigating gene duplicability often focus on individual species and that the influence of genomic context and the timing of duplication events remains unaccounted for. Therefore, we undertake here a large-scale study in which we investigate duplicate retention in 9,178 core gene families that are shared between 37 flowering plant species. The flowering plants or angiosperms are an ideal model system to study the consequences of gene duplication, since both whole-genome duplications as small-scale duplications are common within the angiosperms. We observe a striking consistent pattern of gene duplicability for all 9178 gene families across the 37 species, with gene families being either primarily single-copy or retained in duplicate in all species. Potential ‘dosage-balance' sensitive genes form an intermediate state, since they are often retained in duplicate for periods extending to 50 million years after whole-genome duplication but get lost from the genome eventually.

275C

Adaptive molecular convergence at the level of transcripts, but not individual residues, in protein sequences in Anolis.

Russ Corbett-Detig [1], Shelbi Russell[2], Daniel Hartl[2], Rasmus Nielsen[1], Daniel Hartl[2]
[1] UC Berkeley, Berkeley, CA, USA, [2] Harvard, Cambridge, MA, USA

Convergent evolution of ecologically favorable phenotypes is perhaps the strongest available evidence of adaptive evolution. However, it is unknown whether similar changes at the genetic level are required to produce similar phenotypic outcomes. While, many authors have reported compelling examples of identical—or similar—amino acid replacement mutations that are associated with convergent phenotypic evolution, genome-wide surveys for molecular convergence have yielded mixed results. Hence, it remains unknown how common molecular convergence is across a broad set of phenotypes and organisms. The adaptive radiation of Caribbean anoles is one of the best-studied examples of phenotypic convergence. This radiation is characterized by numerous independent origins of suites of adaptive morphological and behavioral traits—collectively referred to as 'ecomorphs'. Anoles therefore offer a unique opportunity to investigate the genetic underpinnings of phenotypic convergence. To address the frequency of molecular convergence in anoles, we sequenced the genomes of 12 species, four representatives of each of three ecomorphs. In scanning 12,358 genes, we find no evidence supporting convergent evolution at amino acid residues within ecomorphs. However, we do find that many genes are evolving at accelerated rates within a set of ecomorph species, which indicates that a molecular convergence is occurring on a broader scale. Gene ontology analyses show that many of these ecomorph-accelerated genes are functionally related, and cell signaling, in particular, shows evidence of accelerated evolution. We conclude that adaptive molecular convergence at the level of individual mutations is uncommon, convergence in rates of evolution across genes is comparatively more common.

276D

Gentree: a community resource of new gene study

Yi Shao [1], Chunyan Chen[1], Hao Shen[2], Manyuan Long[3], Yong E. Zhang[1]
[1] Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China, [2] Department of Computer Science, Hunan University of Technology, Zhuzhou 412008, China, [3] Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637, USA

Gentree (http://gentree.ioz.ac.cn) is designed as a community resource to facilitate to study evolutionarily new genes in human. For each Ensembl annotated gene, Gentree shows the timing when it gets originated, the mechanism to create this locus and various annotation information. First, Gentree consists of self computed age information based on vertebrate whole genome alignment. As a complement, Phylostratigraphy and ProteinHistorian age data based on protein alignment is also integrated. The web interface enables users to browse when different member from the protein family or biological process of interest gets originated. Second, the origination mechanism is roughly classified as DNA-level duplication, RNA-level duplication and de novo emergence, respectively. Users can view how parental and children locus is diverged regarding both the gene structure and the transcription. Finally, the Ensembl annotation tends to be labile for recently evolved young genes. By integrating numerous proteomics datasets, Gentree showed that at least 64 pseudogenes could be candidate primate-specific protein-coding genes. Furthermore, since lineage- or species-specific young genes are often poorly characterized, Gentree integrated numerous transcriptome data such as BrainSpan or BodyMap 2.0 and genome-wide association data to give a hint on the potential functionality of these genes. The current version of Gentree focuses on human, but the subsequent versions will cover more species including mouse and fruit fly.

277A

THE PANCREATIC LIPASE-COLIPASE COMPLEX: AN UNEXPECTED OLD DUO

Monica Lopes-Marques [1,2], Miguel M. Santos[1,3], L. Filipe C. Castro[1]

[1] CIIMAR-Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal, [2] ICBAS-Institute for the Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal, [3] Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal

Colipase (CPLS) is a small protein cofactor essential for the activation of pancreatic lipase (PL) an enzyme involved in hydrolysis of dietary triglycerides. CPLS gene orthologues have been described exclusively in mammals and birds, in contrast to PL which has a wider phylogenetic distribution. Here we investigate the evolutionary history of the CPLS gene family in metazoans. Our findings indicate that CPLS-like genes first emerged in chordate ancestry. In agreement, we find clear orthologues in the invertebrate chordate species, *B. floridae* and *C. intestinalis*. Similarly, agnathans and chondrichthyans have clear single copy CPLS genes. In contrast, the majority of teleost species examined here lack a CPLS gene. Sequence and structural analysis using comparative homology modeling demonstrated that all CPLSs recovered exhibit a true CPLS profile with conserved cysteine residues involved in the formation of the disulphide bridges and the three finger structural topology. Protein-protein docking analysis demonstrates that the lamprey and spotted gar PL-CPLS complexes show a very similar binding pattern when compared to the crystal structure of *H. sapiens*. Overall our data indicates that this enzymatic system is older than expected and represents a chordate innovation possibly linked with the emergence of the pancreas.

278B

Expanded homolog space in vertebrate genomes: insights from cyclostomes and cartilaginous fishes

Shigehiro Kuraku
RIKEN CLST, Kobe, Japan

Ten years ago, our knowledge of vertebrate genes was confined to human and several laboratory animals for which large-scale sequence resources were available. Now, we can access whole genome sequences of quite a few ‘non-model' vertebrates. Those emerging resource have shed light on additional members of well-studied gene families which were not identified in the genomes of traditional ‘model' vertebrates (‘cryptic pan-vertebrate genes'). The expanded homolog space can further be confirmed with genome-wide sequences of cyclostomes and cartilaginous fishes that split more anciently from the lineage leading to bony vertebrates which most of the vertebrate species with sequenced genomes belong to. This presentation will cover current status of sequence data production for those animals, some examples of revised gene phylogeny, and alleged factors that can mislead phylogeny inference.

279C

Convergent sequence evolution in thick-lipped cichlid fish

Peiwen Xiong, Paolo Franchini, Axel Meyer
Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Konstanz, Germany

Convergent evolution of phenotypic traits supports the hypothesis that selection led to adaptive evolution. The molecular basis of convergence is rarely studied and barely understood so far. In cichlid fish, a notable convergent trait, the thick-lipped phenotype, has evolved repeatedly in independent lineages both in African and the Neotropics. Based on phylogenetic approaches on transcriptomic data, we aimed to detect the homologous genes might account for convergent evolution in Central American and African thick-lipped cichlid fish (*Amphilophus labiatus* and *Lobochilotes labiatus*, respectively). We identified 2,268 single-copy orthologous genes by comparing both thick- and thin-lipped cichlids as well as other nine teleost species. A Bayesian phylogenetic tree was constructed based on nine genes with 8,994 bp. By performing branch comparisons between all pairs of lineages on the orthologous genes, a set of candidate genes was identified, in which the convergent substitutions accumulated in thick-lipped species after they diverged from their stem lineages are more than expected. These genes might have contributed to the convergent evolution of this phenotype, and provide us with new insights into the genetic basis and mechanism of convergent molecular evolution.

280D

Genetic diversity and population structure of the endangered oceanic whitetip shark *Carcharhinus longimanus* in the Atlantic Ocean.

Sâmia Camargo [1,2], Fernando Mendonça[1,2], Rui Coelho[3], Miguel Santos[3], Fausto Foresti[2]
[1] Instituto do Mar, Universidade Federal de São Paulo, UNIFESP, Santos, São Paulo, Brazil, [2] Instituto de Biociências de Botucatu, Universidade Estadual Paulista, UNESP, Botucatu, São Paulo, Brazil, [3] Instituto Português do Mar e da Atmosfera, IPMA, Olhão, Algarve, Portugal

Among the large species of sharks, the oceanic whitetip shark *Carcharhinus longimanus* has high levels of population depletion, and is currently listed as "Vulnerable" on the IUCN Red List. Population genetics is a powerful tool that provides important information regarding to population structure and genetic connectivity contributing to the establishment of good strategies for conservation of threatened species. Considering the limited information about population levels of the whitetip shark, we used partial sequences of the mtDNA control region to determine its population genetic structure across the Atlantic Ocean. We sampled 215 specimens of *C. longimanus* across wide areas of both the eastern and western Atlantic, and obtained 724 bp identifying nine polymorphic sites that resulted in 12 distinct haplotypes. The total nucleotide diversity was $\pi = 0.0013$ and haplotype diversity $h = 0.5953$. These results show a genetic variability slightly below the observed average among other species of pelagic sharks. The Analysis of Molecular Variance (AMOVA) evidenced moderate levels of population structure ($FST=0.1039$, $P<0.001$) with restriction of gene flow between the Western and Eastern Atlantic Ocean. Although such divergence index does not represent distinct genetic stocks, we suggest that priority actions should be directed to specific areas that contain these unique haplotypes leading to the maintenance of the genetic diversity throughout the Atlantic Ocean. Financial support: BIOTA FAPESP 2011/23787-0 and FCT SFRH/BPD/93936/2013

281A

lncRNA-RNA interactions in gene expression regulation

Michał Szcześniak, Oleksii Bryzghalov, Izabela Makałowska
Adam Mickiewicz University in Poznan, Poznan, Poland

Long non-coding RNAs (lncRNAs) are non-protein coding transcripts longer than 200 nucleotides. Although there is possibility that a fraction of lncRNAs are not functional and represent mere transcriptional noise, a growing body of evidence shows they are engaged in a plethora of molecular functions and have a considerable contribution to the observed diversification of eukaryotic transcriptomes and proteomes. Still, however, only ca. 1% of lncRNAs have well established functions and much remains to be done towards decipherment of their biological roles. One of the least studied aspect of lncRNAs biology is their engagement in gene expression regulation through RNA-RNA interactions. By hybridizing with mate RNA molecules, lncRNAs could potentially participate in modulation of pre-mRNA splicing, RNA editing, mRNA stability control, translation activation, or abrogation of miRNA-induced repression. Here, we present a similarity-search based method for identification of RNA-RNA interactions transcriptome-wide, which enabled us to find 18,871,097 lncRNA-RNA base-pairings across the human transcriptome. Further analyses show that the interactions could affect processing, stability and functions of 57,575 transcripts. An extensive use of RNA-Seq data provided support for approximately one third of the interactions, at least in terms of the two RNA components being co-expressed. We also created an online database to store the RNA-RNA interaction data. Altogether, our results suggest that lncRNA-RNA interactions are broadly used to regulate and diversify the transcriptome.

282B

Retrogenes in animal genomes

Michal Kabza, Joanna Ciomborowska, Izabela Makalowska
Department of Bioinformatics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland

Retrogenes are copies of existing genes that arise from insertion of reverse transcribed mRNAs into the genome. Because of the lack of regulatory elements, which are not inherited, such copies used to be considered as non-functional (retropseudogenes) and classified as "junk DNA". Multiple discoveries, most notably in the past 10 years, have challenged that view. Since then, multiple mechanism by which retrogenes may become functional have been proposed. A unique feature of retrogenes - the loss of all or most of the cis-regulatory elements, leads to the lower level of evolutionary constraint, which is the reason why retrogenes may relatively easily undergo neofunctionalization. As a result, retroposition is a vital part of the process of development of lineage- and species-specific traits.

Although multiple attempts have been made to detect retrogenes in the genomes of model organisms, there is still no repository of retrogenes for a broader range of organisms. Here, we present a new database called RetrogeneDB (http://retrogenedb.amu.edu.pl/) that contains high-quality retrogene datasets for 62 genomes from Ensembl release 73. RetrogeneDB allows users to search for retrogenes and their parental genes using numerous criteria. For selected organisms retrogene expression estimation from RNA-Seq data is provided.

Acknowledgments: This work was supported by Polish National Science Center under project 2013/09/N/NZ2/01221

283C

Evolution of chimeric lncRNAs in the human lineage

Carla Bello Cabrera, Fyodor Kondrashov
Centre for Genomic Regulation (CRG), Barcelona, Spain

The evolution of new phenotypic traits are fundamental for the diversity and survival of the species. It is well known that new genes have contributed to the origin of adaptive evolutionary novelties and that the primary source for such novelties is the duplication of genomic sequences. The increasing amount of information regarding long non-coding RNAs (lncRNAs) has raised the question of whether these RNA genes could play important roles in the evolution of the human lineage. Here, we evaluate the generation of new chimeric lncRNA genes that have originated though exon duplication in the human lineage. Utilizing the data generated by the ENCODE project, we show that exon duplication is a common event among lncRNAs and that alternative spliced lncRNA genes undergo significantly more exon duplication events than those that are not. The latter suggests that alternative splicing could be a driver of exon duplications that increases lncRNA diversity. By comparing primate genomes using whole-genome sequencing (WGS) data we have identified novel human specific chimeric lncRNA genes that were recently formed through exon duplication. Moreover, RNAseq expression analysis using publicly available data shows that some of these novel chimeric genes are expressed in the brain, suggesting a role in human brain expansion. Notably, single-nucleotide polymorphism (SNP) analysis utilizing data from the 1000 Genomes Project shows that these lncRNAs are evolutionary constrained and thus under negative selection, indicating that they are in fact functional. Overall, these results contribute to unveil the evolutionary pathway of the human lineage and other apes.

284D

DDT resistance in Drosophila: beyond the mono/polygenic debate.

Joshua Schmidt, Rebecca Smith, Robert Good, Charles Robin
The University of Melbourne, Melbourne, Victoria, Australia

Insecticides impose such a strong selective regime on insect populations that the initial survivors are often conceived of as ‘hopeful monsters’ that carry a rare allele at a single gene that has a major effect. However if such alleles have fitness effects that need to be ameliorated by modifier loci, or if the selection intensity is variable across the species range so that minor effect genes are adequate, or if abundant initial standing variation allows multiple solutions, then over time, the adaptive response to insecticides is likely to involve many genes. Thus If we are to understand this classic model of microevolution at a reductionist level we need to know the number of alleles involved in the selection response, their effect sizes, the physiological pathways involved and how they interact and then assess their adaptive significance. James Crow among others mapped multiple DDT associated variants in Drosophila melanogaster and yet only one gene has previously shown to be of relevance to variance in field populations. Here we use a genome wide association study in Drosophila melanogaster to molecularly identify DDT-associated polygenes and use selective sweep analyses to assess their adaptive significance. We observe that the genetic architecture involves genes from diverse gene ontologies. This includes genes that function in neuronal development and function, one of which we characterize for the first time. Yet we only find compelling evidence of selection at a putative detoxifying gene.

285A

A Database for De Novo Genes

Scott Foy, Joanna Masel
University of Arizona; Dept of Ecology & Evolutionary Biology, Tucson, Arizona, USA

The evolution of noncoding DNA sequences to de novo genes that express proteins is of growing interest in both evolutionary and molecular biology. As the field continues to grow and more data is generated, it will be useful to consolidate de novo gene information into a central database, where it can be stored and easily queried. We are developing such a database containing de novo gene, taxon, transcript, and protein information, along with citation and strength of evidence summaries. Both the strength of evidence that a sequence codes for a protein, and the strength of evidence for de novo birth in a particular time interval, varies greatly among studies. Furthermore, all data will be embedding in a phylogenetic context.

The database will eventually store data for both pure de novo genes born from intergenic sequences, and overlapping genes born from alternative reading frames. Additionally, the database will store genes of very different ages. The inclusion of these will facilitate the study of changes in gene and protein properties (e.g., intrinsic structural disorder or aggregation propensity).

286B

Fitness-valley crossing in subdivided asexual populations

Michael McLaren
Stanford University, Stanford, USA

Adaptations involving sets of interacting mutations that would be deleterious or neutral on their own but are beneficial together may be crucial for the evolution of new biological functions. Acquiring these adaptations requires the population to cross a fitness valley or plateau; such valley crossing may also be a crucial step of cancer development in mammals. Spatial population structure is ubiquitous in nature and in mammalian somatic tissues, and has been shown in some cases to speed the crossing of fitness valleys. Here we present a detailed and general analysis of the effects of spatial structure on valley crossing in asexually reproducing populations. We consider a two-step valley in a population subdivided into discrete subpopulations, or demes. We first use qualitative analysis under a general model to show how subdivision can facilitate valley crossing by increasing the time that an intermediate genotype drifts in the population before going extinct. We then give a complete quantitative analysis of valley crossing when the population is structured as a finite island model with a large number of demes. Our results show that valley crossing can be orders of magnitude faster in a heavily subdivided population than in an unstructured one. However, whether subdivision is beneficial depends on all of the model parameters in a complex way. We also solve for the probability distribution of the valley-crossing time and find that even when subdivision reduces the average valley-crossing time, the probability of crossing the valley very quickly is independent of the level of subdivision.

287C

The opsin repertoire of the jumping bristletail *Machilis hrabei* (Hexapoda: Archaeognatha).

Alexander Böhm, Günther Pass
Department of Integrative Zoology, University of Vienna, Vienna, Austria

Despite the great interest from a phylogenetic point of view no data has been published until now on opsins in primary wingless insects. While some of these ancestral groups are blind the Archaeognatha have highly developed compound eyes and ocelli. In the investigated *Machilis hrabei* two of the ocelli are sole shaped and situated below the compound eyes, while the third one curiously is directed downward at an 90 degree angle.

In ESTs and preliminary genome sequence data of this species we found six opsins. Five of them belong to the rhabdomeric and one to the ciliary class of opsins. The only ciliary-type opsin is an Opn3 homologue. Opn3 opsins became only recently known to be expressed in the brain and various other tissues of insects and vertebrates. The putative spectral sensitivities of the other five opsins found in *Machilis hrabei* are as follows: one UV sensitive, one blue, three long wave (green) sensitive opsins. One of the latter is the only hexapod opsin so far that is placed in a certain cluster of crustacean and chelicerate long wave sensitive opsins.

Furthermore, we examine the expression pattern of the mentioned opsins by in situ hybridization. This should lead to new insights on the evolution of opsins and vision in hexapods and set a basis for future physiological investigations.

288D

Evolutionary innovation in *Escherichia coli* populations

Kathleen Sprouffske [1,2], Paul Sniegowski[3], Andreas Wagner[1,2]

[1] University of Zurich, Institute for Evolutionary Biology and Environmental Sciences, Zurich, Switzerland, [2] The Swiss Institute for Bioinformatics, Lausanne, Switzerland, [3] University of Pennsylvania, Department of Biology, Philadelphia, PA, USA

Understanding how evolutionary innovations arise and spread is key to understanding evolution. Many approaches to studying evolution can only infer the course and dynamics of evolution using existing species or the fossil record. Experimental evolution, on the other hand, allows us to directly observe evolution in action. Here, we use experimental evolution in conjunction with phenotypic analyses to understand the dynamics of innovation in evolving populations of bacteria with increasing mutation rates.

We evolved four *Escherichia coli* strains with different mutation rates (8 replicates each) for 3000 generations in minimal media at 37˚C. We characterized innovation in these populations in over 100 stressful, novel conditions (conditions that were never experienced by the strains during the experiment). We found innovation to be common across all strains, and identified strain-specific patterns of innovation.

289A

Protein trafficking in photosynthetic organelles of the armoured amoeba *Paulinella chromatophora*

Przemyslaw Gagat, Andrzej Bodyl, Pawel Mackiewicz
University of Wroclaw, Wroclaw, Poland

Paulinella chromatopora is an armoured filose amoeba of the supergroup Rhizaria that inhabits sediments of freshwater bodies. It harbours two photosynthetically active endosymbionts of cyanobacterial origin (chromatophores), acquired independently of primary plastids of the supergroup Archaeplastida (photosynthetic organelles of glaucophytes, red algae and green plants). Similarly to primary plastids, chromatophores have lost many essential genes, and transferred substantial number of genes to the host nuclear genome via endosymbiotic gene transfer (EGT), including those involved in photosynthesis. This indicates that, like primary plastids, *Paulinella* endosymbionts must have evolved a transport system to import their EGT-derived proteins. We elaborated a model for protein import into *Paulinella* chromatophores and their thylakoid membranes, based on analyses of the chromatophore genomes, *Paulinella* EST database and presequences of proteins imported to the chromatophores. For comparative studies, we analysed genomes of primary plastids, all sequenced cyanobacteria and some bacteria. Our model involves (i) vesicular trafficking to the outer chromatophore membrane, (ii) a simplified Tic-like complex at the inner chromatophore membrane resembling primary plastid Tic translocon, (iii) a molecular motor responsible for pulling imported proteins into the chromatophore stroma and (iv) thylakoid trafficking based on Sec, Tat and Srp pathways, which are also conserved in primary plastids. Part of the model concerning the vesicular trafficking through the outer chromatophore membrane has already been proved experimentally. Since *Paulinella* chromatophores indeed evolved a protein import system, the endosymbionts should be acknowledged as true call organelles, among primary plastids and mitochondria.

290B

The fate of young and old duplicates from 14q32.1 and 18q21.3 *SERPIN* gene clusters

Susana Seixas
Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal

The superfamily of serine protease inhibitors (*SERPIN*s) has two extended clusters in the human genome originated by duplications occurred at different timings of vertebrate evolution. The 14q32.1 cluster includes 11 members, all sharing a similar gene structure to alpha-1-antitrypsin (SERPINA1) and the 18q21.3 cluster comprises 10 members, characterized by their homology to chicken ovalbumin. Although the majority of these genes are widespread across mammalian species, some are restrained to certain phylogenetic groups, making the repertoire of each species unique. In primates, events of gene duplication and divergence are associated to the origin of novel functions (*SERPINA2* and *SERPINB3*), while other genomic processes led, in some specific lineages, to the pseudogenization of ancient and/or young duplicates (*SERPINA13, SERPINB11 and SERPINA2*). Specifically, *SERPINA2* active isofrom, which has not yet been fully characterized, resembles the misfolding variants of it close homolog SERPINA1 without their pathological consequences. SERPINB3 has accumulated at the reactive site, several amino acid replacements, currently exerting a distinct activity from SERPINB4 against both endogenous and exogenous proteases. Conversely, *SERPINA13* and *SERPINB11* were equally lost early in human evolution, but while *SERPINA13* appears to have been kept in a few primates, *SERPINB11* is conserved in many species and in present day humans it has resurrected as a novel non-inhibitory SERPIN with an unknown function. Overall, gains and losses of genes have had an important value in the long-term evolution of 14q32.1 and 18q21.3 clusters with probable implications in natural history of the human species and in the susceptibility to disease.

291C

Of NUMTs and NUPTs – *Toxoplasma gondii* has the highest content ever observed with >10,000 insertions in a 63Mb genome

Cheng Sun[2], Sivaranjani Namasivayam[1], Assiatu Barrie[2], Wenyuan Xiao[1], Erica Hall[1], Jenna Oberstaller[1], Cedric Feschotte[2], Jessica Kissinger [1], Ellen Pritham[2]
[1] *University of Georgia, Athens, GA, USA,* [2] *University of Texas - Arlington, Arlington, TX, USA*

NUMTs account for 1.6% of the *T. gondii* ME49 nuclear genome, the highest fraction ever reported for any eukaryote. The zoonotic parasite *Toxoplasma gondii* contains three genome sequences: one nuclear (63Mb); one plastid organellar, ptDNA (35kb); and one mitochondrial organellar, mtDNA, whose sequence has been elusive. Traditional mtDNA isolation methods have been hampered by a highly-polymorphic mitochondrion, an elusive genome architecture and nuclear insertions of small mtDNA fragments, called (NUMTs). We employed techniques that allowed the identification of 23 sequence elements that constitute the mtDNA genome of *Toxoplasma*. Homology-based analyses of NUMT and NUPT (ptDNA insertions) revealed that insertion/deletion is a frequent and active process generating polymorphisms between strains. Our analyses indicate that within apicomplexans, NUM/PTs are found throughout the Coccidia. Comparisons with *Neospora caninum* (28 MY divergence) revealed that while most insertions are unique to the genus, a few insertions predate the divergence of these genera. This conservation of sequences is consistent with evolutionary constraint and is suggestive of cellular function. Because most NUMT insertions reside within (67%) or nearby genes (within the 1 Kb flanking region (18%)), we hypothesize that they might impact gene expression. Together these observations demonstrate a role for organellar sequences in shaping the genetic architecture of this important zoonotic pathogen and contributing to strain-specific differences that may be involved in adaptation, virulence and speciation.

292D

Are Insertions and Deletions Key Players in Protein Evolution?

Stephane Emond [1], Nobuhiko Tokuriki[2], Florian Hollfelder[1]
[1] *University of Cambridge, Cambridge, UK,* [2] *University of British Columbia, Vancouver, Canada*

In Nature, proteins evolve and acquire new functions by accumulating mutations. Substitutions and InDels (Insertions and Deletions), as well as circular permutations and rearrangement of protein domains, account for the majority of evolutionary changes. While the effects of substitutions have been extensively studied and documented, understanding the structural and functional effects of InDels still remains a challenge. InDels are assumed to be highly deleterious mutations because they are more likely to disrupt the structural integrity of proteins than are substitutions. On the other hand, they may induce significant structural changes that substitutions alone cannot cause and thus are believed to be key players in many natural evolutionary processes, such as the modification of active site loops to generate new enzyme functions[1] or the emergence of new protein structures[2]. We aimed at performing experimental protein evolution by randomly incorporating InDels to investigate how they would be tolerated and whether they could be selected for functional improvements. Starting from a previously reported methodology based on transposition mutagenesis[3], we developed new tools to generate libraries of single InDel variants. We screened the resulting libraries to identify adaptive InDels (i.e., conferring new functions) using different proteins as starting points. Our results show that, while being generally more deleterious than substitutions, InDels can also lead to functional improvements and may allow access to alternative evolutionary trajectories.

References

1 Park et al. (2006). Science 311, 535-538.
2 Grishin (2001). J Struct Biol 134:167-185.
3 Jones (2005). Nucleic Acids Res 33(9):e80.

293A

Genome-wide adaptation to an extreme desert environment in the rock pocket mouse, *Chaetodipus intermedius*

Noelle K. J. Bittner [1], Megan Phifer-Rixey[1], Matthew Bomhoff[2], Michael W. Nachman[1]
[1] University of California Berkeley, Berkeley, California, USA, [2] University of Arizona, Tucson, AZ, USA

Uncovering the genetic basis of adaptation to different environments is a fundamental goal of evolutionary genetics. The rock pocket mouse, *Chaetodipus intermedius*, has a number of physiological and anatomical adaptations that allow it to live without free drinking water in an extremely arid environment on a granivorous diet. Using ALLPATHS-LG we have assembled an Illumina sequenced *de novo* genome to gain insight into the genetic basis of adaptation to desert environments. We have identified lineage-specific gene family expansions and contractions, pseudogenes, and genes under selection. Comparisons with assembled genomes from other distantly related desert-adapted species provides insight into the genetic basis of adaptation to a desert environments.

L 7C

Evolution of functional and regulatory novelty at the self-incompatibility locus in Arabidopsis

Vincent Castric [1]

[1] *CNRS-University of Lille, UC Berkeley*

Self-incompatibility is a genetic system preventing selfing and enforcing outcrossing in hermaphroditic flowering plants. In the Brassicaceae, the primary phenotype (specificity of the recognition proteins) is encoded by a highly diversified set of molecular lock-and-key combinations, while the inheritance system ultimately determining the phenotype in heterozygote genotypes (dominance/recessivity interactions) is controlled by a dedicated set of non-coding RNAs and their target sites in the same genomic location. The high diversity of the molecular lock-and-key and its complex regulatory machinery involve repeated emergence of functional novelty for both of these aspects, hence providing the opportunity to study recently evolved molecular interactions in a context where the co-evolutionary constraint is an integral part of the mechanism by which the system functions. A major asset is that both the genomic and ecological contexts in which these genetic systems evolve have been explicitly clarified, providing the unique opportunity to investigate the co-evolutionary process down from nucleotide variation all the way up to their fitness consequences. By building upon these two case studies, I will detail several theoretical and functional challenges that arise from the study of this fascinating diversification process.

8 The horizontal component of microbial evolution

8.1

CRISPR-cas systems as a barrier to lateral gene transfer in archaea - comparative genomics and experimental genetics

Israela Turgeman[1], Adit Naor[1], Yuri Wolf[2], David Christensen[2], Eugene Koonin[2], Uri Gophna [1]
[1] Tel Aviv University, Tel Aviv, Israel, [2] National Library of Medicine, NIH, Bethesda, Md, USA

CRISPR-Cas systems provide acquired heritable immunity to bacteria and archaea against invasion by selfish DNA elements. As such, it has been speculated that it may prevent or at least impede horizontal gene transfer (HGT) in specific lineages. Using independent measures of HGT, and the spacer number in each CRISPR-Cas containing genome we explored the dependence between the the activity of the system, and the estimated number of recent gene acquisition events in hundreds of microbial genomes. Surprisingly, CRISPR-Cas activity was positively correlated with lateral gene acquisition in bacteria, while in archaea, where these systems are nearly ubiquitous, the apparent negative association could be attributed to the effects of growth temperature. Data from halophilic archaea indicate that many spacers in archaea target chromosomal replication genes, implying a potential role for CRISPR-Cas in interspecies antagonism and genetic conflicts in archaea. Using experimental genetics we show that indeed during cell fusion experiments, one partner species degrades part of the other's genome and that the activity of CRISPR-Cas can determine the genetic and ecologic consequences of inter-species mating and HGT.

8.2

Parallel evolution of a global regulator ameliorates the cost of plasmid carriage

Ellie Harrison [1], Steve Paterson[2], Andrew Spiers[3], Michael Brockhurst[1]
[1] University of York, York, UK, [2] University of Liverpool, Liverpool, UK, [3] Abertay University, Dundee, UK

Conjugative plasmids play a major role in horizontal gene transfer yet their existence represents a paradox; though many carry useful bacterial genes, all plasmids exert a biosynthetic burden on their host. Therefore selection should lead to the loss of the plasmid backbone and the retention of any useful genes on the bacterial chromosome. Numerous studies have shown that evolution can resolve this paradox by ameliorating the cost of plasmid carriage, yet the mechanisms for this are poorly understood. In this study we combine experimental evolution, gene expression analysis and whole genome sequencing to investigate how evolution resolves this paradox.

We propagated replicate populations of *Pseudomonas fluorescens* carrying mercury resistance plasmid pQBR103 by serial transfer for 450 generations. Gene expression analysis revealed that prior to evolution plasmid carriage was associated with widespread up-regulation of genes involved in protein production, suggesting that pQBR103 exerts a translational burden on its host. Following evolution however these transcriptional changes where reversed. Sequencing of evolved clones revealed widespread parallel evolution targeting two bacterial loci; gacA and gacS, a two-component regulator of post-transcriptional modification that positively regulates the production of numerous secondary metabolites. Knockouts of these genes confirm that loss of gacA/S function results in amelioration of the cost of plasmid carriage. As loss of gacA/S results in increased post-transcriptional repression of bacterial genes, these data suggest that bacteria are able to ameliorate of the cost of plasmid carriage by reducing their own translational workload to accommodate that of the plasmid.

8.3

Interactions between horizontally acquired genes create a fitness cost in Pseudomonas aeruginosa

Alvaro San Millan, Macarena Toll-Riera, Qin Qi, R Craig MacLean
University of Oxford, Oxford, UK

Horizontal gene transfer (HGT) plays a central role in the evolution of microbes. Despite the benefits associated with HGT, newly acquired genetic material can also generate fitness costs that are likely to impose an important constraint on evolution by HGT. In this study, we use the interaction between *Pseudomonas aeruginosa* PAO1 and a costly small plasmid (pNUK73) as a model system to investigate the molecular basis of the cost of HGT. Using RNA-Seq, we show that the acquisition of pNUK73 results in a profound alteration of the transcriptional profile of the PAO1 chromosome, including the activation of the SOS response. Compensatory mutations in either a chromosomal helicase and kinase eliminate the fitness cost of plasmid carriage and reverse the changes in chromosomal gene expression associated with pNUK73 acquisition. Compensatory mutations also decrease the expression of the plasmid replication gene *rep*, and we experimentally demonstrate that high levels of *rep* expression drive the cost associated with pNUK73 carriage and the activation of the SOS response. Both genes involved in compensatory adaptation show no clear biological role in PAO1 and multiple signatures of recent horizontal acquisition in the PAO1 genome, implying that the cost of pNUK73 stems from intragenomic conflict between recently horizontally acquired genes. Our study provides new insights into the evolutionary and molecular mechanisms responsible for the cost of HGT and we argue that interactions between mobile genetic elements may play an important role in the evolution of bacteria via HGT.

8.4

Whole genome sequencing of non-vaccine pneumococcal serotypes from invasive disease reveals widespread but highly variable recombination

Cheryl Andam [1], Patrick Mitchell[1], Alanna Callendrello[1], Lesley McGee[2], Lindsay Kim[2], Bernard Beall[2], Marc Lipsitch[1], William Hanage[1]

[1] Harvard School of Public Health, Boston, Massachusetts, USA, [2] Centers for Disease Control and Prevention, Atlanta, Georgia, USA

A common resident of the upper respiratory tract, the bacterium *Streptococcus pneumoniae* (pneumococcus) is a major cause of pneumonia, bacteremia and meningitis, accounting for high morbidity and mortality worldwide. Recombination and selection play an important role in driving the population dynamics and evolution of different pneumococcal lineages, allowing them to successfully evade the impacts of selective pressures such as vaccination and antibiotic treatment. However, it remains unclear how vaccines, which target only a small subset of the 94 pneumococcal serotypes, affect the evolution of the non-vaccine serotypes that occupy the niche left behind by the vaccine serotypes. We completed whole-genome sequencing of 881 non-vaccine serotype pneumococcal isolates from invasive disease cases across the United States sampled before and after 13-valent conjugate vaccine introduction in 2010. Recombination rates and patterns are highly variable across the population with recombination to mutation parameter per genome ranging from <1-101 and 10-273 recombination events within a single sequence type. At least 11 novel genetic variants have emerged post-vaccine, many of which possess recombinant genomic regions. Evidence of at least four incidents of serotype switching, where genes coding for a specific capsule are altered or exchanged with other types through recombination, has been identified in the post vaccine sample. We conclude that selective pressures due to vaccination can alter recombination dynamics of pneumococcal lineages not specifically targeted by the vaccine.

8.5

Recombination drives the GC-content of bacterial genomes through biased gene conversion.

Florent lassalle[1], Séverine Périan[1], Thomas Bataillon[2], Xavier Nesme[1], Laurent Duret[1], Vincent Daubin [1]

[1] Université Lyon 1, Lyon, France, [2] Aarhus University, Århus, Denmark

Recent results have suggested that a mysterious selective force favouring higher GC-content exists in Bacteria. Here, we show that GC-Biased Gene Conversion (gBGC), a non adaptive process associated with homologous recombination has a great influence on genomic GC-content in most if not all bacterial species. First we find a consistent positive relationship between the GC-content of a gene and evidence of intra-genic recombination throughout a broad spectrum of bacterial clades. Second, we show that the evolutionary force responsible for this pattern is acting independently from selection on codon usage, and could potentially interfere with selection in favor of optimal AU-ending codons. A comparison with data from human populations shows that the intensity of gBGC in Bacteria is comparable to what has been reported in mammals, in which gBGC is considered a significant evolutionary force. We propose that gBGC is widespread among Bacteria and could therefore be an ancestral feature of cellular organisms. We argue that if gBGC occurs in bacteria, it can account for previously unexplained observations, such as the apparent non-equilibrium of base substitution patterns and, in association with lateral gene transfer, the heterogeneity of gene composition within bacterial genomes. Because gBGC produces patterns similar to positive selection, it is essential to take this process into account when studying the evolutionary forces at work in bacterial genomes.

8.6

Origins of major archaeal clades correspond to gene acquisitions from bacteria

Shijulal Nelson-Sathi [1], Filipa Sousa[1], Mayo Röttger[1], Nabor Lozada-Chávez[1], Thorsten Thiergart[1], Arnold Janssen[2], David Bryant[3], Giddy Landan[4], Peter Schönheit[5], Bettina Siebers[6], James McInerney[7], William Martin[1]

[1] Institute of Molecular Evolution, Düsseldorf, Germany, [2] Mathematisches Institut, Düsseldorf, Germany, [3] Department of Mathematics and Statistics, Dunedin, New Zealand, [4] Genomic Microbiology Group, Kiel, Germany, [5] Institut für Allgemeine Mikrobiologie, Kiel, Germany, [6] Faculty of Chemistry, Essen, Germany, [7] Department of Biology, Maynooth, Ireland

Prokaryotic genome evolution entails both tree-like components generated by vertical inheritance and network-like components generated by lateral gene transfer (LGT). The relative contribution of these two processes during the formation of prokaryotic species is the subject of intense debate. While it is clear that LGT within prokaryotic groups such as cyanobacteria, proteobacteria, or halophiles is an important factor for genome evolution, its role, if any, at the origin of such groups still remain as an open issue. To investigate the role of vertical and horizontal evolutionary processes underlying the origin of higher taxa in archaea, we have performed phylogenomic analysis of 134 archaea in the context of their homologues from 1,847 reference bacterial genomes. Our results show origins of archaeal higher taxa unexpectedly correspond to 2,264 group-specific gene acquisitions from bacteria. Interdomain gene transfer is highly asymmetric, transfers from bacteria to archaea are more than 5-fold more frequent than vice versa. These findings uncover a pivotal role for lateral gene transfer in major evolutionary transitions among prokaryotes and implicate bacterial gene acquisitions as key innovation en route to the origin of archaeal higher taxa.

8.7

Evolutionary assembly patterns of prokaryotic genomes

Maximilian Press, Christine Queitsch, Elhanan Borenstein
University of Washington, Seattle, WA, USA

Evolutionary innovation, by definition, occurs in the context of some genomic background, and this background limits available evolutionary paths. For example, protein evolution by sequence substitution is constrained by epistasis between residues. However, in prokaryotes, evolutionary innovation frequently happens by macrogenomic events such as the acquisition of whole genes via horizontal transfer. Previous work has suggested that the acquisition of genes can be similarly influenced by ancestral genomic content, yet the extent of such gene-level constraints has not yet been systematically characterized. Here, we set out to understand the structure and evolutionary impact of such constraints in prokaryotes, using probabilistic ancestral reconstructions from 634 extant prokaryotic genomes and a novel framework for detecting evolutionary constraints on gene-level acquisition events. We identified 10,594 directional dependencies between genes, and demonstrated that such dependencies influence the evolution of genes involved in important metabolic processes and pathogenesis. We further showed that these dependencies are often sufficient for predicting whether or not a given ancestral genome will acquire specific genes. Modeling these dependencies as a network, we also confirmed that detected dependences mirror many known gene-gene functional relationships. Lastly, we applied a graph-theory based approach to detect functional classes of genes that are gained in specific orders relative to one another. The obtained order indicated that early adaptations to physical environments and lifestyles have a strong effect on subsequent evolution in prokaryotes. More generally, our findings suggest that the emergence of specific metabolic and pathological phenotypes in prokaryotes may be predictable from current genomes.

8.8

Horizontal Transfer in E.coli plasmids.

Alice Ledda [2], Catherine Branger[1]

[1] MCU-PH Service de Microbiologie et Hygiène Hôpital Louis Mourier, Colombes, Paris, France,

[2] Imperial College, London, UK

Horizontal Transfer (HT) is responsible for the spread of antibiotic resistances in bacteria. These genes are often carried and spread by plasmids. Unfortunately, not much is known about plasmid evolution in natural environments. Understanding plasmid evolution may be a key factor in fighting the spread of antibiotic resistance.

We applied an innovative analysis method inspired from graph theory to a newly sequenced database of more than 100 deeply sequenced plasmids from a comprehensive set of E.coli spanning the whole history of antibiotic usage.

A common problem in studying plasmid phylogenies is the lack of a core genome. To overcome this difficulty, we developed a technique from graph theory and statistical physics to determine in a data driven way the most reliable set of genes to build plasmid phylogenies. We show that the strongest phylogenetic signal is contained in the Incompatibility Group (IG) genes.

We find that plasmid evolution is dominated by HT: plasmid phylogenies reconstructed from IG genes are unrelated to the phylogenetic history of their E.coli hosts. Plasmid accessory genes are randomly assembled from a pan-plasmidome pool: their presence and phylogeny are unrelated to the plasmid phylogeny reconstructed from IG. HT shapes plasmid evolution both at the level of their gene content and at the level of their spreading across E.coli.

To quantitatively explain our findings, we introduce a population genetics model of plasmid evolution describing the influence of HGT on their gene content. Our theory agrees with experiments and casts new light on other aspects of plasmid evolution and HGT.

8.9

Environmentally co-occurring mercury resistance plasmids have variable, context-dependent fitness effects and are phenotypically and genetically diverse

James Hall [1], Ellie Harrison[1], Andrew Lilley[4], Steve Paterson[2], Andrew Spiers[3], Michael Brockhurst[1]

[1] Department of Biology, University of York, York, UK, [2] Institute of Integrative Biology, University of Liverpool, Liverpool, UK, [3] The SIMBIOS Centre and the School of Science, Engineering and Technology, Abertay University, Dundee, UK, [4] Pharmaceutical Science Research Division, King's College London, London, UK

Conjugative plasmids are important agents of horizontal gene transfer, enabling genetic exchange and facilitating local adaptation of microbial communities. To understand the genetic composition of co-occurring plasmids and the effects of carriage on their hosts, we analysed four pQBR mercury resistance plasmids and measured their effects on a Pseudomonas fluorescens host which had also been isolated from the same site. The overall effects of carriage varied between plasmids and were strongly context dependent, with the relative fitness of plasmid-bearers changing with environmental mercury, growth medium, and whether the competitor also carried a plasmid. Conjugation rate, and rate of loss by segregation, also varied. Despite the large sizes of the plasmids, few accessory genes could be ascribed functions, although on two plasmids we identified a conserved chemotaxis operon, a type IV pilus-encoding cluster, and a region encoding putative arylsulfatase enzymes, which were also identified in geographically distant isolates. Transposons, particularly the Tn5042 mer transposon, appear to have been transferred horizontally between the plasmids relatively recently, indicating the importance of recombination in plasmid evolution. Together our findings show extensive genetic and phenotypic variation amongst environmentally co-occurring plasmids and suggest that environmental heterogeneity plays a role in the maintenance of plasmid diversity.

8.10

Molecular evolution of the pathogenicity and carbohydrate metabolism in mosquito-associated *Spiroplasma*

Wen-Sui Lo[1,2], Tean-Hsu Chang[1], Chuan Ku[1], Ling-Ling Chen[1], Chih-Horng Kuo [1,2]
[1] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan, [2] Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung Hsing University and Academia Sinica, Taipei, Taiwan, [3] Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan, [4] Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

Phylogenetic character mapping and comparative genomics are powerful tools to study the evolution of traits and their underlying genetic mechanisms. Using four divergent mosquito-associated *Spiroplasma* species as the study system, we conducted genome sequencing and transcriptomics experiments to investigate their evolution of pathogenicity and metabolic pathways. Our results indicated that the glycerol-3-phosphate oxidase gene (*glpO*), which is responsible for the production of reactive oxygen species, appeared to be the main virulence factor in the pathogenic *S. culicicola* and *S. taiwanense*. The independent losses of this gene in *S. diminutum* and *S. sabaudiense* provided an explanation of their lack of pathogenicity. Moreover, phylogenetic analyses revealed possible recombination of *glpO* among lineages belonging to different genera and local rearrangements of adjacent genes. In addition to glycerol metabolism, genes involved in other carbohydrate utilization also exhibited highly variable patterns of presence and absence among these four species. For gene expression, our transcriptome analysis indicated that the *glpO* in *S. taiwanense* is up-regulated upon glucose depletion in artificial medium, suggesting that its pathogenicity might be a stress response. Furthermore, the glucose depletion triggered opposite responses between the pathogenic *S. taiwanense* and the non-pathogenic *S. diminutum* regarding other carbohydrate metabolism pathways. Taken together, our results suggested that the genetic variations in metabolic pathways among these symbiotic bacteria could influence their interactions with the mosquito hosts.

8.11

Origin and spread of integron cassettes and how their fixation or loss shapes bacterial evolution

Jean Cury [1,2], Thomas Jové[3], Eduardo PC Rocha[1,2]
[1] Microbial Evolutionary Genomics, Institut Pasteur, 28, rue du Dr Roux, Paris, France, [2] CNRS, project EVOMOBILOME, UMR3525, 18, rue du Dr Roux, Paris, France, [3] Inserm, U1092, Université de Limoges, Limoges, France

Integrons are genetic platforms allowing gene acquisition by a specific recombination process. They also facilitate the emergence of diverse phenotypes by aggregating new genes and by providing new combination of them. Integrons are known to play a major role in the spread of antibiotic resistance genes, but their other roles in the evolution of bacteria are poorly understood. In particular, it is yet unknown how new cassettes are generated and how they can ultimately be fixed in genomes. We produced computational models for the integron integrase and for the recombination sites (attC). This allowed us to detect nearly 500 integrons and their gene repertoires with high sensitivity in 2643 bacterial genomes. The comparative analysis of the integrons clearly suggests that these can be divided in two types. First, core integrons, present at high frequency in a species and which are responsible for the generation of new gene cassettes. Second, integrons mostly present in mobile genetic elements, containing cassettes obtained from a multitude of other integrons. These mobile integrons seem responsible for the spread of cassettes created by the core integrons among bacteria. Accordingly, mobile integrons show higher diversity in their attC sites. This trait is exacerbated in attC cassettes without integron integrase in their genome, suggesting they are relics of former integrons. The study of these gene cassettes sheds lights on the fate of integron cassettes in bacterial genomes and how they may ultimately be integrated in the new host genetic background.

726A

Molecular Clocks Constrained by Horizontal Gene Transfers Predict an Archaeal Common Ancestor circa 3.9 Ga

Greg Fournier[1], J. Wolfe
[1] *Department of Earth, Atmospheric, and Planetary Science, Massachusetts Institute of Technology, Cambridge, USA*

The use of molecular clocks for dating early events in microbial evolution is an ongoing challenge for paleogenomics, as, generally, most ancient microbial lineages lack diagnostic fossil records for the absolute dating of nodes on phylogenetic trees. Therefore, the confidence in estimates for the divergence times of major microbial lineages has rested mostly upon the confidence in the unconstrained rate models used, which are sensitive to lineage-specific model violations, as well as other artifacts.

We attempt to surmount these obstacles by using horizontal gene transfer (HGT) events between archaeal lineages and clades with established fossil records and more reliable calibration dates. This propagates the age constraints of recipient nodes to the ancestors of donor lineages within the archaeal phylogeny, improving the precision and accuracy of molecular clock models. We present the results of this novel method, called TARDIS (Transfer Assisted Rate and Divergence time InferenceS).

Using TARDIS, we show that the most likely age for the last common ancestor of Archaea is ~3.9 Ga, strikingly close to the end of the proposed "Late Heavy Bombardment" series of impact events. This provides independent support for the "bottleneck" hypothesis of early life evolution, as proposed by Gogarten-Boekels, and is consistent with previous work in ancestral sequence reconstructions by Groussin and Boussau predicting hyperthermophilic Domain ancestors, but a mesophilic LUCA. In this model, early life diversified some time before 3.9 Ga, but only hyperthermophilic lineages survived major near-sterilizing impact events, presumably in deep-sea hydrothermal systems. These surviving lineages gave rise to the bacterial and archaeal Domain ancestors. This hypothesis explains the shape, depth, and ecological distributions observed in microbial phylogenetic trees. While substantial uncertainty remains in these estimates, expanding the HGT dataset, adding cross-calibration constraints, and investigating the ancestor age of Bacteria using similar methods will increase the utility of this approach.

727B

The contribution of genetic recombination to CRISPR array evolution

Anne Kupczok, Giddy Landan, Tal Dagan
Christian-Albrechts-University of Kiel, Kiel, Germany

CRISPR (clustered regularly interspaced short palindromic repeats) is a microbial immune system against foreign DNA. Recognition sequences (spacers) encoded within the CRISPR array mediate immunity in a sequence-specific manner. CRISPR array evolution is thought to be determined by spacer acquisition at the beginning of the array and deletion of successive spacers. Here we study the contribution of genetic recombination between homologous CRISPR arrays to the evolution of spacer repertoire. Acquisition of spacers from exogenic arrays may confer the recipient with immunity against unencountered antagonists. Spacer recombination can be detected by spacer content similarity not expected under vertical evolution alone. We analyze spacer content and order in four eubacterial species. We find that CRISPR array evolution in *E.coli* and *S.agalactiae* can be explained solely by vertical inheritance and differential spacer deletion. In *P.aeruginosa*, we find an excess of single spacers potentially incorporated into the CRISPR loci during independent acquisition events. In *S.thermophilus* we find evidence for the acquisition of spacers by recombination in numerous strains. A strong signal is present in five out of 70 unique strains. Genetic recombination has been proposed to accelerate adaptation by combining beneficial mutations that arose in independent lineages. However, for most species under study, we find that CRISPR evolution is shaped by spacer acquisition and loss rather than recombination. Since the evolution of spacer content is characterized by a rapid turnover, it is likely that recombination is not beneficial for improving phage resistance, or that it cannot be detected in the resolution of intra-species comparisons.

728C

The genome evolution and host adaptation of two arthropod-symbiotic *Spiroplasma*

Wen-Sui Lo [1,2], Gail E. Gasparich[3], Chih-Horng Kuo[1,2]

[1] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan, [2] Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung Hsing University and Academia Sinica, Taipei, Taiwan, [3] Department of Biological Sciences, Towson University, Towson, MD, USA

Spiroplasma are wall-less bacteria associated with arthropods in natural environments. To investigate their genome evolution and possible mechanisms of host adaptation, we completed the genome sequences of *Spiroplasma eriocheiris* (crab pathogen) and *S. mirum* (tick symbiont) for comparative analysis. As expected for symbiotic bacteria, their genome sizes are quite small compared to free-living bacteria, with the chromosome size ~1.2 Mb and containing ~1,200 protein-coding genes. The chromosomal organizations of these two closely related species are highly syntenic and their genes exhibited high sequence similarities (93% at the nucleotide level and 94% at the protein level). Interestingly, gene phylogenies indicated that 4% of the *S. eriocheiris* genes might have originated from horizontal transfers. These putatively transferred genes could be classified into four major clusters based on their chromosomal locations, metabolic functions, and possible donors, suggesting that each cluster has arisen from a single transfer event. The two genomes differed considerably in term of their pseudogene abundance. While only ~1% of the genes were pseudogenized in *S. eriocheiris*, the estimate is ~9% for *S. mirum*. Furthermore, the *S. mirum* pseudogenes include several of those putatively transferred genes, indicating that these horizontal transfer events occurred in the common ancestor of these two species. After the species divergence, these acquired genes might have contributed to the metabolic needs of *S. eriocheiris*, thus were maintained by selection. On the other hand, the genome degradation observed in *S. mirum* might be explained by the relaxation of selection or an increase in genetic drift.

729D

Lateral Gene Transfer is an overlooked mechanism in the adaptation of microbial eukaryotes to new environments

Laura Eme[1], Courtney W. Stairs[1], Anna Karnkowska[2], Vladimir Hampl[2], Andrew J. Roger[1]
[1] Dalhousie University, Halifax, Canada, [2] Charles University, Prague, Czech Republic

The importance of lateral gene transfer (LGT) in shaping eukaryotic genomes is controversial. Using a robust large-scale phylogenomic protocol, we have investigated the impact of LGT in shaping the genomes of three anaerobic microbial eukaryotes belonging to distinct eukaryotic super-groups: Blastocystis, a stramenopile human gut parasite, Pygsuia, a deep-branching Obazoan and the excavate Monocercomonoides. Phylogenomic analyses of Blastocystis revealed at least 65 very recent LGTs, unique to this lineage. A number of these genes are involved in adaptation to parasitism, as well as to anaerobiosis. Surprisingly, at least 10 were clearly acquired from other eukaryote lineages, a phenomenon that is usually considered to be extremely rare. In addition, genes originally acquired by LGT from prokaryotes seem to be often transferred between anaerobic protists, and are also generally massively duplicated after their acquisition. Several examples of LGT that were crucial to anaerobic adaptations were also found in the breviate microbe Pygsuia, and the bona fide amitochondrial excavate Monocercomonoides. Pygsuia, in the evolutionary remodelling of its mitochondria to anaerobiosis, has acquired an iron-sulfur cluster formation pathway (i.e., SUF) by LGT from Archaea, providing the only known example of the replacement of an essential ancestral mitochondrial system by an archaeal one. Similarly, Monocercomonoides, which is entirely devoid of mitochondria (or related organelles), has also replaced ancestrally mitochondrial pathways by cytosolic functional analogues, laterally acquired from prokaryotes or other anaerobic protists. These underscore the major importance of LGT in forging “metabolic mosaics” of ancestral and newly acquired pathways in eukaryotic protists.

730A

Endosymbiotic gene transfer in light of prokaryotic genome evolution

Chuan Ku, Shijulal Nelson-Sathi, Mayo Roettger, Filipa Sousa, William Martin
Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Genome sequences have shed much light on the origin of eukaryotic genes. It is now widely accepted that the bioenergetic organelles of eukaryotes, mitochondria and plastids, were once free-living prokaryotes whose genes were mostly relinquished to the nuclear genome in a process called endosymbiotic gene transfer. More recently, it has also been proposed that there were additional prokaryotes that donated genes to eukaryotes based on the observation that, in phylogenetic trees, the prokaryotic sister taxa are not always alphaproteobacteria or cyanobacteria, the lineages from which mitochondria and plastids are supposed to have originated. To date, the mechanisms behind such apparent lateral gene transfer (LGT) from prokaryotes outside the context of organelle origins remain largely unknown and, more importantly, its contribution relative to transfers from organelles is still unclear. To approach the latter problem, we examined the context for eukaryotic gene origins – prokaryotic genome evolution. There is strong evidence for pervasive LGT among prokaryotes based on genomics and phylogenetics. Prokaryote genomes are characterized by pangenomes: collections of genes shared by a taxonomic group that far exceed the number of genes found in any individual genome. LGT and gene loss, mechanisms for redistribution of prokaryotic genes, compounded by incomplete sampling, have resulted in apparent transfers from non-proteobacterial or non-cyanobacterial lineages to even organelle genomes, which are largely immune to LGT. Comparative analyses provide evidence that the vast majority of eukaryotic genes stem from the genomes of the host and two organelles, instead of cryptic symbionts or individual prokaryote-to-eukaryote LGT events.

731B

Why do eukaryotes have bacterial membranes?

Victor Sojo
University College London, London, UK

Recent phylogenetic evidence supports the endosymbiotic hypothesis for the origin of eukaryotes, by which a bacterium embedded within an archaeal host gave rise to the first eukaryotic cell. If so, this original eukaryote must have had an archaeal plasma membrane and bacterial proto-mitochondrial membranes; yet all known modern eukaryotes have exclusively bacterial membranes, both in their mitochondria and other organelles as in their boundary with the exterior. For some reason the archaeal phospholipid-synthesis machinery was lost and the bacterial one kept. The membranes of archaea and bacteria are utterly different, but given that archaea and bacteria coexist in many environments it is not clear why one type of phospholipid would have been favoured over the other in eukaryotes. In fact, genes for bacterial lipid biosynthesis had to be transferred from the proto-mitochondrion into the archaeal genome that now is the eukaryotic nucleus. Why wasn't the archaeal lipid machinery kept instead, given that it was already in the proto-nucleus? We hypothesise that this was due to bioenergetic reasons: as mitochondria became the powerhouses of the eukaryotic cell, energy production came to rely largely on them. The adaptation of bioenergetic mitochondrial proteins to their membrane meant that the bacterial phospholipids had to be kept. Replacing them with archaeal analogues would have led to maladaptation, loss of energy-conversion efficiency, and even deleterious leakage of reactive-oxygen-species. I will present phylogenetic evidence that supports this hypothesis, gathered from genome-wide comparisons across all three domains of life.

732C

Three's a crowd: Plastid establishment did not require a chlamydial partner

Daryl Domman [1], Matthias Horn[1], T. Martin Embley[2], Tom A. Williams[2]
[1] *University of Vienna, Vienna, Austria,* [2] *Newcastle University, Newcastle upon Tyne, UK*

The plastids of plants and algae trace their origin to a single, primary endosymbiosis, but the selective driver of this foundational event remains unclear. A recent hypothesis proposes that the initial role of the cyanobacterial endosymbiont was to provide energy to a host cell whose reserves were depleted by a chlamydial pathogen, and that proteins secreted by the chlamydia forged the initial metabolic link between host and symbiont through integration of bacterial metabolites into host energy stores. Here, we test this hypothesis against phylogenies of the key enzymes involved using better-fitting models of sequence evolution that account for heterogeneity in the evolutionary process, such as compositional differences across sites and among taxa. These factors may result in artifacts in the resulting phylogenies when not properly accounted for. Our analyses reveal a mosaic origin for plant energy metabolism in which genes were obtained from the host lineage, the cyanobacterial endosymbiont, and other bacterial groups. Crucially, we find no compelling evidence for a chlamydial origin for these genes under the best-fitting phylogenetic models, suggesting there is no need to invoke a chlamydial partner in the establishment of the primary plastid endosymbiosis. This study highlights the possible pitfalls of model violation when inferring ancient divergences and gene transfer events.

733D

LGT and chemical clues in metabolic evolution

Filipa L. Sousa, Shiju Nelson-Sathi, William F. Matrin
Molekulare Evolution Heinrich-Heine-Universität, Duesseldorf, Germany

Prokaryotes evolve by vertical and horizontal gene exchange. At the start of life, both components should have contributed to prokaryotic evolution as well. Can these ancient HGT events entail clues about the type of chemical reactions ancient organisms were performing to support their metabolism?

Recently, we have performed an analysis of 1981 prokaryotic genomes were HGT was shown to contribute to the emergence of the major archaeal domains(1).Taking these HGT events as our initial dataset, we propose to give some lights into the metabolism of ancient organisms. Since the oxygenation of the atmosphere is a biological invention, one way to distinguish between recent and ancient interdomain HGTs, is to consider the oxygen utilization of the organisms involved in the transfer. This allows us to separate the 1050 HGT occurring between more than one archaeal and bacterial group into “ancient” (occurring between anaerobic organisms) and “recent” (aerobic organisms). An inspection of the most frequent taxa involved in ancient LGT shows a high representation of both methanogenic and acetogenic organisms supporting the ancestry of methanogenic-like forms of life at the start of biological evolution. On the other hand, methyltransferases and SAM proteins are within the most represented protein families of these ancient HGTs. This indicates the importance of methyl groups in ancient metabolism, a feature still present in modern organisms, as shown by the high frequency of RNA and protein methylations in extant genomes.

1-Nelson-Sathi et al, Nature 2015

734A

Comprehensive analysis of the genomic localization of bacterial toxin-antitoxin systems

Francesc Peris-Bondia, Laurence Van Melderen
Laboratoire de Génétique et Physiologie Bactérienne, IBMM, Faculté des Sciences, Université Libre de Bruxelles (ULB), Brussels, Belgium

Bacterial toxin-antitoxin (TA) systems are composed of a stable toxin and an unstable antitoxin. TA are involved in plasmid addiction. These systems are widely spread in bacterial and archeal genomes. Although a variety of functions have been assigned to chromosomally-encoded systems, their biological roles are still unclear or limited to particular cases, making the 'selfish' behaviour an attractive hypothesis to explain the evolutionary success of TA. They are thought to invade bacterial genomes through horizontal gene transfer but it still remains unclear.
We have performed an exhaustive search of TA and a comprehensive analysis of their distribution, genomic organization and context. We have searched for similar genes in all the bacterial, viral and plasmidic genomes available in NCBI. We have determined if toxin and antitoxin ORFs co-localize and look for the presence of signature of mobile elements in the neighbouring genes.
Although the number of TA is big in chromosomes, the number of TA per gene is much bigger in plasmids. Also a big number of TA are pseudogenes. That indicates that in the chromosome either TA has not role, either its role is temporary. In any case, this is only possible with high rates of horizontal gene transfer. In general, TA are part of the accessory genome but we can see few families being part of the core genome. Showing that the latter ones have probably acquired different roles in bacterial physiology throughout evolution. Understanding TA distribution is necessary for a full comprehension of mobile genetic elements evolution.

735B

Characterization of antimicrobial resistance gene exchanges between antibiotic-exposed and environmental microbiota following soil manure-amendement.

Sebastien Leclercq, Chao Wang, Xiaochen Du, Zhihai Sui, Jie Feng
State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

The emergence of multi-resistant pathogenic bacteria in the last 30 years is considered as one of the largest threat of modern medecine. Antibiotic resistance genes (ARGs) are easily horizontally transferred between bacteria through mobile genetic elements (MGEs). Moreover, the last few years have revealed that an unexpected high frequency of ARG exchanges occur between clinical pathogens and environmental bacteria.
One preferred route for these exchanges is the fertilization of agricultural fields with manure from antibiotic-exposed animals. However, the mode and frequency of these exchanges are still unclear. In this project, agricultural soils were experimentally fertilized with selected swine manure to study the dynamics of ARG-carrying MGE horizontal transfers between the two bacterial communities. The combination of functional metagenomics, high throughput sequencing, and high-level molecular biology methodologies provided us valuable information on how and at which frequency these MGEs may be transferred between bacterial communities. These results can significantly contribute to our understanding of the impact of agronomic processes, such as manure amendment, on the evolution of antibiotic resistance.

736C

The loose evolutionary relationships between transcription factors and other gene products across prokaryotes

Gabriel Moreno-Hagelsieb
Wilfrid Laurier University, Waterloo, ON, Canada

Using patterns of co-occurrence between genes whose products are experimentally known to interact in Escherichia coli, we show results suggesting that genes coding for transcription factors (TF genes) and genes coding for their target genes have a weak association when compared to other types of associations. The other types of associations were genes in the same operons, genes coding for proteins that work in the same biochemical pathways, and genes coding for proteins that physically interact. We further show that TF genes in prokaryotes other than E. coli tend to have lower co-occurrences with any other genes, than the rest of genes, which suggest that the results obtained with experimentally known interactions might hold for most if not all other prokaryotes. Since these results suggest quick transfer of TF genes, we checked their codon adfaptation indexes (CAI) and found that they tend to be significantly lower that the CAI of other genes. Overall, these resuults suggest that TF genes might be among the most rapidly evolving interactions in prokaryotes.

737D

Measuring phylogenetic bias in gene transfer from genome-scale data

Zsófia Nóra Kéri [1], Gergely János Szöllősi[1], Vincent Daubin[2]
[1] ELTE-MTA 'Lendület' Biophysics Research Group, Budapest, Hungary, [2] Laboratoire de Biométrie et Biologie Évolutive, Unité Mixte de Recherche CNRS, Université Claude Bernard – Lyon 1, Villeurbanne, France

Lateral gene transfer (LGT) is a well-known evolutionary process that is known to occur between both closely and distantly related bacterial species. Although several methods have been developed to detect LGT by comparing gene and species phylogenies, a quantitative assessment of how the rate of transfers depends on the evolutionary distance of the host and donor species is lacking. Using a maximum likelihood method implemented in ALE (Szollosi et al. 2013) I inferred LGT events for more than 4900 gene families from over 36 cyanobacterial genomes. Analysing the distribution of transfers using a statistical test I developed, I found that LGT rates are biased towards more closely related species. I am currently working on extending the analysis to other groups of bacterial genomes in order to establish the generality of the results.

738A

Experimental evolution uncovers physical barriers to Horizontal Gene Transfer

Shimon Bershtein [1,2], Adrian Serohijos[1], Sanchari Bhattacharyya[1], Michael Manhart[1], Eugene Shakhnovich[1]

[1] Harvard University, Cambridge MA, United States Minor Outlying Islands, [2] Ben-Gurion University of the Negev, Beer-Sheva, Israel

Horizontal gene transfer (HGT) is a major force in bacterial evolution, yet the molecular constraints to accommodation of new genes remain largely unknown. Here we generated 36 E. coli strains whereby chromosomal folA gene encoding essential metabolic enzyme dyhidrofolate reductase (DHFR) has been replaced by orthologs from other bacteria. We found that molecular properties (thermodynamic stability and enzymatic activity) of the orthologous DHFR are not good predictors of fitness - despite being stable and fully active, most horizontally transferred orthologous DHFRs caused a marked drop in fitness (15-90%) compared to wild-type strain. However, the ‘'enzyme fitness", i.e., the product of the intracellular DHFR abundance and enzymatic activity (kcat/KM) did correlate with growth rate, indicating that the drop in DHFR abundance upon HGT is the major cause of the observed fitness effects. Using high-throughput experimental evolution we propagated the orthologous strains for approximately 3,000 generations under standard growth conditions and found that fitness of the evolved strains has improved dramatically. Strikingly, after the evolution, there appeared a highly statistically significant correlation between growth rate of an evolved strain and the catalytic activity (kcat/KM) of its DHFRs (Spearman R= 0.57, p = 0.0007). Whole genome sequencing revealed that the evolved strains that showed the most dramatic increase in fitness have accumulated genomic rearrangements causing inactivation of the ATP-dependent Lon protease followed by an increase in the intracellular DHFR abundance. Thus, protein quality control (PQC) constitutes the first "crude" barrier to HGT by decreasing the intracellular abundance of the newly acquired proteins.

739B

Protease inhibitors in pathogenic bacteria are novel virulence factors

Minca Ferlin, Dusan Kordis
Josef Stefan Institute, Ljubljana, Slovenia

Cystatins are natural inhibitors of cysteine proteases. During the analysis of cystatin superfamily, we have found that eukayotic cystatins and stefins have been horizontally acquired by a few bacteria. Their role in bacteria could be to evade host immunity. We performed a detailed phylogenomic analysis of the cystatins and stefins in ~30.000 prokaryotic genomes. Functional and structural characterization of bacterial stefins and cystatins has shown that they possess conserved cystatin fold and the inhibitory motif QXVXG. To demonstrate the biochemical activity of bacterial stefins we expressed *Vibrio cholerae* stefin (VCA0935) and *Bacteroides fragilis* fusion inhibitor containing chagasin and stefin (BF1388). We explored the inhibitory properties of recombinant VCA0935 and BF1388 proteins and determined their interaction constants with diverse cysteine proteases, cathepsins L, S, K, V, B and papain. Both VCA0935 and BF1388 were found to act as fast and tight binding inhibitors of endopeptidases cathepsins K, S, V, L and papain. Interestingly, the pathogen stefins inhibits the endopeptidase activity of cathepsins S, K, L and V, which are all important players in the host adaptive and innate immunity. The acquisition of novel virulence factors for pathogenic bacteria by horizontal gene transfer from eukaryotes is very rare. Very few cases are known where protease inhibitors assist pathogens in invading the eukaryotic hosts by inhibiting their proteases. Bacterial stefins and cystatins are especially suited to inhibit the numerous host cysteine proteases during infection. Therefore, they are novel virulence factors that function in the invasion and dissemination of the pathogens.

740C

A novel method for selective sweep detection in bacteria

Oren Avram, Yaara Oren, Eli Levy Karin, Tal Pupko
Tel Aviv University, Tel Aviv, Israel

Selective sweep is a process in which a new mutation that confers a selective advantage is dispersed throughout the population.

While widely investigated in eukaryotes, selective sweeps received little to no attention in microbial studies. It was commonly assumed that bacteria have low recombination rates and that once a bacterium gains an adaptive mutation, clonal expansion is more likely to happen than recombination. Thus, searching for sweeps among bacterial populations was considered to be fruitless.

However, in our previous study (Oren et al. PNAS 2014), which examined regulatory transfer across different bacterial clades, we detected several cases of promoters that underwent a clear selective sweep. These surprising observations suggest that selective sweeps in bacteria may be much more common than previously thought.

We are currently developing novel methodologies to detect selective sweep events in bacteria. Our approach accounts for bacterial-specific characteristics, such as the high sequence variability between strains. The methodologies are characterized on simulated data and are applied to detect sweeps among several bacterial species, for which enough strains were sequenced.

Detecting selective sweeps in bacteria would allow for a better understanding of the evolutionary mechanisms underlying bacterial infections. Such understanding may contribute to the ongoing effort to detect and fight pathogens.

741D

Large-scale study of genetic exchange through bipartite graphs

Eduardo Corel, Raphaël Méheust, Philippe Lopez, Eric Bapteste
UPMC UMR7138, Paris, France

Introgressive events are recognized as an important driving force in the evolution of prokaryotes. A convenient way of representing the complexity of this exchange of genetic material is to construct
large-scale similarity networks. I will specifically focus on the use of algorithms on bipartite graphs and apply them to the study of the adaptation to lifestyle in prokaryotes, and to the characterisation of their pathogenicity.

742A

VESICLE MEDIATED DNA TRANSFER IN FRESH-WATER CYANOBACTERIA

Claudia Walda, Karina Stucken, Ben Krause-Kyora, Tal Dagan
Christian-Albrechts-University, Kiel, Germany

Outer membrane vesicles (OMV) are described as discrete, spherical blebs having a bilayer or double bilayer membrane. Their average diameter ranges between 20 and 250 nm. OMVs are secreted by all gram-negative bacteria tested so far, as well as several gram-positive bacteria (e.g. *Acinetobacter baumannii*, *Staphylococcus aureus*), and marine cyanobacteria (e.g. *Prochlorococcus*). OMVs are thought to contribute to various extracellular functions including intercellular communication. Accumulating evidence indicate the presence of DNA in OMVs of various microbial species, suggesting a putative role as a DNA transfer mechanism. However, the cellular mechanisms involved in OMV production, and regulation and their possible contribution to DNA transfer are yet unknown. Here we report the finding of OMVs in the freshwater cyanobacteria *Synechocystis* sp. PCC 6803, *Anabaena* sp. PCC 7120, and *Chlorogloeopsis fritschii* PCC 6912. The vesicles were isolated from stationary growth cultures by filtration and ultracentrifugation. The luminal content was fluorescent when stained with SYBR gold, indicating the presence of nucleic acids. Interestingly, TEM images of *C. fritschii* PCC 6912 revealed also a phage that is probably temperate. OMV genomics of *Anabaena* sp. PCC 7120 revealed that DNA stored in the vesicles matches the whole genome in a low coverage. Two transposase genes were overrepresented in the OMV genomics with x100 to x200 fold higher coverage in comparison to the remaining genes. Current research is focused on the characterization of OMV content under different environmental conditions and OMVs transfer efficiency.

743B

Evolution of bacterial regulatory RNAs from vestigial bacteriophage genes

Fenil Kacharia[1], Justin Merritt[2], Rahul Raghavan [1]
[1] Portland State University, Portland, USA, [2] Oregon Health and Science University, Portland, USA

RNAs that do not code for proteins are critical to gene regulation in all domains of life. In bacteria, regulatory RNAs, including small RNAs (sRNAs) control the expression of protein-coding genes by modulating transcription, translation or mRNA stability. The application of technologies that interrogate entire transcriptomes has revealed unexpectedly large numbers of sRNAs in bacteria. However, unlike the mechanisms of protein evolution, the mechanisms driving the evolution of regulatory RNAs are poorly understood. The *Escherichia coli* genome contains several bacteriophage genomes (prophages) in various stages of degradation. In this study using a combination of comparative and experimental analyses we identified a novel sRNA (EcsR2) that arose *de novo* in *E. coli* from a vestigial prophage tail fiber assembly gene (*tfa*). The *E. coli* chromosome contains numerous intact and pseudogenized copies of *tfa*, revealing the cyclic process of *tfa* acquisition and loss in the bacterium. However, one of the copies of *tfa* appears to have escaped this fate by evolving into a functional RNA. Similar to most other sRNAs in *E. coli*, EcsR2 is dependent on the Hfq protein for stability, and it binds to target mRNAs to regulate their abundance. We further show that evolution of new regulatory RNAs from erstwhile prophage genes is a widespread phenomenon in bacteria by defining the evolutionary history of a dual-function riboregulatory mRNA in *Streptococcus mutans* that emerged from a bacteriophage *cro* gene.

744C

Tracing natural competence in the bacterial domain using a function aware phylogenetic profiling.

Holger Bergmann, Julian Dosch, Ingo Ebersberger
Goethe University, Frankfurt, Germany

The dissemination of genomic DNA via horizontal gene transfer plays an important role in the rapid adaptation of bacteria to changing environments. Natural competence - the ability to uptake and utilize DNA directly from the environment - is the most versatile way of mining genetic diversity across species borders. To date, five different DNA uptake machineries have been described on the molecular level, and about 120 species are known as naturally competent under varying conditions. Initial approaches to trace the individual components of the competence machinery across the bacterial tree of life suggest, however, that natural competence might be far more common among bacteria. Yet, these surveys have been limited to the analysis of local sequence similarities leaving the question about functional equivalency of the identified candidates unaddressed. Here we introduce a novel phylogenetic profiling approach to overcome this limitation. We integrate a targeted ortholog search with a scoring of pairwise protein similarity on the feature architecture level. Features comprise, among others, functional protein domains, secondary structure elements, transmembrane domains and low complexity regions. In the case of overlapping, redundant annotations in the architecture, the highest scoring linear path resolving the redundancy is used. The resulting feature architecture similarity score serves then as a proxy for the degree of functional equivalence between the identified ortholog pairs. We apply our procedure to trace the DNA uptake machineries of five naturally competent bacteria in more than 1,000 species shedding light on the distribution and evolution of natural competence in the bacterial domain.

745D

Ongoing recombination in a natural population of *Legionella pneumophila*

Leonor Sanchez-Buso[1,2], Iñaki Comas[1,2], Fernando Gonzalez-Candelas [1,2]
[1] *Unidad Mixta "Infección y Salud" FISABIO-Universidad de Valencia, Valencia, Spain,* [2] *CIBERESP, Valencia, Spain*

Legionella pneumophila is a strictly environmental pathogen and the etiological agent of legionellosis. It is known that non-vertical processes play a major role in the short-term evolution of pathogens but little is known about the relevance of these and other processes in environmental bacteria. We have used next generation sequencing to obtain nearly complete genome sequences of 69 *L. pneumophila* strains linked to recurrent outbreaks in a single location (Alcoy, Spain) along 11 years. Many (n=45) of these isolates belonged to the ST578, usually found in clinical samples from this location but rarely isolated from environmental samples or from other locations. Despite the relatively short time and the limited geographical area of sampling we found that none of these isolates was identical to any other. The number of SNPs along the complete genomes spanned between a minimum of 5 to a maximum of 1807 and were not related to the dates or outbreak when they had been isolated. Our analyses reveal that 16 recombination events are responsible for almost 98% of the SNPs detected in their core genome and an apparent acceleration of the evolutionary rate. These results have profound implications for our understanding of microbial populations and for public health interventions in *Legionella* outbreak investigations

746A

Toward understanding how mobile genetic elements rendered *Klebsiella pneumoniae* hyper-virulent or multi-resistant

Camille Blin [1,2], Virginie Passet[1,3], Marie Touchon[1,3], Sylvain Brisse[1,3], Eduardo P Rocha[1,3]
[1] Institut Pasteur, Microbial Evolutionary Genomics, Paris 75015, France, [2] Sorbonne universités, UPMC Univ Paris 06, Paris 75005, France, [3] CNRS, UMR3525, Paris 75015, France

Klebsiella pneumoniae (Kpn) is an opportunistic pathogen frequently responsible for nosocomial infections and is emerging as a cause of invasive, community-acquired infections. Infections by multidrug-resistant (MDR) Kpn isolates are becoming major therapeutic challenges. Many of their resistance traits are carried and spread by mobile genetic elements (MGE). Invasive community-acquired isolates possess virulence genes that are also carried by MGE. Interestingly there are almost no isolates carrying both virulence and resistance genes, suggesting the existence of sign epistasis between the two. We analyzed 34 complete and 18 draft genomes of genus Klebsiella to study the emergence of virulent and multi-resistant Kpn within this genus (which has mostly commensal, environmental or plant-associated bacteria). We computed core and pan genomes of the species, analyzed patterns of selection and detected events of homologous recombination. These analyses provided an overview of the genetic population structure and how it was shaped by recombination and natural selection. We have established an exhaustive inventory of Kpn mobile genetics elements including plasmids, ICE, and prophages, most of which have never been described. We are now analyzing the distribution of these elements in the light of the population structure, the genome organization and the accessory traits they encode. We will then be able to assess their contribution to the evolution of virulence and resistance to antibiotics and their possible dependency on specific genetic backgrounds.

747B

The phylogeny of *Rickettsia* using different evolutionary signatures: How tree-like is bacterial evolution?

Gemma Murray, Lucy Weinert, Emma Rhule, John Welch
University of Cambridge, Cambridge, UK

Rickettsia is a genus of bacterial endosymbionts whose hosts and transmission strategies are both impressively diverse. This is reflected in a dynamic genome, which varies greatly in size and content. Several previous studies have used molecular data to estimate this phylogeny, however, these reconstructions differ. One possible reason for this uncertainty is the non-tree-like or reticulate nature of bacterial evolution. In this study, we reconstruct the *Rickettsia* phylogeny using whole-genome data, including two new genomes of strains from previously un-sampled host groups. Our study uses several approaches to identify and correct for systematic error. These include the use of non-stationary models, multiple quasi-independent sources of phylogenetic signal, and simulations over known topologies. We also test for evidence of a plurality of evolutionary histories across the genome, resulting from recombination or horizontal gene exchange. Taken together, our results suggest that much of the apparent evidence for reticulate evolution in *Rickettsia* stems from systematic error, sometimes resulting from evolutionary processes that are not well described by current phylogenetic models. After accounting for this error, we argue that the evolution of *Rickettsia* is largely well described by a single topology. Our results demonstrate the need to consider sources of bias in phylogeny construction, particularly over long evolutionary distances, or with dynamic genomes, when testing for reticulate evolution.

9 Inference of demography and selection under violations of the Kingman coalescent assumptions

9.1

The effect of past demography and natural selection on the rate of evolution

Francois Balloux
Department of Genetics, Evolution and Environment, University College London, London, UK

Over recent years, there has been considerable interest in the effect of violations to the underlying assumptions to the Coalescent. However most this work has focused on fixation probabilities and genetic diversity. There has been less interest in the effect of such deviations on the substitution rate (i.e. the rate of evolution), despite the fact that this is a critical quantity for the analysis of genetic data in population genetics and phylogenetics. Indeed, it is widely accepted that the substitution rate at neutral genes is unaffected by past population demography. This classical result has implications for the analysis of genetic data in population genetics and phylogenetics, and provides a justification for the concept of the molecular clock. In this talk I will present theoretical and empirical evidence on the effect of past demography and natural selection on the substitution rate in a variety of biological systems. I will in particularly show that the substitution rate at neutral genes is not independent on population size fluctuations in organisms with overlapping generations.

9.2

Estimating effective population size under inbreeding

Benjamin Redelings, Seiji Kumagai, Marcy Uyenoyama
Duke University, Durham, NC, USA

We present Bayesian estimates of effective population size in organisms reproducing through a mixture of self-fertilization and random outcrossing. We consider two indices of effective number: N_i, determined by the probability of non-identity between genes randomly sampled from distinct individuals, and N_p, determined by the rate of parent-sharing between lineages ancestral to the sampled genes. Using a modified coalescence framework, we describe the dependence of these indices on the rate of selfing. Under pure hermaphroditism, effective number monotonically declines with selfing rate. Under androdioecy and gynodioecy, however, both indices of effective number depend on the relative proportions of sexual forms (males, females, hermaphrodites), and effective number can show a non-monotonic dependence on the selfing rate. Our Bayesian method uses genetic data as the basis for estimates of selfing rate, effective size, and other aspects of reproduction. Results obtained from our analysis of a dataset from an androdioecious organism indicate a 20% decline in N_p relative to the number of reproductives and an even greater decline in N_i. Results from a gynodioecious organism indicate a decline in N_i, but not N_p, relative to the number of reproductives. To our knowledge, these quantities constitute the first direct estimates of effective population size.

9.3

The seed-bank coalescent

Jochen Blath, Bjarki Eldon, Adrian Gonzalez Casanova, Noemi Kurt, Maite Wilke Berenguer
TU Berlin, Berlin, Germany

Many microorganisms can enter (and leave) a dormant state of vanishing

metabolic activity, for example when faced with unfavourable environmental con-

ditions. Such dormant forms may stay inactive for extended periods of time and

thus create a seed-bank that strongly affects the interplay of evolutionary

forces driving the genetic variability of the population. In this paper, we investi-

gate the effect of `srong' seed-banks on the genealogy and patterns of genetic

variability in large populations over macroskopic timescales.

It turns out that under appropriate conditions, a new coalescent model arises naturally

as limiting genealogy of a Wright-Fisher model if one takes seed-banks explicitly into account.

This new probabilistic structure, which we call the `seed-bank coalescent',

exhibits qualitatively new features, some related to the so-called structured coalescent with two islands,

others similar to effects visible in the so-called `Bolthausen Sznitman coalescent'. In particular, our model leads beyond results

on `weak' seed-banks effects previously investigated by Kaj, Krone and Lascoux.

After deriving the seed-bank coalescent with mutation, we present implications on the evolutionary parameters and patterns of genetic variability reflected in the site-frequency spectrum. We also discuss inference methods for determining the presence and `strength' of the seed-bank.

9.4

Selection efficiency in chronic and acute malaria infections: The effect of overlapping generations

Hsiao-Han Chang, Lauren Childs, Caroline Buckee
Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Controlling malaria requires an understanding of how malaria parasites evolve to defeat the host immune system, but the evolutionary dynamics of the parasite within and between hosts are poorly understood. Malaria infections may be acute or chronic in humans. Acute infections are short, analogous to non-overlapping generations. Chronic infections persist for several months, allowing multiple transmission events and are equivalent to overlapping generations. Using non-overlapping generations for the malaria life cycle has been shown to skew the null distribution of allele frequency and lead to false signals of population expansion or selection. Although the importance of chronic infections has been recognized, it is unclear how overlapping life cycles may influence our understanding of malaria evolution. Here, we develop a population genetic model of acute and chronic infections, and we show that the probability of fixation for beneficial alleles is higher in the chronic-infection model. We propose that greater selection efficiency in the chronic-infection model is because parasites can be transmitted at multiple time points due to the longer within-host persistence. Furthermore, we combine population genetic and epidemiology models, and show that, in a population with both acute and chronic infections, beneficial mutations that have occurred in chronically infected patients have higher fixation probability. Interestingly, the fixation probability increases as the proportion of chronic patients and transmission intensity decreases. This offers a solution on why most drug resistance mutations have emerged from Southeast Asia but not Africa although transmission is lower and the effective population size is smaller in Southeast Asia.

9.5

Disentangling demography and selection effects in the cattle genome - new insights from the 1000 bull genomes project.

Simon Boitard [1,2], Marlies Dolezal[4], Bertrand Servin[3], Daniel Fischer[6], Jared Decker[8], Iona MacLeod[9], QianQian Zhang[7], Bernt Guldbrandtsen[7], Mogens Lund[7], Allessandro Bagnato[5], Johanna Vilkki[6], The 1000 bull genome project na[10]

[1] Génétique Animale et Biologie Intégrative, INRA & AgroParisTech, Jouy en Josas, France, [2] Institut de Systématique et Evolution, MNHN & EPHE & CNRS & UPMC, Paris, France, [3] Génétique, Physiologie et Systèmes d'élevage, INRA & ENSAT, Toulouse, France, [4] Vetmed Uni Vienna, Vienna, Austria, [5] Universita Degli Studio di Milano, Medicina Veterinaria, Milano, Italy, [6] MTT Agrifood Research, Jokioinen, Finland, [7] Aarhus University, Aarhus, Denmark, [8] University of Missouri, Columbia, USA, [9] University of Melbourne, Melbourne, Australia, [10] http://www.1000bullgenomes.com, na, Australia

Modern cattle genomes have been shaped by a complex interplay of human mediated demographic and selective effects. The initial domestication of taurine cattle around 10,000 years before present, its subsequent importation to Europe through different migration routes, or the much more recent creation of modern breeds, have resulted in a series of bottleneck, isolation and admixture events. In parallel, animals have been subjected to artificial selection for a wide range of traits, including milk and meat production, fertility, stature, coat color or mild temper. The recent sequencing by the 1000 bull genomes project of about 1,400 bulls from more than 40 taurine breeds, offers a great opportunity to explore and disentangle these processes. Based on these data, we first estimate a history of population size changes for each of the breeds, from the last few generations before present back to 100,000 years before present, using a new Approximate Bayesian Computational approach allowing to take advantage of large samples of genomes. These estimations confirm the decline of cattle population size since domestication, and outline interesting differences between breeds or groups of breeds. We then scan the cattle genome for regions whose genetic diversity cannot be explained by this estimated demographic model, using the CLR and iHs statistics. This leads to detect a series of genes related to caot color, stature or production traits, as clear targets of positive selection.

294A

Influence of drift and large variance in offspring production on coevolutionary dynamics and likelihood to observe balancing selection

Aurelien Tellier
Technische Universitaet Muenchen, Freising, Germany

Coevolution between hosts and their parasites is expected to follow a range of possible dynamics, the two extreme cases being called trench warfare (or Red Queen) and arms races. Long-term stable polymorphism at the host and parasite coevolving loci is characteristic of trench warfare, and is expected to promote molecular signatures of balancing selection, while the recurrent allele fixation in arms races should generate selective sweeps. We compare these two scenarios using finite size haploid coevolutionary models that include both mutation and genetic drift with large variance in offspring production. Indeed, many pathogen micro-organisms may typically undergo recurrent bottlenecks during disease transmission and/or produce large amount of spores or infectious propagules (on the order of the population size), thus violating the classic assumption of the Wright-Fisher model. Polymorphism in host and parasite populations can be maintained when using an appropriate time rescaling under the model with large variance in offspring production. We subsequently perform coalescent simulations using the Beta-coalescent under these dynamics to generate site-frequency spectra at both host and parasite loci. Genomic footprints of recurrent selective sweeps are often found, whereas trench warfare yield signatures of balancing selection only in parasite sequences, and only in a limited parameter space. Our results suggest that it may be best to study parasite rather than host populations to find genomic signatures of coevolution, such as selective sweeps or balancing selection.

295B

Inferring population demography and mutation rate using jointly pedigree information and genomic data

Florence Parat [1], Daniel Wegmann[2], Arélien Tellier[1]
[1] Technische Universität München, Freising, Germany, [2] University of Fribourg, Fribourg, Switzerland

Inferring complex demographies of populations, that potentially violate common model assumptions, using genetic data is a challenging task. Part of the solution might be to add an other source of information: Pedigrees. To apply this idea, we develop a Likelihood method that allows the joint use of information contained in pedigrees and genetic data to increase the power of inference and allow the estimation of parameters out-of-reach with usual methods, as the mutation rate or the sex ratio of a population.
Our reasoning is as follows. On the one hand, the pedigree of a random sample of individuals contains information on the recent demography of the population. On the other hand, the genetic data, summarized as a Site Frequency Spectrum (SFS), is the result of both the demography, ancient (time of MRCA) and recent, and the mutational process. Our method allows inference of both demographic and mutational parameters from the genetic data by tracing back the possible genetic histories of our sample constrained by the pedigree under a given mutation model, and by calculating the probability of this pedigree given the demographic model.
Our method is particularly suitable for inference of past events, population sizes and mutation rates in domesticated species (animals or plants) for which large pedigrees and full genome sequences are available. We apply our inference method to cattle populations with life history traits which are known to violate the classic coalescent assumptions such as overlapping generations, unbalanced sex ratio, small effective population size, and skewed offspring distribution.

296C

Statistical identifiability of non-Kingman coalescent processes from molecular data

Jere Koskela, Paul Jenkins, Dario Spano
University of Warwick, Coventry, UK

Skewed offspring distributions, natural selection and range expansion of populations have all been shown to give rise to genealogies described by a family of coalescent processes known as Lambda-coalescents. The family provides a rich modelling framework, which is able to capture skewed branching of ancestral trees, simultaneous mergers of more than two lineages and associated insensitivity of nucleotide diversity to population size. Such features have been observed recently in genealogies of Atlantic cod and Pacific oysters populations, whale louse mitochondrial DNA as well as influenza viruses under strong selective immune pressure.

In contrast to the universality of the Kingman coalescent, the statistics of Lambda-coalescent trees depend on the details of the skewness in offspring distribution, the strength and mechanism of natural selection or population expansion etc. Hence it is of great interest to ask whether they can be correctly learned from molecular data. In this talk I will show that such identifiability fails in general when observations are taken over relatively few generations, such as a single time point, even when many unlinked loci are available. A long time-series of molecular data results in much stronger identifiability of Lambda-coalescent processes from molecular data under verifiable modelling assumptions. Recent advances in sequencing technology have resulted in time series data becoming available, and these results provide a strong motivation for its use, particularly whenever the usual Kingman assumptions are violated.

297D

Reference Models in Molecular Evolution

Marguerite Lapierre [1,2], Cécile Delaporte[3,4], Amaury Lambert[3,4], Guillaume Achaz[1,4]
[1] UMR 7138 Évolution Paris Seine, équipe Évolution des génomes eucaryotes, IBPS & UPMC & CNRS, Paris, France, [2] Atelier de Bioinformatique, UPMC, Paris, France, [3] UMR 7599 LPMA, équipe Probabilités, Statistiques et Biologie, UPMC & CNRS, Paris, France, [4] UMR 7241 CIRB, équipe SMILE, Collège de France, Paris, France

The current model of reference in Molecular evolution is the Neutral Theory introduced by M Kimura, which states that the vast majority of mutations that segregate through generations are neutral. This theory and the mathematical model it relies on have four major hypotheses: neutrality, panmixia, constant population size and a low variance of offspring number. In practice, to account for real situations, the two first hypotheses are often amended with demography and sub-populations, but the two other hypotheses are almost never reconsidered. Alternative models have been theoretically described: the birth-death model (branching process) and the coalescent with multiple mergers. These models allow to take into account variation in the population size, a foundation event, selection or variability in the number of descendants. I will present these alternative models and how they fit the data of large population sequencing projects, such as the 1000 Human Genomes project. Moreover, we study how these different models can converge to similar ones in a certain range of parameter.

10 Speciation genomics

10.1

Speciation genomics: divergence continuum and beyond

Christian Lexer [1], Margot Paris[1], Dorothea Lindtke[1], Celine Caseys[1], Camille Christe[1], Raphael O. Wüest[2], Eligio Bossolini[3], Michael Kessler[4], Nicolas Salamin[5], Kai N. Stölting[1]
[1] University of Fribourg, Fribourg, Switzerland, [2] Univ. Grenoble Alpes LECA, Grenoble, France, [3] Bayer CropScience, Gent, Belgium, [4] University of Zurich, Zurich, Switzerland, [5] University of Lausanne, Lausanne, Switzerland

Rapid recent progress in ecological & evolutionary genomics is imparting fresh perspectives to the study of population divergence and speciation, i.e. the origin and maintenance of biological diversity. Our group's research interests revolve around the use of novel laboratory and computational tools for studying adaptive evolutionary responses, speciation, and species radiations in plants (e.g. New Phytologist 196:652-654, 2012; Molecular Ecology 22:842-855, 2013; Journal of Biogeography 40, 1013–1022, 2013; Heredity 111:474–485, 2013; Ecology Letters 16:1515–e7, 2013; Molecular Ecology 23:4373–4386, 2014; Molecular Ecology 23:4316–4330). In my talk, I will highlight recent progress of our work. I will focus on phenome and whole-genome perspectives on population divergence along the entire 'speciation continuum' in a model plant group (*Populus* spp.), and related research on species-rich radiations in biodiversity hotspots that have not been on the "radar" of speciation genomics thus far. This will include selected examples from our ongoing work on South African and South American plant radiations. I will highlight the gap that currently exists between genomic research at the micro-evolutionary scale (i.e. population divergence and speciation) and research at the macro-scale of entire species radiations in most groups of animals and plants. I will suggest how this gap might potentially be closed, in model and non-model groups, by integrating speciation genomics more closely with phylogenomics and spatially explicit approaches from ecology.

10.2

Rapid and complex genomic divergence during adaptive speciation in wild Solanum

Leonie Moyle [1], James Pease[1], David Haak[1,2], Matthew Hahn[1]
[1] *Indiana University, Bloomington, Indiana, USA,* [2] *Virginia Tech, Blacksburg, Virginia, USA*

Examining genomic divergence across closely related species can reveal the changes underpinning adaptive divergence and speciation. We have generated multi-tissue transcriptomes to examine genomic divergence among all species in the diverse plant group *Solanum* section *Lycopersicon* (wild tomatoes). This clade of 13 diploid species is ~2 MYA, and extant species have substantial ecogeographic, morphological, and physiological differences, in addition to well-characterized pre- and postzygotic isolating barriers. Transcriptomes from 29 genotypes (1-3 accessions from each species, plus outgroups) generated >190 Gb of sequence (~6.6 Gb per accession). *De novo* and reference-based assembly generated a set of >16,000 high quality orthologs across all lineages (>45% of known coding regions). Average pairwise sequence divergence among accessions was <1%, consistent with very recent differentiation among species. Phylogenomic reconstruction revealed a consensus topology that agreed with relationships previously inferred from a smaller number of loci. However, we also observed rampant phylogenetic incongruence across this tree. Statistical tests of two alternative explanations of this incongruence strongly support an inference of widespread incomplete sorting of shared ancestral variation rather than widespread ongoing or recent introgression. Our genomic data support very rapid recent diversification in this group, and suggest that adaptive and isolating changes between species are drawn from a relatively modest number of observed fixed differences between lineages. In conjunction with ecological and genetic data, these phylotranscriptomic data are being used to reveal the genomic substrate of adaptive diversification in this group, as well as the specific loci that contribute to ecological and reproductive differentiation among lineages.

10.3

Heterogeneous genomic differentiation during repeated sympatric ecological speciation in Midas cichlid fishes

Axel Meyer, Andreas Kautt, Gonzalo Machado-Schiffiano
University of Konstanz, Konstanz, Germany

Genome-wide data reveal an often highly heterogeneous pattern of genomic divergence during speciation. Disentangling the dynamic effects of divergent selection and gene flow from the stochastic effects of a population's demographic history remains difficult. Cases of recent sympatric speciation seem promising in this regard, since selection had to be strong enough to overcome gene flow and may thus leave distinct signatures in the genome, while the confounding effect of genetic drift is usually assumed to be negligible. Midas cichlid fishes (*Amphilophus sp.*) inhabiting small and isolated crater lakes in Nicaragua form young (only < 2 - 22,000 years old) and monophyletic flocks of endemic species and no geographic barriers exist in these lakes. Thus, sympatric ecological speciation is the most likely mode of speciation. Moreover, Midas cichlids provide natural replicates of this process and several species seem to be at different stages of the speciation continuum. Based a comprehensive RADseq data set (of > 700 individuals) the source population and the crater lakes to (i) infer the demographic history of the speciation processes of small-scale radiations of Midas cichlids based on the joint site frequency spectrum and full-likelihood coalescent simulations and (ii) take this information into account when describing their differentiation at a genomic level. Overall, we find evidence for colonizations by small founder populations, speciation in sympatry, and document a highly heterogeneous landscape of genomic differentiation. Particular focus was on patterns of differentiation in previously identified QTL regions presumably underlying traits involved in ecological speciation in these species.

10.4

Genomics of Divergence along a Continuum Incipient Speciation

Philine Feulner
Eawag Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland

The patterns of genomic divergence during ecological speciation are shaped by a combination of evolutionary forces. Processes such as genetic drift, local reduction of gene flow around genes causing reproductive isolation, hitchhiking around selected variants, variation in recombination and mutation rates are all factors that can contribute to the heterogeneity of genomic divergence. Here, I will summarize a comprehensive assessment of genomic patterns of divergence in an ecogenomic model species, the three-spined stickleback. Our data set contains 60 fully sequenced genomes sampled across five replicate parapatric lake-river population pairs varying in their degree of genetic differentiation. We find that genomic regions of exceptional differentiation are generally different for each population pair, while their size and abundance are not correlated with the extent of genome-wide population divergence. An analysis of allele frequency spectra in those regions reveals evidence for different processes, namely hitchhiking around positively selected variants, background selection, and a local reduction of gene flow, shaping individual regions of exceptional differentiation. We provide empirical evidence that alternative mechanisms determining the evolution of genomic patterns of divergence are not mutually exclusive, but rather act in concert to shape the genome during population differentiation, a first step towards ecological speciation.

10.5

Speciation in cichlid fishes of Lake Malawi

Milan Malinsky [1,2], Richard J. Challis[3], Alexandra Tyers[3], Eric A. Miska[2], Martin J. Genner[4], Richard Durbin[1], George F. Turner[3]

[1] Wellcome Trust Sanger Institute, Cambridge, UK, [2] Gurdon Institute, Cambridge, UK, [3] University of Bangor, Bangor, Wales, UK, [4] University of Bristol, Bristol, UK

Understanding the causes and consequences of speciation, including at the genetic level, requires investigation of multiple taxon pairs at different stages on the speciation continuum. The Lake Malawi cichlid radiation is one of the largest adaptive radiations on earth. It has generated over 500 species in less than five million years, involving divergence in habitat, feeding apparatus, nuptial colour and many other traits, thus presenting an opportunity to observe hundreds of varied, recent, and in some cases ongoing, speciation events. However, investigation of the early stages of speciation in the large cichlid radiations is hampered by the difficulty in estimating sister species relationships, in reconstructing past geographical situations, and in accommodating introgression among non-sister taxa. We have obtained high quality whole genome sequences for: a) 117 individuals from over 50 representative Lake Malawi cichlid species, in order to understand the genetic structure, assess the extent of introgression and reconstruct a phylogenetic tree/network, thus clarifying the relationships between the species and providing a solid basis for research on speciation and adaptation in the system. b) 150 individuals of the Astatotilapia calliptera lineage that are the subject of more recent radiation in six crater lakes just north of Lake Malawi. These include cases of within-lake adaptive divergence, providing convincing examples of lineage diversification driven by selection, without significant geographic isolation. Together with ecological and morphological data, the genomes paint a picture of selection on specific traits such as jaw morphology driving the emergence of divergent genomic islands in the process of speciation.

10.6

Deserts of diversity in the Human-Chimpanzee ancestor links the X chromosome to speciation.

Julien Dutheil [1,3], Kasper Munch[2], Kiwoong Nam[2], Thomas Mailund[2], Mikkel Schierup[2]
[1] Max Planck Institute for Evolutionary Biology, Plön, Germany, [2] Bioinformatics Research Center, Aarhus, Denmark, [3] Institut des Sciences de l'Evolution - Montpellier, Montpellier, France

The human-chimpanzee divergence is lower on the X chromosome than expected based on autosomal divergence and an effective population size of 3/4 that of the autosomes. This has led to a controversial hypothesis, invoking a secondary hybridization event and proposing a unique role of the X chromosome in human-chimpanzee speciation. We study a complete genome alignment of human, chimpanzee, gorilla and orangutan and infer patterns of incomplete lineage sorting (ILS) between these species using a coalescent hidden Markov model. The amount of ILS depends on the ancestral effective population size and is therefore a proxy for ancestral diversity. We show that the previously inferred low divergence of the X chromosome is entirely due to megabase-sized regions comprising one-third of the X chromosome.
We demonstrate that background selection can only explain 10% of this reduction at most. We performed simulations to assess the effect of selective sweeps in the ancestral species and show that recurrent selective sweeps need to be invoked to explain such large regions deprived of ILS. In agreement with this hypothesis, we report evidence of population specific diversity reduction in extant humans and Great Apes that overlap with the regions of low diversity in the ancestral species. These regions also significantly overlap with regions devoid of Neanderthal introgression into modern humans. We propose that these X-linked regions are involved in the formation of reproductive barriers and therefore played a role in the speciation of human and chimpanzee, and possibly other Great Ape species.

10.7

Disrupted imprinting, the large X-effect, and extreme hybrid overgrowth in hamsters

Thomas D Brekke, Colin Callahan, Jeffrey M Good
University of Montana, Missoula, MT, USA

Extreme abnormal growth is a recurrent phenotype in mammalian hybrids, indicating that disruption of development may play an important role in mammalian speciation. Disrupted genomic imprinting - the parent-specific epigenetic silencing of one allele - resulting in dosage imbalances between growth factors and repressors has been hypothesized to be the predominant cause of abnormal hybrid growth in mammals. However, genetic imprinting may also expose hybrid incompatibilities through the hemizygous expression of genes. Here we combined genome-wide transcriptomic and quantitative genetic experiments to dissect the genetic underpinnings of extreme overgrowth that manifests in hybrids between two closely related species of dwarf hamsters. Specifically we tested for disrupted imprinting and differences in the expression of growth-related genes in overgrown hybrid placenta. We found that large hybrids show evidence for disrupted paternal imprinting and differential expression of imprinted genes in general. Surprisingly, the disruption of paternal imprinting is associated with reduced gene expression. As paternally imprinted genes tend to repress offspring growth, these data suggest that overgrowth is associated with a reduction in growth repressors rather than an excess of growth factors. Using QTL mapping, we have further identified the X chromosome as the predominant factor explaining hybrid placental overgrowth. However, expression and imprinting status of the X chromosome is not significantly disrupted, indicating that X-linked genetic incompatibilities are not caused by chromosome-wide misregulation. Collectively, our data underscore a central role for both disrupted epigenetic processes and the X chromosome in driving the evolution of abnormal hybrid growth in mammals.

10.8

From sequences to molecular phenotypes: a simple model of transcription factor-DNA binding that predicts a higher rate of speciation in small populations.

Bhavin Khatri [1], Richard Goldstein[2]

[1] Mathematical Biology Division, National Institute for Medical Research, London, UK, [2] Division of Infection and Immunity, University College London, London, UK

Speciation is fundamental to the huge diversity of life on Earth. Yet the underlying genetic mechanisms that lead to reproductive isolation are still poorly understood. Increasing advances in genomic technologies have allowed more detailed sequencing and understanding of the specific interactions between loci, for example, using ChipSeq analysis of transcription factor binding sites. However, there is currently a lack of theory that can bridge the potential data from such cross-species studies and the actual process of speciation. In this work, we examine a biophysical model of allopatric speciation based on transcription factor-DNA binding. This allows study of a simple molecular basis of hybrid incompatibilities with predictions that can be tested using ChipSeq analysis. Both our theory and simulations give a new prediction for the monomorphic regime of evolution, consistent with data, that smaller populations should develop incompatibilities more quickly. This arises as: 1) smaller populations have a greater initial drift load, as there are more sequences that bind poorly than well, so fewer substitutions are needed to reach incompatible regions of phenotype space; 2) slower divergence when the population size is larger than the inverse of discrete differences in fitness. The biophysical model thus represents a robust mechanism of rapid reproductive isolation for small populations that does not require peak-shifts or positive selection. In particular, our results directly predict how hybrid binding energies change with divergence time, which we suggest can be directly tested using ChipSeq analysis of transcription factor intra- and inter-species genomic binding.

10.9

Reticulate speciation and adaptive introgression in the *Anopheles gambiae* species complex

Jacob E. Crawford [1], Michelle M. Riehle[4], Wamdaogo M. Guelbeogo[5], Awa Gneme[5], N'Fale Sagnon[5], Kenneth D. Vernick[3], Rasmus Nielsen[1], Brian Lazzaro[2]
[1] University of California, Berkeley, Berkeley, CA, USA, [2] Cornell University, Ithaca, NY, USA, [3] Institut Pasteur, Paris, France, [4] University of Minnesota, St. Paul, MN, USA, [5] Centre National de Recherche et de Formation sur le Paludisme, Ougadougou, Burkina Faso

The *Anopheles gambiae* species complex is composed of ecologically specialized subgroups, including several primary human malaria vectors in sub-Saharan Africa, that exhibit varying degrees of reproductive isolation. Comparisons across the speciation continuum within this complex afford a valuable opportunity to understand the nature of speciation and species boundaries in this system. We sequenced 32 complete genomes from field-captured individuals of *Anopheles gambiae*, *Anopheles gambiae* M form (recently named *A. coluzzii*), the recently discovered "GOUNDRY" subgroup of *A. gambiae*, and *A. arabiensis*. We use a combination of existing and novel approaches to show autosomal introgression has recently occurred among taxa along the speciation continuum, in some case facilitating adaptation, despite a genomic backdrop of strong reproductive isolation and adaptive differentiation. The X chromosome, however, is strongly differentiated among all species and subgroups, pointing to a disproportionately large effect of X chromosome genes in driving speciation among anophelines. We also find a large region of the X chromosome that has swept to fixation in the GOUNDRY subgroup within the last 100 years, which may be an inversion that serves as a partial barrier to contemporary gene flow. We show that speciation with gene flow results in genomic mosaicism of divergence and introgression, sometimes leading to striking contrasts in divergence along the genome. Our results point to a model of diversification that leads to a reticulate gene pool connecting vector species and subgroups across the speciation continuum with important implications for malaria control efforts.

10.10

Insights into speciation and introgression from 765 genomes of the main malaria vectors in Sub-Saharan Africa

Giordano Botta [1], Alistair Miles 2, Beniamino Caputo [1], Alessandra della Torre [1], Dominic Kwiatkowski [2-3] and the Anopheles gambiae 1000 Genomes Consortium [4]
[1] Department of Public health and Infectious Diseases, University of Rome "La Sapienza", Rome, Italy, [2] Centre for Genomics and Global health, The Wellcome Trust Centre for Human genetics, University of Oxford, Oxford, UK, [3] Wellcome Trust Sanger Institute, Hinxton, UK, [4] Affiliations listed at http://www.malariagen.net/projects/vector/ag1000g

The *Anopheles gambiae* species complex is known to exhibit highly dynamic patterns of reproductive isolation and introgression, and is a model system for the study of evolutionary and speciation genomics. *An. gambiae* and *An. coluzzii* are also the primary vectors of human malaria in Africa, thus their evolutionary history are inextricably linked to human populations. Furthermore, the recent scale-up of malaria control interventions created strong selective pressures that has driven the spread and introgression of insecticide resistance mutations in contemporary populations.

The *Anopheles gambiae* 1000 genomes project (Ag1000G) is a consortial project involving 13 institutions which aims to provide the foundation for a new era of research into the evolutionary and population genomics of malaria vectors, as well as improved methods for their control. In phase 1 of the Ag1000G project, whole genome deep sequence data has been generated for 765 *An. gambiae* and *An. coluzzii*, collected from 8 countries spanning sub-Saharan Africa.

We observe an astonishing level of nucleotide diversity, an average of 1 SNP every ~5 base pairs. There is clear evidence for population structure correlating both with known species boundaries and major geographical features. The genomic landscape of divergence between species seems to have been shaped by different patterns of selection, gene flow and ancestral polymorphisms. Mutations conferring insecticide resistance have freely crossed species and geographic barriers, providing insights into mechanisms of introgression. Site frequency spectra and coalescent-based methods recover contrasting demographic histories between populations. We discuss open questions, including links between mosquito and human demographic histories.

10.11

Same, same, but different: Quantitatively similar but qualitatively divergent paths to extremophile speciation in H2S tolerant fish

Markus Pfenninger [1], Simit Patel[1], Manfred Schartl[3], Lenin Arias–Rodriguez[4], Martin Plath[2]
[1] Biodiversity and Climate Research Centre, Frankfurt, Hessen, Germany, [2] Northwest A&F University, Yangling, Shaanxi, China, [3] Universität Würzburg, Würzburg, Bayern, Germany, [4] Universidad Juarez Autonoma de Tabasco, Villahermosa, Tabasco, Mexico

Populations that diverge by repeatedly adapting to the same environmental stressor offer a unique opportunity to study ecological speciation. Several populations of the Poecilia mexicana complex have independently colonized hydrogen sulphide (H2S) containing springs, even though it is a potent respiration poison for the species. We analysed the genome-wide traces of adaptation of two clearwater/H2S population pairs on the way to ecological speciation in the face of gene-flow. Using population genomic approaches, we find that much more identical SNPs and genes are involved in the parallel adaptation to the hostile environment in both population pairs than expected by chance. However, the presence of a mitochondrial key adaptation in one population determines a dramatic difference in level of divergence at adaptive loci among the respective populations. Our study thus demonstrates that even closely related lineages follow both parallel and disparate molecular evolutionary paths to ecological speciation in response to the same selection pressure.

10.12

Genetics and molecular mechanisms affecting polyploidy-dependent speciation

Luca Comai, Isabelle Henry
Department of Plant Biology and Genome Center, University of California at Davis, Davis, California, USA

Genome-wide duplication events are ubiquitous in the angiosperm evolutionary tree and might have contributed to diversification. We study incipient and recent polyploidy in Arabidopsis to understand the molecular basis of successful hybridization, establishment, and adaptation of polyploids. The considerable regulatory disruption, sterility, and genome instability observed in neo-allopolyploids must be reconciled with the success and apparent stability of natural allopolyploids. We study molecular and genetic mechanisms using the Arabidopsis model system, as well as other species, when suitable. We will present information emerging from this ongoing analyses and discuss possible mechanisms involved in genome stabilization and regulatory innovation, which could contribute to the evolutionary success of polyploids.

10.13

Recent allopolyploid origins and the early stages of diploidization in the highly successful tetraploid *Capsella bursa-pastoris*

Kim A. Steige [1], Gavin M. Douglas[2], Gesseca Gos[2], Adriana Salcedo[2], Karl Holm[1], Emily B. Josephs[2], Ramesh Arunkumar[2], J. Arvid Ågren[2], Khaled M. Hazzouri[2,3], Wei Wang[2], Adrian E. Platts[4], Robert J. Williamson[2], Barbara Neuffer[5], Douglas G. Scofield[1], Martin Lascoux[1], Stephen I. Wright[2], Tanja Slotte[1,6]

[1] Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden, [2] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, [3] Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates, [4] McGill Centre for Bioinformatics, McGill University, Montreal, Canada, [5] Department of Botany, University of Osnabrück, Osnabrück, Germany, [6] Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden

Whole genome duplication and hybridization are key factors for plant speciation, and can have major effects on genome structure and gene expression evolution. There is growing evidence for epigenetic changes following polyploidization in plants, but the relative importance of hybridization and polyploidy for rapid changes in gene expression is not well characterized. Here we conduct whole genome and transcriptome sequencing in the widespread tetraploid weed *Capsella bursa-pastoris*. Whole genome data provide strong support for recent hybrid origins of *C. bursa-pastoris* within the last 100-300,000 years from two diploid progenitors in the *Capsella* genus. We find major-effect inactivating mutations were mainly inherited from the parental species and we observe a decrease in the efficacy of natural selection genome-wide, due to the combined effects of demography, selfing, and genome redundancy from whole genome duplication. We further test to what extent duplicate gene expression in *C. bursa-pastoris* is mainly affected by hybridization vs genome duplication by contrasting allele-specific gene expression in diploid hybrids of the two parental lineages of *C. bursa-pastoris* with homeolog-specific expression in the allopolyploid. Our results suggest that the earliest stages of diploidization are associated with quantitative genome-wide decreases in the strength and efficacy of selection rather than rapid gene loss, and that evolutionary processes in this recent polyploid are greatly affected by the mating system and evolutionary history of the parental species.

10.14

Allopolyploidy-associated changes of genome composition in genus *Melampodium* : does evolution repeat itself?

Jamie McCann [1], Jiři Macas[2], Petr Novák[2], Tod Stuessy[1,3], José Villaseñor[4], Hanna Weiss-Schneeweiss[1]

[1] Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria, [2] Biology Centre ASCR, Institute of Plant Molecular Biology, Ceské Budejovice, Czech Republic, [3] Herbarium; Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, Ohio, USA, [4] Instituto de Biología, Universidad Nacional Autónoma de México. Departamento de Botánica, Mexico, D.F., Mexico

Allopolyploidy has played an important role throughout the evolution of the flowering plants. Genome mergers have been shown to be accompanied by often significant and rapid alterations of genome size and structure via chromosomal rearrangements and altered dynamics of tandem and dispersed repetitive DNA families. Recent developments in sequencing technologies and bioinformatic methods have allowed for a comprehensive investigation of the repetitive component of the genome. Herein, we investigate repetitive DNA dynamics in a group of closely related allopolyploids and their progenitor species in the genus *Melampodium* using the RepeatExplorer pipeline. The largest section of the genus (sect. *Melampodium*; $x = 10$) encompasses six diploid species, two of which participated in the origin of four allopolyploids ($4x$ and $6x$). While two allotetraploid species originated from different parental taxa, two allohexaploids originated via the cross of the same maternal and paternal species, a diploid and one of the allotetraploids, respectively. Significant differences in genome composition in diploid parental taxa manifested in genome size variation have been investigated using low-pass NGS. The composition of genomes of parental taxa is further used as a basis for comparative analysis of genomic repeat dynamics in polyploids. The ultimate goal of this study is to infer the patterns of genome dominance in allopolyploids with respect to their parental taxa and to test the hypothesis of evolution repeating itself in allopolyploids that share both maternal and paternal parents.

10.15

The Effects of Pollination and Range Shifts on the Diversification of the Tribe Antirrhineae

Ezgi Ogutcen, Jolan Theriault, Daniel B. King, Jana C. Vamosi
University of Calgary, Calgary, AB, Canada

Range shifts are considered an important precursor to evolutionary divergence because they place populations in different environments that favour different characters. Long-distance dispersal promotes an expansion in niche breadth in terms of pollination syndromes in angiosperms, potentially explaining a wide variety of pollination syndromes. Antirrhineae, a tribe under Plantaginaceae, is a useful group for studying the interplay between dispersal and pollination in macroevolution because it has members in the Old World and the New World, and exhibits numerous transitions in major pollinating groups.
By integrating predictive modeling and range reconstructions with phylogenetic analysis, We aim to: i) reconstruct where major range shifts have occurred within Antirrhineae; ii) determine whether range overlap increases or decreases with time since divergence within the tribe; and iii) examine whether shifts in geographic distribution and pollination syndromes are associated with differences in speciation rates in Antirrhineae.
We find that distribution or pollination alone does not affect speciation rates, but the interaction of these two characters have a significant effect on speciation rates in Antirrhineae. Our age-range correlation analysis also suggest that sympatric speciation is predominant in the tribe, with a trend towards young nodes having more range overlap than older nodes in the phylogeny.

10.16

Pollinator-mediated divergent selection drives reproductive isolation, speciation, and the evolution of barrier genes in sexually deceptive orchids

Philipp Schlüter [1], Khalid Sedeek[1], Edward Whittle[2], John Shanklin[2]
[1] *University of Zurich, Zurich, Switzerland,* [2] *Brookhaven National Lab, New York, USA*

Sexually deceptive orchids of the genus *Ophrys* are a good example of ecological speciation driven by shifts in highly specific pollinators. They achieve high pollinator specificity by chemical mimicry of the pollinator female's sex pheromone. Genes underlying pseudo-pheromone production therefore are targets of divergent selection by pollinators. In particular, stearoyl-ACP desaturase (SAD) homologues control alkene double-bond position, which constitute an important component of variation in the pseudo-pheromones, and have a large effect upon pollination. These genes appear to diverge early in speciation and are more highly differentiated between species than the majority of the genome, which displays extensive shared polymorphism among species. *SAD* genes evolved by gene duplication from a housekeeping-desaturase-like ancestor and have diverged in gene expression and protein function. Changes in protein-protein interaction and subcellular localisation may minimise epistatic interactions between housekeeping and alkene-biosynthetic desaturases, and changes in the substrate binding pocket may have altered the protein's substrate specificity so as to free it from deleterious pleiotropy associated with pollinator-mediated selection, thereby enabling rapid species divergence in this group.

375A

Bayesian species delimitation combining multiple genes and traits

Cecile Ane [1], Claudia Solis-Lemus[1], L. Lacey Knowles[2]
[1] University of Wisconsin - Madison, Madison, WI, USA, [2] University of Michigan, Ann Arbor, MI, USA

Distinguishing whether a group of subspecies forms one single species or separate species is a crucial problem for evolutionary and conservation biology. Here, we propose a Bayesian statistical method for species delimitation from genetic and/or morphological data, which we implemented in the software iBPP. Our method is based on an evolutionary framework, which uses the coalescent model for gene evolution. Morphological data is modeled with a multivariate normal distribution whose covariance matrix is determined by the species tree. For this model, we develop a new conjugate prior adapted to two levels of dependency: among measured variables and among individuals sampled from a set of putative species. I will describe the robustness gained by using multiple data types in a unified framework, when some model assumptions are violated, such as because of gene flow or phenotypic plasticity.

376B

Genome divergence during early stage speciation with gene flow

David Marques [1,2], Laurent Excoffier[1,3], Ole Seehausen[1,2]

[1] Institute of Ecology & Evolution, University of Bern, Bern, Switzerland, [2] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland, [3] Swiss Institute of Bioinformatics, Lausanne, Switzerland

The number of speciation studies using population genomics approaches has increased quickly, yet few have compared replicate events of the same process with variation in geographical opportunity for gene flow between incipient species. We studied genome-wide differentiation among threespine stickleback (Gasterosteus aculeatus) lake and stream ecotypes that evolved in the last 150 years within the Lake Constance drainage system. We tested predictions of speciation with gene flow theory by comparing genomic differentiation between lake and stream ecotypes breeding in sympatry versus in parapatry. Consistent with higher gene flow, we found fewer genomic islands of differentiation and lower genome-wide levels and heterogeneity of differentiation in the sympatric case. However, 15 genomic islands of differentiation resisting gene flow in sympatry showed parallel divergence, and are thus candidate regions for adaptive divergence among lake and stream ecotypes. Interestingly, 12 of these islands cluster in a low-recombination region on chromosome VII. Furthermore, islands showed non-random overlap with many known quantitative trait loci, most of them controlling phenotypic traits that have diverged between Lake Constance ecotypes. The genomic co-localization of parallel differentiation, trait architecture and recombination landscape suggests that both divergent natural selection and the genomic arrangement of heritable traits are important in adaptation and early stage ecological speciation with ongoing gene flow.

377C

Where morphology fails: Unravelling the true structure and diversity of the zebra using large scale sequencing

Casper-Emil Tingskov Pedersen [1], Anders Albrechtsen[1], Lounes Chikhi[2], Hans Redlef Siegismund[1], Rasmus Heller[1]

[1] University of Copenhagen, Copenhagen, Denmark, [2] Université Paul Sabatier, Toulouse, France

The plains zebra (Equus quagga) is an iconic species of the African savannah. It is usually subdivided into five subspecies according to geographic location and morphology. However, these groupings have not been resolved by mtDNA or microsatellite data and remain contentious. Furthermore, the relation between the extinct quagga and the plains zebra is controversial, with some evidence suggesting they are sister species and other evidence pointing at the quagga as an extinct subspecies of the plains zebra. We used a combination of newly generated RADseq and publicly available full genome data to examine the intra- and interspecific admixture patterns in the plains zebra.

Our results show that the currently recognized subspecies do not correspond to genetic clusters. We did find a geographical pattern of genetic clustering, but it deviated from the subspecies designations in several cases. Furthermore, our results suggest that the extinct quagga is not a sister species to the plains zebra; rather it is a derived population most closely related to southern populations of plains zebra. We found that the three zebra species are genetically distinct, however, we found a signal of ancient gene flow between the plains and the mountain zebra.
Our results show the power of GBS and genomic data to unravel subtle patterns of genetic structure. The findings are relevant for conservation and management of the plains zebra as well as for understanding the deeper evolutionary history of the zebras.

378D

Genomic Signatures of Hybrid Speciation in Invasive Sculpins (*Cottus*)

Fritz Sedlazeck[1], Jie Cheng[2], Janine Altmüller[3,4], Arndt von Haeseler[1,5], Arne Nolte [1]
[1] Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, A-1030 Vienna, Austria, [2] Department for Evolutionary Genetics, Max-Planck-Institute for Evolutionary Biology, D-24306 Plön, Germany, [3] Cologne Center for Genomics, University of Cologne, D-50931 Cologne, Germany, [4] Institute of Human Genetics, University of Cologne, D-50931 Cologne, Germany, [5] Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, A-1090 Vienna, Austria

Models of homoploid hybrid speciation predict that traits from both parental species are combined in an emerging hybrid species making it fitter than either parent. Hybridization between *Cottus rhenanus* and *Cottus perifretum* has produced an invasive hybrid species. Here, we determine which alleles from one or other parent species have reached high frequency in the hybrid population by natural selection, rather than by drift or migration. We first measured allele frequencies for 52,211 SNPs from one or other parent in invasive *Cottus* by high-throughput sequencing a pool of genomic DNA. We used a diffusion approximation to model the expected spectrum of allele frequencies across the genome under neutral drift with gene flow. 649 high frequency alleles (linked to 212 annotated genes) from *C. rhenanus* did not fit our neutral expectation, suggesting that they were targets of natural selection in the hybrid lineage. In contrast, high frequency alleles from *C. perifretum* could be explained by our neutral model, without invoking selection, presumably because the time since admixture is too long or because the overall contribution of this species to the admixed gene pool is too large. Only 12% of selected alleles are within previously identified hybrid-incompatibility QTL, suggesting either that most intrinsic incompatibilities have not yet been discovered, or that extrinsic selection is important in shaping the hybrid genome. The evolution of invasive *Cottus* involves a diverse genomic basis that presumably involves genotypes that originate from *C. perifretum* as well, but which could not be detected using our experimental method.

379A

Investigating the genetic basis of high altitude adaptation in Andean hummingbirds

Marisa Lim [1], Christopher Witt[2], Catherine Graham[1], Liliana Davalos[1]
[1] Stony Brook University, Stony Brook, NY, USA, [2] University of New Mexico, Albuquerque, NM, USA

Discovering the genetic basis for key adaptations involved in species diversification and distribution is a fundamental challenge in evolutionary biology. We investigate convergent evolution to high altitude adaptation in hummingbirds (Trochilidae), a diverse family that occurs from sea level to 5000 meters in elevation. Species from four of the nine monophyletic hummingbird clades independently colonized high altitude in the South American Andes Mountains ~10 Mya resulting in relatively recent adaptation to high elevation. Under the hypoxic conditions at high altitude, we predict that 1) limited routes to adaptation to high altitude leads to parallel evolution and 2) increased metabolic costs affect pathways involved in energy expenditure and transport of glucose and oxygen (e.g., hypoxia-inducible factor and fatty-acid metabolism pathways). To test these hypotheses, we sequenced transcriptomes for 12 species that span the phylogeny and occur at high or low altitude using RNA-seq. Transcriptomes were aligned and annotated with genomes from the zebra finch (Taeniopygia guttata) and Anna's hummingbird (Calypte anna) to identify orthologous genes. We compiled a panel of candidate genes from those under positive selection, including HBB (beta-globin production), MT-CO3 (oxidative phosphorylation), HIF1AN (oxygen sensor), and ACOX3 (fatty-acid metabolism), and compared them across our study species. HBB shows signals of both positive selection and convergence across high altitude lineages. This work will facilitate future research on convergent genetic adaptation to high altitude in hummingbirds and the tree of life.

380B

Selection Vs. drift in bottlenecked populations along an environmental gradient

Daniel Jeffries [1], Lori Lawson-Handley[1], Gordon Copp[2], Bernd Hänfling[1]
[1] *University of Hull, Hull, UK,* [2] *CEFAS, Lowestoft, Suffolk, UK*

Effective species conservation is underpinned by phylogeographic knowledge, and to this end, we produced a phylogeography of the threatened European freshwater fish, the crucian carp (*Carassius carassius*). For this we used mitochondrial DNA, microsatellites and approx. 20,000 SNP loci (RAD-seq), and compared the ability of these data types to resolve spatial genetic structure. We therefore add perspectives from real biological data to the question "are more samples or more markers better?" Something that, until now, has mostly been addressed with simulated data.

The genome-wide SNP dataset also allowed us to re-examine a fundamental hypothesis in population and speciation genetics, which is that low effective population size results in inefficient selection, due to overriding levels of genetic drift. Crucian carp populations are prone to genetic bottlenecks, therefore, we addressed this hypothesis by identifying loci under selection across a temperature gradient within our sample sites, and comparing signatures of selection, across this gradient, in non-bottlenecked and bottlenecked populations.

381C

Functional organization of the genome and speciation in the house mouse

Václav Janoušek [1,2], Pavel Munclinger[1], Liuyang Wang[3,4], Katherine C. Teeter[5], Priscilla K. Tucker[3]

[1] Department of Zoology, Faculty of Science, Charles University in Prague, Prague, Czech Republic, [2] Institute of Vertebrate Biology, ASCR, Brno, Czech Republic, [3] Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan, USA, [4] Department of Molecular Genetics and Microbiology, School of Medicine, Duke University, Durham, North Carolina, USA, [5] Department of Biology, Northern Michigan University, Marquette, Michigan, USA

Despite considerable progress in understanding how living organisms evolve, the process of their formation still remains unresolved. Some recent studies have suggested that functional architecture of the genome might be an important factor in species divergence. However, none of these studies explored the importance of genome organization in the evolution of reproductive isolation. To address this issue, we use data consisting of 1316 SNP markers genotyped in three hybrid populations of the house mouse (*Mus musculus musculus* X *M. m. domesticus*) in combination with publicly available mouse genomic resources. Partial reproductive isolation between these two subspecies evolved mainly in allopatry with intrinsic isolation being a major contributor to the total barrier. Our goal was to explore genomic features associated with various levels of gene flow across the subspecies boundary where these two subspecies overlap and interbreed. While rate of recombination and differentiation have been previously suggested to play a role in species divergence, we discovered that functional organisation of the genome may be an important feature driving the evolution of reproductive isolation. Specifically, we determined that various rates of gene flow correlate with the location of gene functioning in the cell. Genomic regions containing genes functioning on the cell periphery are more likely to cross the subspecies boundary, whereas genomic regions enriched for genes functioning intracellularly tend to be selected against in hybrid mice. These findings suggest that functional organization of the genome may play a role in species divergence and more specifically in the evolution of reproductive isolation.

382D

Genetic differentiation in the widely distributed species *Arabidopsis thaliana* : A first step in speciation?

Dounia Saleh, Karl J. Schmid
Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany

Geographic separation of disjunct populations in combination with local adaptation likely explains numerous speciation events in plants. One possible speciation mechanism is limited gene flow between allopatric populations leading to the evolution of genetic incompatibilities among rapidly evolving genes involved in reproductive functions, host-pathogen interaction and other aspects of local adaptation.
We investigated whether such speciation processes occur among populations of the widely distributed species *Arabidopsis thaliana* based on the polymorphism data of 80 Eurasian accessions from the 1001 genomes project. Multivariate analyses of population structure revealed two genetic clusters that separate the accessions along a West-East gradient. The two clusters represent the deepest evolutionary split within A. thaliana and likely are caused by a history of glacial refugia. Using F_{ST} and related statistics we estimated genetic differentiation of all SNPs between the two clusters and tested whether certain groups of genes defined by gene expression patterns and regulatory or metabolic pathways are overrepresented. Additionally, we compared topologies of individual gene trees with the expected structure under a population model with two clusters. Our results showed no strong evidence for a similar differentiation of particular gene classes but allowed us to identify several candidate genes for reproductive isolation. We conclude that despite the high level of self-fertilization in *A. thaliana*, there is sufficient recombination and gene flow occurring to counteract, at least partially, the evolution of genetic incompatibilities between the two major genetic clusters in this species and raises some questions regarding the required conditions for speciation mechanisms.

383A

Population genomics of the hybrid, asexual, tropical root-knot nematodes (Meloidogyne)

Amir Szitenberg [1], Laura Salazar Jaramillo[2], Dominik Laetsch[2], Mark Blaxter[2], David Lunt[1]
[1] *University of Hull, Hull, UK,* [2] *University of Edinburgh, Edinburgh, UK*

Meloidogyne is a genus of plant parasitic nematodes causing ~5% loss to world agriculture per annum. The most aggressive species are obligatory asexual apomicts whose phylogenetic relationships and species boundaries have been confounded by both recent origin and a complex hybridization history. Our recent genome analysis shows that M. incognita is a double hybrid between the hybrid species M. floridensis, and an unknown third species. The origin of the other tropical apomicts is less clear. Here we have taken a population genomics approach to study the species and population structure within the apomict group. We sample both across the species complex, and globally within species, using whole genome sequence data to elucidate parental species origins and species boundaries. Using phylogenomics we determine that the key agricultural species M. javanica and M. incognita are both double hybrids sharing the same hybrid and non-hybrid parents.

384B

Using rigorous demographic modeling to interpret patterns of genomic divergence in a recent avian radiation

Ilan Gronau [1], Leonardo Campagna[2], Adam Siepel[3], Irby Lovette[2]
[1] Herzliya Interdisciplinary Center (IDC), Herzliya, Israel, [2] Cornell University, Ithaca, NY, USA, [3] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Our ability to utilize high throughput sequencing technologies in the study of speciation relies on the adaptation of computational methods developed for model organisms to non-model species commonly used in speciation studies. We demonstrate this through a study of the recent radiation of southern capuchino seedeaters from the genus *Sporophila*. We used RAD-seq to obtain large scale sequence data for sixty individuals from six *Sporophila* species and explored patterns of genomic divergence in this recent radiation. A simple examination of F_{ST} values across loci indicates that a small minority of loci are differentiated and that divergence in some of these might have been driven by natural selection. Suspecting that neutral processes can generate similar signatures, we inferred a detailed demographic history for these species using a Bayesian inference method called G-PhoCS. G-PhoCS was originally developed for whole-genome sequence data, but since it bases its inference on patterns of variation in thousands of unlinked loci, it is also well suited for data generated from reduced representation approaches such as RAD-seq or GBS. Our inference suggests a complex process of speciation involving a very large ancestral N_e for five of the six species, followed by a rapid series of divergence events and subsequent gene flow. We show that the dramatic changes in N_e and gene flow explain the large variance in divergence levels observed across loci. We conclude by suggesting ways of using rigorous demographic modelling to produce null models for finding islands of divergence in the genomes of recently diverged species.

385C

Chromosomal background of reproductive isolation in a nascent species pair of Lake whitefish

Anne-Marie Dion-Côté[1], Radka Symonova [2], Fabien Lamaze[3], Petr Rab[4], Louis Bernatchez[1]
[1] Department of Biology, Institut de Biologie et des Systèmes (IBIS), University Laval, Quebec City, Canada, [2] Research Institute for Limnology, University of Innsbruck, Mondsee, Austria, [3] Laval University Cancer Research Center, Quebec City, Canada, [4] Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Libechov, Czech Republic

Salmonids are interesting model organisms for speciation cytogenomics: they are pseudo-tetraploids, which, along with small population sizes, is expected to facilitate fixation of chromosomal rearrangements. In northeastern North America, two Lake whitefish lineages have colonized post-glacial lakes ~12,000 YBP following ~60,000 years of allopatry. A dwarf limnetic form has evolved repeatedly since then from the normal benthic form. Evidence suggests that the accumulation of genetic and chromosomal incompatibilities in allopatry might have facilitated this divergence.
We screened for chromosomal instability in hybrids between nascent species (dwarf and normal) and examined chromosomes in pure embryos, and healthy and malformed (unfit) backcross embryos with the emphasis on the ploidy level. Further, we applied cytogenetic methods to three natural Lake whitefish sympatric pairs, focusing on more detailed chromosomal features. Our results showed evidence of mitotic instability through an increased variance in intra-individual chromosome numbers in healthy backcrosses. In malformed backcrosses, extensive aneuploidy corresponding to multiples of the haploid number (n=40, 2n=80, 3n=120) was found, indicating meiotic breakdown in their F1 parent. Karyotypes of parental forms did not show obvious differences at the whole-chromosome level, whereas more detailed cytogenetic analyses revealed an extensive polymorphism mostly shared by the three species pairs. Our results indicate ongoing genomic rearrangements between these incipient species, consistent with chromosomal divergence in allopatry followed by gene flow in sympatry. As such, cytogenetics is highly complementary to genomic approaches and should not be neglected towards a full understanding of the genomic bases of speciation.

386D

Genome scans identify the genomic architecture of sympatric speciation in cichlids

Christoph Hahn, Domino Joyce
University of Hull, Kingston upon Hull, UK

The East African lake Malawi harbors over 700 endemic species of haplochromine cichlid fish, representing one of the largest and most diverse adaptive radiations on earth. As such, it is an important model system with which to understand ecological diversification, sexual selection and speciation. We present a genome-wide SNP data set obtained using a RAD-sequencing approach. It comprises more than 250 individuals representing the taxonomic diversity in the Lake Malawi cichlid radiation and allows for unprecedented resolution of the phylogenetic relationships within the flock. We focus further in-depth analyses on four sympatric populations of Diplotaxodon species from an endemic deep water genus. Previous studies have found significant genetic differentiation between these populations and revealed phenotypic differentiation in head morphological traits. We apply and compare Fst outlier scan approaches to detect signals of divergent selection between the sympatric species. We find 59 outlier candidate regions, and gene ontology enrichment analyses of the functional classes of genes in these regions strongly supports the hypothesis that the machinery of transcriptional regulation plays a central role in species differentiation. More detailed analyses of the genes likely involved in the process reveal e.g. transcription factors for ecologically relevant phenotypic traits such as head morphology.

387A

Identification of the olfactory receptor genes involved in male sex pheromone perception in the butterfly *Bicyclus anynana* : a step forward for understanding the evolution of chemical communication.

Veronique Baumlé [1], Nicolas Montagné[2], Emmanuelle Jacquin-Joly[2], Caroline Nieberding[1]
[1] Evolutionary Ecology and Genetics Group, University Catholic Louvain, Louvain-la-Neuve, Belgium, [2] Physiologie de l'Insecte : Signalisation et Communication, INRA, UMR-A 1272, Versailles, France

Olfactory communication using sex pheromones is important in mate choice, species recognition and sexual selection. The composition of the sex pheromone has been identified for hundreds of moth species, and Lepidoptera have thus become the main taxonomic group for studying the mechanistic basis, and the evolution, of sex pheromone communication. Stabilizing selection supposedly maintains the "lock-and-key" specificity of the interaction between sex pheromone components (the keys) on the one hand, and olfactory receptors and associated proteins of the perception machinery (the locks) on the other hand. Yet, the huge diversity of chemicals forming the sex pheromone of closely related species in Lepidoptera, including the *Bicyclus* genus (Bacquet *et al.*, in press), leads us to question that stabilizing selection is the main selective force acting on the evolution of sex pheromone communication in Lepidoptera. This project aimed at identifying the genes involved in male sex pheromone reception in the model butterfly B. anynana and assessed how sex pheromone reception evolved across closely related species of the Bicyclus genus. For this, we used the transcriptome recently produced for B. anynana (in preparation) and selected 6 candidate olfactory receptor genes. First, we identified their full length coding sequence and determined their intra- and inter-species sequence polymorphism. Secondly, we tested their odorant binding specificity by using a heterologous expression system in Drosophila coupled with electroantennography assay in order to identify the specific OR genes for the three male sex pheromone components.

388B

Detecting selection with haplotype-based methods: benchmarking and application to tropical butterflies

Angeles de Cara [1], Annabel Whibley[2], Marianne Elias[3], Mathieu Joron[3,4], Frederic Austerlitz[1]
[1] UMR 7206 CNRS/MNHN/UPD, Paris, France, [2] John Innes Centre, Norwich, UK, [3] UMR 7205 CNRS/MNHN/UPMC/EPHE, Paris, France, [4] CEFE, Montpellier, France

The vast amount of genome-wide polymorphism data available has led to considerable efforts to develop methods to detect natural selection at the molecular level. Finding regions under selection is one of the first steps towards understanding the processes of adaptation and speciation. Our ability to detect selection depends critically on the data available and on the robustness of the methods to the underlying assumptions. Several commonly used methods look for FST outlier loci. However, these methods sometimes fail to identify loci under weak selection. Conversely, some neutral markers can be inferred to be under selection. Alternatively, we can use haplotype-based methods to infer selection within populations. These methods rely on the idea that positive selection on a position in the genome will create a region of extended homozygosity. We study here the efficiency of three such methods (iHS, nSL and H12) in simulated data obtained by performing artificial selection on a polygenic trait. We show that these methods work mainly when selection is strong and the traits are only mildly polygenic. Furthermore, we analyse sets of individuals of the tropical butterflies *Heliconius* and *Ithomiini* of different species and morphs, to test the power of these methods on known regions under selection and to infer new candidates of selection.

389C

Zinc accumulation and transcriptional response in the allopolyploid Arabidopsis kamchatica

Timothy Paape[1], Masaomi Hatakeyama[1], Jun Sese[2], Teo Cereghetti[1], Rie Shimizu-*Inatsugi*[1], Kentaro Shimizu[1]

[1] *Institute of Evolutionary Biology and Environmental Studies and Institute of Plant Biology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland,* [2] *Computational Biology Research Center, AIST, Japan*

In allopolyploid species, ancestral gene copies can result in the retention of phenotypes that were present in either of the diploid parents. Heavy metal hyperaccumulation in plants represents a highly tractable quantitative trait that has tremendous potential for detailed analyses of gene regulation between diploid and polyploid relatives. While most studies of additive or non-additive expression following polyploid hybridization have analyzed differential expression between diploid orthologs and homeologs from base line expression, heavy metal treatments allow for quantifiable differential responses of gene expression among species and hybrids. *Arabidopsis kamchatica* is an allopolyploid species derived from the metal hyperaccumulator *A. halleri* and non-metal accumulator *A. lyrata* diploid species. We have developed a bioinformatics pipeline to separate homeologous RNAseq reads and have demonstrated that zinc hyperaccumulation in *A. kamchatica* appears to be the result of gene regulation and polymorphism derived from the diploid hyperaccumulating ancestor A. halleri despite hybridization with a non-accumulating *A. lyrata* ancestor. Several candidate metal homeostasis genes show strong expression bias favoring the *A. halleri* derived copies. Specifically, the zinc transporter ATPase gene HMA4 shows 15 – 70 fold increased expression in *A. halleri* vs. *A. lyrata* derived copies in the allopolyploid and a large genomic region (290 kb) surrounding this locus shows significantly different patterns of sequence diversity consistent with relaxed selective constraint on the *A. lyrata* derived homeologs and strong constraint or hitch-hiking on the *A. halleri* derived HMA4 region.

390D

Genomic despeciation between two invasive Asian carps, bighead carp (Hypophthalmichthys mobilis) and silver carp (H. molitrix)

Guoqing Lu [1], Jun Wang[1,2], Lifeng Zhu[1,3], Sarah Gaughan[1], Chenghui Wang[2]
[1] University of Nebraska at Omaha, Omaha, NE, USA, [2] Shanghai Ocean University, Shanghai, China, [3] Nanjing Normal University, Nanjing, Jiangsu, China

Interspecific hybridization may diversify species through speciation; on the other hand, it can fuse parental species through the breakdown of barriers to gene flow. Two Asian carps, bighead carp *(Hypophthalmichthys mobilis*) and silver carp (*H. molitrix*) are reproductive isolated within their native ranges; however, when introduced to the United States, the two invasive species exhibit extensive hybridization. We address this hybrid swam issue through comparative genomic and transcriptomic analysis. We conducted *de novo* genome assembly using approximately 80 Gb Illumina short reads and 10 Gb PacBio long reads. The assembled genome was estimated 1.00 and 0.97 Gb for bighead carp and silver carp, respectively, with 53.5% and 35.1% repeats. Around 60 Gb Illumina reads from two F1 hybrids were mapped to the assembled genomes and a number of genomic variation patterns and genetic inheritance profiles (paternal or maternal) were identified. Comparatively analysis of the genomes detected numerous gene expansion and contraction events among the fish lineage. The transcriptomic analysis of pure and hybrid Asian carps resulted in over 40,000 transcripts and more than 23,000 annotated proteins for each species. The parental origin analysis of F_1 hybrid sequences demonstrated a bias toward the expression of silver carp genes. The high genomic/transcriptomic similarity between backcrosses and parental species and the degree of similarity increasing with successive generations, along with the above genomic and trnascriptomic findings, strongly support our hypothesis that bighead carp and silver carp, formed 3 MYA in their native rivers, are under despeciation in novel invasive environments.

391A

Coupling genomics with experiments to study divergence-with-gene-flow in trees

Luisa Bresadola [1], Kai N. Stoelting[1], Céline Caseys[1,2], Dorothea Lindtke[1,3], Christian Lexer[1]
[1] University of Fribourg, Fribourg, Switzerland, [2] University of British Columbia, Vancouver, BC, Canada, [3] University of Sheffield, Sheffield, UK

Rapid recent progress in ecological and evolutionary genomics is imparting fresh perspectives to the study of speciation, i.e. the origin and maintenance of biological diversity. A particularly active field of research at the current time is the study of "divergence-with-gene-flow", that is, divergence that involves episodes of sym- or parapatry and thus genetic contact during some stage of the process, before reproductive isolation is complete. In this project, we address key questions regarding the ecological and evolutionary genomics of "divergence-with-gene-flow" in *Populus alba* and *P. tremula*, two widespread Eurasian tree species related to *Populus trichocarpa*, the first completely sequenced forest tree. In particular, defined key questions and hypotheses concerning the following topics are addressed: (1) the roles of early vs. late-acting reproductive barriers in the maintenance of species boundaries, (2) the genomic architecture and selective value of species differences maintained in the face of gene flow, (3) the role of genetic incompatibilities as early post-mating barriers in species isolation. We address these topics with the help of high-throughput "genotyping-by-sequencing" approaches in natural and experimental populations, and by coupling genomics with evolutionary ecology experiments. The results are expected to advance our understanding of the origin and maintenance of reproductive barriers in hybridizing tree species representing keystone or foundation species in terrestrial habitats.

392B

Hominidae-specific coding and conserved Non-coding sequences

Morteza Mahmoudisaber [1], Naruya Saitou[1,2]

[1] University of Tokyo, Tokyo, Japan, [2] Graduate University for Advanced Studies (SOKENDAI), Mishima, Japan

Family Hominidae, which includes humans and great apes, is recognized for unique complex social behavior and intellectual abilities. Despite the increasing genome data, however, the genomic origin of its phenotypic uniqueness has remained elusive. Clade-specific coding and conserved non-coding sequences (CNSs) are among the high-potential evolutionary candidates being involved in driving clade-specific characters and phenotypes. On this premise we analyzed various data retrieved from major DNA databases to find Hominidae-specifc (HS) genes and conducted whole-genome analysis on Hominidae family members' CNSs along with those of gibbon, rhesus macaque and marmoset to identify CNSs restricted to Hominidae clade members. We found one HS protein coding and four HS ncRNAs. The evolution of HS protein-coding gene, DSCR4, placed in Down syndrome critical region of chromosome 21 with no structure homology to any known protein, was shown to have happened in several steps through LTR/ERV1, LTR/ERVL retrotransposition and transversion. Using genomic distance as threshold for neutral evolution, we found 1852 HCS CNSs. Polymorphism coverage and derived allele frequency analysis of HS CNSs showed that these CNSs are under purifying selection, indicating that they may harbor important functions. HS CNSs are overrepresented in promoters and untranslated regions often flanked by genes involved in developmental process and sensory perception of sound; they also showed a significantly lower nucleosome occupancy probability. Low HS coding gene to HS regulatory CNS ratio suggests regulation alteration of existing protein-coding genes to have played a more significant role in Hominidae evolution than emergence of novel genes.

393C

Population genomics and ecological diversification of crossbills (*Loxia curvirostra*) in the Iberian Peninsula

Ainhoa Agorreta, Diego San Mauro
Complutense University of Madrid, Madrid, Spain

Finches and relatives have long been the focus of evolutionary studies, and include the famous island radiations such as Darwin's finches of the Galapagos and the Hawaiian honeycreepers, in which considerable knowledge has been gained on the origin of species. We are studying a group of finch-like passerines, the crossbills (*Loxia curvirostra*) to investigate patterns and processes generating and maintaining diversity in populations of the Iberian Peninsula. In particular, we aim to investigate whether ecological speciation has been an important driver in their diversification by testing if there is a correlation between population structure and habitat and trophic resource (pines cones). We are using next-generation sequencing (NGS) techniques to generate hundreds of thousands of single nucleotide polymorphisms (SNPs). These data are being analysed in an evolutionary context to test hypotheses on the origin and differentiation of Iberian populations and their connection with other European populations of crossbills. We are also investigating the role and relative importance of ecological adaptation and drift in generating divergence, and the association of genomic polymorphisms and particular phenotypic variations. Preliminary results indicate the presence of structure in the Iberian crossbill populations that appears associated to trophic resource. Some distinct populations are identified and interesting relationships of Iberian crossbills with other European populations are being revealed.

394D

Comparative genomics and gene-specific genotyping in mouse: new insights into reproductive isolation genes

Laia Capilla[1,2], Rosa Ana Sánchez-Guillén[1], Marta Farré[3], Jessica Alföldi[4], Kerstin Lindblad-Toh[4,7], Roberto Malinverni[5], Jacint Ventura[2], Aurora Ruiz-Herrera[1,6]

[1] Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, [2] Departament de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, [3] Department of Comparative Biomedical Sciences, The Royal Veterinary College,, London, UK, [4] Vertebrate Genome Biology. Broad Institute of MIT and Harvard, Cambridge, USA, [5] Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Barcelona, Spain, [6] Departament de Biologia Cel•lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, [7] Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

Efforts have been made in identifying genes involved in reproductive isolation (speciation genes) and in determining in which degree chromosomal rearrangements act as an impermeable barrier to gene flow, facilitating the accumulation of genetic incompatibilities that will, in the long term, contributing to speciation. In this work, we define the evolutionary genomic landscape in the mouse genome, by whole-genome comparisons of eleven mammalian species, including five rodent species. We identify novel lineage and clade-specific breakpoint regions within Rodentia and analyze their gene content. We detect an accumulation of protein-coding genes in evolutionary breakpoint regions; especially genes implicated in reproduction isolation, such as Krueppel genes, a group of transcription factors with zinc finger (ZnF) domains, among which the *Prdm9* gene is the most representative. Given the role of this gene in hybrid sterility, we study the evolutionary constrains that may affect the *Prdm9* gene across a natural population of house mice with chromosomal fusions in polymorphic state. Our results reveal an extreme allelic diversity in both ZnF copy number and sequence with the characterization of 12 different alleles in a population of 180 mice. The analysis of meiotic recombination in a subset of these mice reveals that mean recombination rates were positively correlated with a decrease in the number of ZnF domains. Overall, the presence of genes related to species-specific phenotypes (such as reproductive isolation) and low recombination rates in evolutionary breakpoint regions reinforces the adaptive value of genome reshuffling.

395A

Selection and assortative mating in hybrid swordtail fish populations

Molly Schumer [1], Mattie Squire[2], Gil Rosenthal[2], Peter Andolfatto[1]
[1] Princeton University, Princeton, NJ, USA, [2] Texas A&M University, College Station, TX, USA

Hybrid populations are a window into the mechanisms of reproductive isolation between species. Hybrid populations between the swordtail fish *Xiphophorus birchmanni* and *X. malinche* formed approximately 40 generations ago as a result of anthropogenic environmental disturbance and have experienced strong selection due to many genetic incompatibilities distinguishing the parental species. In the present study, we analyze ~500,000 ancestry informative markers throughout the genome to investigate how hybrid populations have evolved over the last 20 generations. We show that incompatibility selection has played a role in the current genomic structure of hybrid populations. Interestingly, in one hybrid population, two hybrid genotype clusters have emerged, one biased toward each parent. We show that these clusters have persisted over many generations, and are sustained by nearly 100% assortative mating. Our results highlight the joint effects of selection and behavior in structuring genomic variation in hybrid populations.

396B

The secrets of a two-billion-years marriage: Mito-nuclear coevolution affects protein-protein interactions, human health and speciation.

Dan Mishmar
Ben-Gurion University of the Negev, Beer-Sheva, Israel

Mitochondrial activity became essential for cell life since the emergence of Eukaryota, more than 2 billion years ago. This requires interactions between factors encoded by the mitochondrial DNA (mtDNA) and by the nuclear genome (nDNA). Since the mutation rate of animal mtDNA exceeds that of the nDNA by an order of magnitude, mito-nuclear coevolution maintains mitochondrial function. We argue that mito-nuclear coevolution is important for three main biological aspects: (A) Maintaining mitochondrial function and protecting the health of the individual; (B) protecting the structure and function of machineries relying on mito-nuclear physical interactions and (C) to protect the fitness within a given species in diverse environments. We hypothesize, that interfering with mito-nuclear co-evolution will affect the individual's health, disrupt physical interactions between mtDNA and nDNA-encoded factors, and may lead to speciation events. Firstly, we demonstrate that interfering with mito-nuclear coevolution leads to genotype combinations that alter the risk to develop diabetes, thus underlining the importance of mito-nuclear interactions for human health. Secondly, while mutating mtDNA and nDNA-encoded protein subunits of NADH ubiquinone oxidoreductase (complex I) that coevolve, we show that protein-protein interactions were disrupted. Finally, we found that morphologically identical vertebrates (chameleons) differentiated into two populations across an ancient marine barrier (the Jezreel Valley), that sharply diverged in mito-nuclear genotype combinations, which comprise of highly functional mutations. A mathematical model explained this finding as the emergence of hybrid breakdown - the first stage of a speciation event. Taken together mito-nuclear interactions and co-evolution affect human health and major evolutionary transitions.

397C

Molecular correlates of hybrid seed failure: exploiting variable strengths of postzygotic isolation in wild tomatoes

Morgane Roth [1,2], Ana Marcela Florez-Rueda[1,2], Margot Paris[1,2], Markos Tsinganis[1,2], Thomas Städler[1,2]

[1] Institute of Integrative Biology, ETH Zurich, Zürich, Switzerland, [2] Zurich Basel Plant Science Center, Zürich, Switzerland

Hybrid seed failure, a common postzygotic barrier in angiosperms, is likely mediated by abnormal endosperm development, highlighting the pivotal role of the endosperm for proper seed development. Epigenetic processes modulate the parent-of-origin specific expression of so-called "imprinted" genes in the endosperm, some of which may be important for resource allocation to the developing embryo. These considerations underlie the hypothesis that the normal (within-species) imprinting pattern may be disrupted in hybrid, abortive seeds. This conjecture has recently been supported through endosperm transcriptome sequencing in reciprocal crosses between the wild tomato species *S. peruvianum* and *S. chilense*, characterized by a strong postzygotic barrier. A disrupted imprinting pattern as well as genome-wide disturbances of gene expression were found in hybrid endosperms. Guided by these pilot results, we are further investigating the wild tomato system and combine laser-assisted microdissection and RNAseq technologies to identify genes imprinted in wild tomatoes in normally developing endosperm from crosses within three species: *S. peruvianum*, *S. chilense* and *S. arcanum* var. *marañón*. The latter taxon was chosen because it exhibits complete hybrid seed failure in reciprocal crosses with *S. peruvianum* but variable levels of hybrid inviability in reciprocal crosses with *S. chilense*, including asymmetric seed failure phenotypes. Based on such phenotypic data, we will determine whether (perturbed) genomic imprinting and overall expression modulation in hybrid endosperms are correlated with different proportions of hybrid seed failure in various interspecific crosses. Our project will provide new insights into the molecular manifestations of an important class of postzygotic barriers in angiosperms.

398D

Genomic regions and genes responsible for species-specific traits and speciation in Lake Victoria cichlid fishes

Yohey Terai [1], Ryutaro Miyagi[2], Shohei Takuno[1]
[1] SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan, [2] Department of Biological sciences, Tokyo Metropolitan University, Hachioji, Japan

Lake Victoria has dried up and has been refilled roughly 15,000 years ago. This lake harbors several hundreds endemic species of cichlids. They have undergone very recent and rapid speciation events during this short period. These fish species are genetically closely related and share nucleotide polymorphisms among species. So far, only one gene, a long-wavelength sensitive opsin (*LWS*), was identified as a gene bearing fixed genetic differences between species. Further analysis showed *LWS* was responsible for adaptation and speciation in cichlids. However, other genes with fixed genetic differences have not been identified from Lake Victoria cichlids. In this study, we identified 20 differentiated genomic regions with fixed genetic differences across the genome of Lake Victoria cichlids. We analyzed genomic DNA sequences from two Lake Victoria cichlids (20 individuals each), and extracted 20 highly differentiated regions (~20 kb) with fixed differences by Fst sliding-window analysis. These regions contained one to three coding genes, one of which included *LWS*. Several of those genes were associated with photoreceptors, morphogenesis, and sex differentiation. The amino acid sequences of the photoreceptor-associated gene were identical but expression level was different between in the eyes of the two species. Three genes showed completely different expression level in the lateral skin of two species. Thus, at least the genes in five regions including *LWS* out of 20 regions were indeed responsible for the species differences, suggesting the genes in differentiated regions may be responsible for the adaptation and speciation of Lake Victoria cichlids as was demonstrated in *LWS*.

399A

Genomic architecture of parallel contact zones along the speciation continuum of an avian superspecies complex

Christen Bossu[1], Alexey Kryukov[2], Jelmer Poelstra[1], Nagarjun Vijay[1], Matthias Weissensteiner [1], Jochen Wolf[1]

[1] Evolutionary Biology Centre, Ecology and Genetics,Department of Evolutionary Biology, Uppsala, Sweden, [2] Institute of Biology and Soil Science, Far East Branch of the Russian Academy of Sciences, Laboratory of Evolutionary Zoology and Genetics, Vladivostok, Russia

During population divergence, certain regions of the genome accumulate genetic differences more rapidly than others. However, the underlying evolutionary forces shaping heterogenous differentiation landscapes are not yet fully understood. In particular, genomic signatures of background selection are especially difficult to disentangle from divergent selection with gene flow promoting reproductive isolation. Here, we sequenced the genomes of 128 individuals across the speciation continuum of the crow superspecies complex (*Corvus* [*corone*] ssp.) which is characterized by parallel evolution of a sexually selected plumage phenotype (populations with all black versus black-and-grey plumage separated by narrow hybrid zones). For most genomic regions, patterns of genetic differentiation were consistent with linked selection shared by all populations, most likely as a result of reproductive isolation in allopatry. Yet, several regions in the genome exhibited signatures indicating divergent selection against gene flow. Despite an apparent parallelism in phenotype, these signatures differed across replicate hybrid zones, pointing towards context dependent selection.

400B

Adaptive genetic diversity of Swiss *Apis mellifera* populations using whole-genome sequencing

MELANIE PAREJO [1,2], David Wragg[3], Peter Neumann[2], Laurent Gauthier[1], Markus Neuditschko[1]

[1] Agroscope, Swiss Bee Research Center, Bern, Switzerland, [2] Institute of Bee Health, Vetsuisse Faculty, University of Bern, Bern, Switzerland, [3] Institut National de la Recherche Agronomique (INRA), Castanet Tolosan, France

The original distribution of the honey bee *Apis mellifera* ranges from Europe, Africa and the Middle East encompassing a large native range with diverse environmental conditions. In total there are more than 29 *Apis mellifera* subspecies differing in phenotypic traits such as morphology, behavior and resistance to disease. Based on phylogenetic analyses the species can be clustered into four major geographically and genetically different lineages (FST between lineages previously estimated on average 0.42). This evolutionary history offers a great model system to study speciation and molecular diversity of local adaptions.
The native honey bee subspecies of Northern Europe *A. mellifera mellifera* is adapted to short foraging seasons and comparatively long winters by developing the ability to form winter clusters. Global trade of this important pollinator for commercial operations in combination with its highly polyandrous mating system pose a risk on the genetic integrity of locally adapted ecotypes through introgression potentially resulting in the loss of valuable local adaptations shaped by natural selection.
We sequenced 120 Swiss *Apis mellifera* drones from different lineages with a sequencing depth of 10x per individual. The haploid male genome allows us to confidently identify single nucleotide polymorphisms (SNPs) with less coverage than an equivalent diploid individual. We will present preliminary results of the adaptive genetic diversity of Swiss *A. mellifera* populations from different lineages by identifying signatures of selection and comparing neutral versus adaptive molecular markers inferred from whole-genome next-generation sequencing data.

401C

Speciation process in Silene nutans

Hélène Martin[1], Fabienne Van Rossum[2], Jean-François Arnaud[1], Pascal Touzet [1]
[1] University of Lille 1, Evolution Ecology Paleontology, Villeneuve d'Ascq, France, [2] Meise Botanic Garden, Meise, Belgium

Silene nutans (Caryophyllaceae), an herbaceous long lived perennial plant, presents two distinct genetic lineages in Western Europe. Due to post-glacial re-colonization, the two lineages came in secondary contact in Belgium, southern England, and western France. Interestingly, cross experiments between plants found in these contact zones and belonging to different lineages, exhibit a strong reproductive isolation. Therefore, genetic differentiation, post-zygotic barrier to hybridization, as well as contrasting morphologies, suggest the occurrence of cryptic species in this taxon. We will present: i) the most relevant demographic scenario that could have led to this speciation event, using Approximate Bayesian Computing on chloroplastic sequences and nuclear RNAseq data on a representative sampling, and ii) the result of a transcriptomic scan in order to characterize the genetic architecture of hybrid incompatibilities in S. nutans.

402D

How the speciation gene *PRDM9* evolved in speciose tarsiers

Sacha Heerschop [1], Hans Zischler[1], Stefan Merker[2], Dyah Perwitasari-Farajallah[3,4], Christine Driller[1]

[1] Institute of Anthropology, Johannes Gutenberg-Universität, Mainz, Germany, [2] Department of Zoology, State Museum of Natural History Stuttgart, Stuttgart, Germany, [3] Primate Research Center, Bogor Agricultural University, Bogor, Indonesia, [4] Department of Biology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia

The search for genetic factors underlying reproductive isolation has attracted increasing scientific attention, and often focusses on genes involved in hybrid sterility. One of these so-called speciation genes is *PRDM9* encoding a zinc finger protein that specifies hotspots of meiotic recombination. The *PRDM9* zinc finger (Znf) domain evolves rapidly in number of repeats and in its sequence resulting in altered DNA-binding specificity, and, therewith, is probably causative for genetic incompatibilities driving species divergence in mammals including primates. We here for the first time explore the evolutionary dynamics of the *PRDM9* Znf array in tarsiers, small nocturnal primates representing the deepest offshoot of extant haplorhines. Living species of tarsiids are endemic to insular Southeast Asia and can be segregated into the three major and geographically isolated lineages of Western, Philippine, and Sulawesi tarsiers. We describe allelic variation and similarities among these lineages and, in particular, within the Indonesian island of Sulawesi, hosting by far the most species-rich tarsier clade. Our data support the different levels of evolutionary distinctiveness for each of the three lineages being in concordance with their ancient divergence dated to the Miocene. The high allelic diversity among comparatively young Sulawesi tarsier populations, however, may point to multiple local effects of genetic drift. Our characterization of the tarsier *PRDM9* ZnF domain not only confirms the diversification history of this ancient primate lineage but also strengthens its long-standing remote evolution over 80 million years.

403A

A screen for the Bateson-Dobzhansky-Muller incompatibility regions in yeast

Gönensin Ozan Bozdag, Emre Karakoc, David Rogers, Arne Nolte, Duncan Greig
Max-Planck Institute for Evolutionary Biology, Plön, Germany

Biological species are described as reproductively isolated entities. Since mutations causing reproductive isolation (RI) are in various types, different models of RI has been proposed. One of these models, the Bateson-Dobzhansky-Muller (BDM) 'genic incompatibilities', has received great attention since 1930s. However, scarce evidence is mainly limited to *Drosophila*. Thus further extension of evidence to different taxa is needed. Yeast species, *S.cerevisiae* and *S.paradoxus*, are good candidates to test the BDM model. When diploid hybrids undergo meiosis, only 1% of the offspring survive. By overcoming the anti-crossover problem between two species, we increased hybrid-offspring viability remarkably (up-to ~30%). Next, we genotyped 372 euploid hybrid-offspring to discover the BDM regions between two genomes.

404B

Analysis of prevalence of epistasis on the basis of huge phylogenies

Galya Klink [1,2], Georgii A Bazykin[1,2]
[1] *IITP, Moscow, Russia,* [2] *MSU, Moscow, Russia*

Epistatic interactions between amino acid sites shape the site-specific fitness landscapes, affecting the site-specific probabilities of fixations of different amino acids. There is abundant evidence that epistasis has a major role in shaping the evolution of protein sequences; however, it is hard to quantify its contribution.
Here, we reconstruct the phylogeny of several mitochondrial proteins from ~3,000 metazoan species, and use this data to obtain high-resolution site-specific distributions of times between points of occurrence of every amino acid observed at each site. We show that substitutions to the same amino acid are clustered on the phylogenetic tree, and that the extent of clustering is higher in conservative sites.

405C

The transcriptional trajectory of parallel pharyngeal jaw evolution in African cichlids

Pooja Singh, Christine Boerger, Nina Znidaric, Christian Sturmbauer
University of Graz, Graz, Styria, Austria

Cichlid fishes epitomise the most spectacular cases of adaptive radiation in vertebrates. These radiations are particularly dense in the Great African Lakes (Tanganyika, Malawi and Victoria). One of the major drivers of speciation in cichlids has been the key innovation of flexible and efficient oral and pharyngeal jaws, which facilitated trophic diversification. However, the transcriptional basis of cichlid oral and pharyngeal jaw development is sparsely resolved, with only a handful of genes implicated in this process. A striking facet of cichlid radiations is evolutionary parallelism of similar trophic morphologies, which presents an exciting opportunity to address fundamental questions regarding the genomics of adaptation. We hypothesise that the short diversification time makes it plausible that ecological adaptations may have primarily been brought about by changes in gene expression. With the aim of dissecting the regulatory network underlying pharyngeal jaw development to elucidate broader patterns of parallel evolution in African cichlid radiations we utilised a comparative mRNA-sequencing approach. Preliminary data analysis of differentially expressed genes in after yolk-sac absorption cichlids young from Lake Tanganyika with different trophic specialisations (algae browser, invertebrate feeder and piscivore) revealed a suite of differentially expressed genes, some previously implicated and some novel. A cluster of *c-fos* cichlid paralogues featured prominently in gene expression differences between algae browsers versus the carnivorous species. Gene ontology annotation of highly differentially expressed genes between the species revealed an enrichment of myoglobin and cytoskeleton ontology terms. In the future correlation networks of candidate genes will be elaborated to study broader developmental pathways.

406D

What does *Aquilegia* genome tell us about speciation?

Gokce Akoz [1,2], Magnus Nordborg[2]

[1] Vienna Graduate School of Population Genetics, Vienna, Austria, [2] Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria

The columbine genus *Aquilegia* (Ranunculaceae) first attracted the attention of Verne Grant more than 50 years ago. His influential studies on the role of floral isolation in plant speciation inspired further studies on *Aquilegia*. Several findings suggest that columbines display recent adaptive radiation across the Northern Hemisphere. This recent and rapid spread makes *Aquilegia* an ideal model organism to study changes that occurred during the early stages of speciation at a genomic level. To get a glimpse of the patterns of genetic variation in the comparatively less studied European taxa, we initially whole-genome sequenced seven individuals of *A. vulgaris,* which as the most widespread species of the genus is found from southern Scandinavia throughout Europe, and one individual of *A. pyrenaica* native to the Pyrenees. Principal Component Analysis and Neighbour-Joining trees on genotypes both indicate that *A. pyrenaica* is separated from all *A. vulgaris* individuals, and also that *A. vulgaris* individuals are separated by geography, following a pattern of Isolation by Distance. We aim to sequence other *Aquilegia* species to extend our understanding of genetic variation, which will, in turn, be decisive for further sampling required to delve more deeply into within population dynamics.

407A

Evolution of the D. immigrans species group revealed by using big sequence data

Yosuke Seto [1], Masanori Toda[2], Koichiro Tamura[1]
[1] Tokyo Metropoilitan University, Hachioji, Tokyo, Japan, [2] Hokkaido University Museum, Sapporo, Hokkaido, Japan

The subgenus Drosophila is a major subgenus in the genus Drosophila. Nevertheless, the phylogenetic relationships among species within this subgenus are still controversial. Within this subgenus, the D. immigrans species group is a large species group morphologically characterized by an array of spinules on the inner side of fore femur. However, previous molecular phylogenetic studies suggested that D. annulipes belonging to the D. immigrans species group is more closely related to Hawaiian drosophilids than other species in the D. immigrans species group. The Hawaiian drosophilids include two groups, endemic Hawaiian Drosophila and the cosmopolitan sister genus Scaptomyza. Therefore, we determined nucleotide sequences of a large number of genes for 11 species belonging to the D. immigrans species group and the genus Scaptomyza using RNA-seq. Then, we conducted a phylogenetic analysis together with the published genome data of Drosophila species to clarify the phylogenetic relationships of the species between the D. immigrans species group and the Hawaiian drosophilids. As the results, we found that D. annulipes and D. curviceps of the D. immigrans species group are not closely related to the major cluster of this species group but the cluster of Hawaiian Drosophila, D. grimshawi, and Scaptomyza graminum and S. flava. This result strongly supported the polyphyly of the D. immigrans species group and suggested the necessity of revising the classification of the species belonging to the D. immigrans species group.

408B

An upgrade and revision of the chimpanzee reference genome

Lukas Kuderna [1], Chad Tomlinson[2], Andrew J Sharp[3], Lars Feuk[4], Wesley C Warren[2], Tomas Marques-Bonet[1]

[1] Institute of Evolutionary Biology (UPF-CSIC), Barcelona, Spain, [2] The Genome Institute, Washington University School of Medicine, St. Louis, USA, [3] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA, [4] Department of Immunology, Genetics and Pathology, Rudbeck Laboratory and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Comparative and evolutionary genomics currently highly rely on genome reference assemblies in order to detect variation between different organisms or species. However, the vast majority of current assemblies are based on whole genome shotgun (WGS) sequencing, a technique known to systematically underrepresent certain complex, and potentially biologically relevant, regions of the genome. Furthermore, these assemblies are highly fragmented in terms of contig and scaffold continuity, leading to fragmented or fully missing gene models and a general loss of information.

The genome of our closest living evolutionary relative, the chimpanzee, is of key importance to study the evolutionary trajectories of our own species. Nevertheless, its current reference assembly is also based on a WGS and therefore only a draft and a highly fragmented representation of the genome with still over 25,000 scaffolds and 100,000 gaps.

Here, we will present our efforts to apply and integrate different 3rd generation sequencing technologies as well as novel genome assembly strategies to resolve most of the complex and missing sequences of the current chimpanzee genome reference, trying to bring its quality closer to the human reference.

409C

The effect of gynodioecy on molecular evolution of nuclear and mitochondrial genomes

Roberta Bergero, Deborah Charlesworth
University of Edinburgh, Edinburgh, UK

The occurrence of interactions between mutations in plant mitochondrial DNA (causing male sterility) and nuclear genes (restoring male fertility) represents a classical example of "selfish genetic elements". In mitochondrial DNA sequences from a few gynodioecious genera, including Plantago, substitution rates appear to be highly elevated for synonymous, but not non-synonymous, sites, and this has been attributed to high mtDNA mutation rates in these taxa. Alternatively balancing selection acting on mtDNA haplotypes in species with long-term male sterility polymorphisms (gynodioecious species) might produce high nucleotide diversity within species, which could lead to high sequence divergence between single alleles sampled from different species. To test whether mitochondrial polymorphisms are maintained for long evolutionary times, we are sequencing mtDNA and nuclear genes and analysing diversity in gynodioecious Plantago species.

410D

Patterns and dynamics of genomic divergence in *Aquilegia*

Daniele Filiault [1], Elizabeth Cooper[2], Evangeline Ballerini[3], Nathan Derieg[3], Scott Hodges[3], Magnus Nordborg[1]

[1] Gregor Mendel Institute, Vienna, Austria, [2] University of Southern California, Los Angeles, California, USA, [3] University of California Santa Barbara, Santa Barbara, California, USA

Due to a recent adaptive radiation, the genus Aquilegia is an excellent model for exploring genomic differentiation along the speciation continuum. In the first part of this work, we focused on understanding the genomic changes involved in establishing and maintaining reproductive isolation between two very closely-related species: *Aquilegia formosa* and *Aquilegia pubescens*. Despite marked differences in pollination syndrome and habitat, these species are interfertile and in fact, genome-wide divergence between the two species is very low. We sequenced five population pools of both A. pubescens and A. formosa with next generation sequencing and performed population-based F_{ST} scans. On the basis of these tests, we identified genes that were highly-differentiated in inter- but not intra-specific population comparisons. Interestingly, several highly-differentiated loci co-localized with a single, major QTL for flower color, suggesting that adaptation to pollinators is one factor driving the divergence of these two species. Continuing work focuses on understanding the genomics of the entire radiation event itself. To this end, we resequenced 10 Aquilegia species from across the Northern Hemisphere and examined patterns of allele sharing between and among species throughout the genus. These patterns revealed that decidedly non-tree-like relationships among species extend throughout the genus. Interestingly, the segregation patterns of some variants reach back into the Semiaquilegia outgroup, opening the possibility that selection on specific loci may have played a key role in the Aquilegia radiation.

411A

GBS data reveal population contraction and expansions in three African rainforest trees during the climatic oscillations of the Pleistocene

Rosalia Pineiro
Natural History Museum, University of Copenhagen, Copenhagen, Denmark

Ï∞îÑp≠§«)÷9ÚHÀ⁄n,Lte‘EhK]ÁÎπáË’Õ‘©!¢ˇ¸˜ÚWŒt9Ê+-Ê/æÚãFù˚Önt[g…«>J±≈≈qÏª»à*ƒ*
Èçd(KPj{`.Ó¶›•ìuÓ≈U¸ET?å8≈cø¯€WΔπXò€º ûäáì.®~7é *ƒ*¥6≠A“ˆ–X,„ë€ ¨˚5n≈iúÿ·™á«!ûøıÕüvs#
SÏ“µoSÙtº ¸ùx“°ãaæØ`fPfiSÙ„,5«2ï,ŒΩ˙•ôd—¸≈R`t-≤{%æ¡[“ÒÒÏìYX¶p#∑µ¸\Y5õM©˙∞–
ŸRü÷’SIäELÀ79˙lFF˝[@Å˝hãfã¥+ xgFît1<-…fi€}¯.˘3¸éÉ
≈¨¢4ëáô`GÄ≈î=>Tê¸SÕ:Rfl5Ò˜Xb…"fñQÖ^q†°BÖ.Œd›x“πütʃ¥ÿåŒZ2c!tÖâ˙ÿì v¸C˘ˆäÅ`Å-
Ã¶S˝µc8xO']®ò&g|˜t8ã
ÿÒt®¬¯±£,„*ƒ*,∑^o∂C§€F*I2_ı∑ΩRB¿i7¬°iÄÂnÈ=vÎJ˝y∂Ú˝˜4¸"ÈjÀ≠úgàèü∏hbÑFçM¯§npkœ.‰
flß„,Ôø◊"®‡&Ñ∏p2˘•–º¨õt¸s≠: Œ®Œà[Ú Óè±ʃÇ,ˆÍÕÆ˘£Öq‰˘q`’åê`0flÕÑéÑ{¸`ª2¯u~
Û#;Á\jÿç!õúk∂»¿„e¥ñm°«Ü—
ıøQ±Ø◊ı°„l?*KeÌıøa¸ï®nTãèV¯ØΩŒêʃ>!ë˘ÌCKMM1âù»fiıÉÛøj*È&_< U†éEö‰AÌï—
ñµIöyËw}HôΩõ˜
Œsåz´7ØÿssÏÑ»'·U6ªøñoLÙKW*Q„›£Å‡)?ÃÃ≤˙ÊÁʃÿ˘…iáø†8Ø?}ª£π˙1‹ÁKfT´2,ìgú?£=nG,û¸˝yπ9
ë\!uG¶Æsª&fiöv¥,Ù√1|Ù@u^Ti@°“Úv8"µ±…›ûu©ú˝f+K≤xåÊU≈Æ†HûÁG*ƒ*!‰Ôâl‡§≈J$«kÌ⁄¡™
∂B€ü¢fl¥Ùº!Ÿº ùÅüú™°,'ü◊<Ä¶-%<’Ó@ œóÅóÖË˘k‡˘Ma'Æb1≤øÏFŒØª¸—
¡Sfófß@·Q•0@åCT;Öd≠CBGïâ∏8@2STmE"HÏ≠Ü€¥\wÄF∞Br~ÈNfQ©dIY$›FO◊#*.∑sÉª~É*Q˜a
C‹º:Åg¶◊–nÉ"<€¯/£7ñÍr≤”ïèΔkÀ‘ ’k¯"]xQ«√$E±¸¡Œi
vZ%G3Ó’rkrõ®\ˆ@gøËÇû•˝"ŸFLç·¸‡kÑ™ìÜF¬ÁóºDÂ+Bº¢≤√i¶`∏H¬ʃ«0TC+R\¥óEEpïäSé¿ÊnŒ
nïK¯©S™∑µtI;»Û"k@xbÃ÷iN˜Ìª¯Ì⁄Àmú∞¥*hX+ÆíÀæô¢»;)Ùç-
·†∂.D4{b@PNÃÜ◊l∑≠ÕÃzøıπgôG©°ÏëÈ∞o√Jb¸©`”iÈfi— áy0*Z(@lù\OìYÚä’ÉzIflE[l#ˆÒh¥öf-
í]{›j®X≥Á˘qfÆu⁄ê V ÜÔËI,®Ãö‘óã‘ù.nËj¡°ΔíËOã}ofiÕqº∞^…¯\±-e‘–C –fl“£÷«Ó‘ıÍ€t8ÊÀ~¯1-
{t¢&˘Cä“ñ∫ÌÔ”dÁÂVîg¿eêŸT:cØ*i≈»›¥›ÿ5%h‡üêïn«◊ΩfiZñ˜†˝ÑZ™èâÛ⁄õ
{Ãò©°∂…5Y}mgØΩ}õfÍü)O¿ôÙ™»0¢≈#›#óZu√PÂ#˜Æ˝>6ôQÿ”•3O¡h‹IV*ƒ*∫êùp∞˙zk—
¯˜Ì≠V‹ÿÊö8oåºÅÏ¥…?}z‰˜√‡øy3D]=Δñ’¯ÿe÷Hé´î_…ôÂ"Ì’°Δæ]∂*^q°≠¯0Ö C˝Zç_wCGz‰
wæËsÚÁ,m¢£≠Ù3
E4Äú¿O]TK˚à¥ò}ó§"ôhP^¬\yÈÀ^¯Õ4K<’`qügVÜZ‘ÙVd∂q{äˆŸ·ñäƱ¸Úû[f†5˜”ÿ>˝fi¡!Ÿoë8œhïFû
(◊!ìà9rxns¥9œ∑à‰¥á"ë ıï¿˚oú©fi7C∑ßgı>9X∂îÔ5ÔkµqiİØø“⁄≥ycj9œyvm˜[Ÿóï ê†?≈{–
Ù…Æ¢b}ıÿ∏∏∏N¢¶íX
ÈÆ≤K{É8)îøC„BòâLÊ*?W8ºK{Yõ˘œS1sÅP(∂nÍø14Ñ\WFï°3éœD€ŸT˘ìfª!T|7c{(Ù˝82

412B

Island evolution in Amazon Parrots

Pavel Dobrynin [1], Sofia Kolchanova[1], Taras Oleksyk[2]

[1] Theodosius Dobzhansky Center for Genome Bioinformatics St.Petersburg State University, Saint Petersburg, Russia, [2] Department of Biology, University of Puerto Rico at Mayagüez, Mayaguez, Puerto Rico

Our research objective is to use a local model of island speciation centered on the endemic Caribbean group of parrots, starting with Puerto Rico's own critically endangered Amazona vittata. Inferences about genome variation and population structure of this species are used to help decisions in the captive breeding programs. Crowd funding has allowed for addition sequence data from the closely related A. ventralis and A. leucocephala. We de novo assembled and annotated genomes from all three species. For gene annotation we used homology-based approach and pre trained HMM gene models. We are interested in question about island evolution, speciation and adaptation, so we looked both for signatures of selection and gene families' expansion and contraction events. We used M. budgerigar and T. guttata as outgroup for the current analysis.

413C

Improved assembly and annotation of the Puerto Rican parrot genome

Sofia Kolchanova [2], Pavel Dobrynin[2], Taras K. Oleksyk[1]

[1] University of Puerto Rico, Mayaguez, Puerto Rico, [2] Theodosius Dobzhansky Center for Genome Bioinformatics at St.Petersburg State University, St.Petersburg, Russia

Despite the advances in sequencing and assembly technologies working with non-model organisms is still a challenge. We improved sequence of the A vittata genome, produced an assembly of scaffolds, and evaluated this assembly as compared to budgerigar, chicken and zebrafinch. In this study we used de-novo annotation as well as transcriptome sequencing of 4 distinct tissues (brain, heart, blood and gut) to produce a draft definition of the genes and proteins present in the genome. Using a database of proteins from 8 other bird species catalogued by NCBI, the function of the genes were characterized to begin to address questions of natural selection and drift. Using the gene annotations we looked for SNPs and genetic variation that distinguishes A. vittata from A. leucocephala and A ventralis to be characterized from both evolutionary and systems biology perspectives.

414D

Investigating genome-wide patterns of introgression in experimental hybrid swarms

Wynn Meyer, Dat Mai, Doris Bachtrog
University of California, Berkeley, Berkeley, CA, USA

The process of speciation occurs through the buildup of barriers to gene flow between diverging populations. Before the evolution of complete reproductive isolation, different genomic regions may be differentially able to introgress, or to be shared across species boundaries. Comparing patterns of introgression across the genomes of recently diverged lineages therefore provides a way of identifying commonalities among the types of genes and regions that are important in the early stages of speciation. We here use genome-wide data from experimental hybrid swarms, derived from species or subspecies pairs from throughout the *Drosophila nasuta* complex, to identify patterns associated with the evolution of reproductive isolation in these species. This complex consists of multiple recently diverged species, many pairs of which are inter-fertile, enabling experimental crosses and maintenance of hybrid populations for multiple generations. In these hybrid populations, we compare patterns of introgression between the autosomes and the sex chromosome, which has frequently been proposed to play a strong role in the evolution of reproductive isolation. We also identify regions that appear to be resistant to introgression in multiple species pairs, providing strong candidates for the loci contributing to the early evolution of reproductive isolation. Finally, we evaluate patterns of linkage disequilibrium between unlinked loci to search for evidence of negative epistatic interactions between different parental alleles, or candidate Bateson-Dobzhansky-Muller incompatibilities. Our study demonstrates how such experimental systems may be used to identify both broad-scale patterns as well as individual candidate loci involved in reproductive isolation in incipient species.

11 PopGen in space! Theory and inference in spatial population genetics

11.1

Consequences of spatial expansions on population functional diversity

Laurent Excoffier [1,2], Stephan Peischl[1,2]

[1] *University of Berne, Berne, Switzerland,* [2] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*

It is known that spatial expansions have had a major influence on population genetic diversity: some neutral variants can increase in frequency and spread over large areas in newly occupied territories. This is the phenomenon of gene surfing. However, selected variants can also surf and thus modify the fitness of expanding populations. We have studied this phenomenon by simulations and by analytical derivations in relatively simple models of expansions in homogeneous environments. Very generally, we find that the fitness of populations located on the expansion front decreases as a function of their distance from the origin of the expansion. The creation of this expansion load happens in 1D or 2D expansions, in case of hard or soft selection, in presence or absence of recombination and for different distribution of fitness effects. However, the evolutionary dynamics of the expansion load differs between cases, and also depends on the level of dominance between variants. All these cases will be briefly presented, and we will conclude by showing some evidence that this phenomenon also occurred in human populations.

11.2

EEMS: a spatially explicit method to visualize geographic population structure and analyze non-stationary isolation by distance

Desislava Petkova [1,2], John Novembre[1], Matthew Stephens[1]
[1] The University of Chicago, Chicagol, IL, USA, [2] Wellcome Trust Centre for Human Genetics, Oxford, UK

Genetic data often exhibit patterns that are broadly consistent with "isolation by distance" - a phenomenon where genetic similarity tends to decay with geographic distance. In practice, however, the relationship between genetics and geography might not be stationary, as predicted under isolation by distance; instead, it can vary across the habitat of a species due to a combination of environmental and historical factors.

We describe EEMS (Estimation of Effective Migration Surfaces) - a recently developed method that characterizes global patterns of genetic differentiation in terms of local rates of migration and diversity. EEMS uses the concept of "effective migration" to model the relationship between genetics and geography: in this paradigm, effective migration between nearby locations is low in regions where genetic similarity decays quickly.

EEMS uses a spatially explicit population genetic model to relate underlying migration and diversity rates to expected pairwise genetic dissimilarities. The inference procedure, in a Bayesian framework, utilizes priors that impose spatial smoothness with Voronoi tessellations and Markov Chain Monte Carlo techniques to sample from the posterior of the rate parameters. EEMS can quantify variation in effective migration and genetic diversity across the habitat, without splitting the samples into genetic clusters to either analyze the data or interpret the results. We demonstrate the potential and limitations of EEMS, and compare it to other methods in terms of its effectiveness to highlight important features of spatial genetic variation, by analyzing several geographically indexed datasets, including a densely sampled collection of Caucasian individuals from the United Kingdom.

11.3

The power of painting: using haplotype patterns to infer history

Garrett Hellenthal [1], Liam Quinn[1], Daniel Lawson[2], Daniel Falush[3], Simon Myers[4], Simon Mead[1,5]

[1] University College London, London, UK, [2] University of Bristol, Bristol, UK, [3] Swansea University, Swansea, UK, [4] University of Oxford, Oxford, UK, [5] National Prion Clinic, London, UK

The widespread availability of densely genotyped individuals representing hundreds of world-wide geographic locations has enabled researchers to reconstruct the ancestral history of numerous populations in unprecedented detail. The currently most powerful approaches to infer history use haplotype information, which exploit correlation patterns among Single-Nucleotide-Polymorphisms (SNPs) to increase precision over the far more commonly-used programs that ignore this information. Here I focus on the haplotype-based technique "chromosome painting", demonstrating how this method can be used to generate a fingerprint that uniquely defines present-day and ancient genomes. Using this approach, I highlight the specific scenarios where incorporating haplotype information provides the greatest benefits over non-haplotype techniques, for example in particular applications to elucidate population sub-structure, identify ancestral relatives, and discover/date past interbreeding events (i.e. admixture) among distinct populations. I further show how different painting protocols can be implemented in order to peel back layers of history when studying populations, for example distinguishing whether different genetic patterns among groups are attributable to recent isolation, periods of admixture and/or ancient substructure. I illustrate the power of these techniques in applications to world-wide human samples, including a new data collection with hundreds of individuals sampled from 18 groups of the Eastern Highlands of Papua New Guinea. We show the recent and ancient genetic influences of this uniquely isolated region that has been subjected to past outbreaks of the prion-based neurological disorder "kuru", showcasing how the rich information in DNA can help resolve existing controversies among anthropologists, historians and linguists.

11.4

New Routes to Phylogeography

Nicola De Maio, Daniel Wilson
Oxford University, Oxford, UK

Phylogeographic methods aim to infer migration trends and the history of sampled lineages from genetic data. Applications of phylogeography are broad, and in the context of pathogens include the reconstruction of transmission histories and the origin and emergence of outbreaks. Phylogeographic inference based on bottom-up population genetics models is computationally expensive, and as a result faster alternatives based on the evolution of discrete traits have become popular.

In this paper, we show that inference of migration rates and root locations based on discrete trait models is unreliable and sensitive to biased sampling. To address this problem, we introduce BaStA (BAyesian STructured coalescent Approximation), a new approach implemented in BEAST2 that combines the accuracy of methods based on the structured coalescent with the computational efficiency required to handle more than just few populations.

We illustrate the potentially severe implications of poor model choice for phylogeographic analyses by investigating the zoonotic transmission of Ebola virus. Whereas the structured coalescent analysis correctly infers that successive human Ebola outbreaks have been seeded by a large unsampled non-human reservoir population, the discrete trait analysis implausibly concludes that undetected human-to-human transmission has allowed the virus to persist over the past four decades. As genomics takes on an increasingly prominent role informing the control and prevention of infectious diseases, it will be vital that phylogeographic inference provides robust insights into transmission history.

11.5

Bayesian estimation of neighborhood size using composite marginal likelihoods

Tara Furstenau [1,2], Reed Cartwright[1,2]

[1] Arizona State University, Tempe, AZ, USA, [2] The Biodesign Institute, Tempe, AZ, USA

Wright's neighborhood size characterizes the degree of genetic differentiation in a population and is important for understanding evolution in realistic spatial dimensions. Several methods have been devised that make reasonable point estimates of neighborhood size but so far none have attempted estimates in a Bayesian framework. We propose a Bayesian composite marginal likelihood (CML) approach to estimate neighborhood size. The Wright-Malécot formula is useful for estimating neighborhood size; unfortunately, this formula is computationally intractable for any reasonably sized data set. A properly calibrated CML is a good proxy in this situation because it is able to capture the asymptotic properties of the full likelihood with a reduced computational burden. The CML is calculated as the weighted product of the probability of identity in state (IIS) between individuals who are sampled certain distances apart. We approximate the probability of IIS using a Taylor series for short distances and a modified Bessel function of the second kind for larger distances. The CML and a prior distribution on neighborhood size are combined to obtain the posterior distribution and samples from this posterior are obtained through MCMC simulations. To test our method we simulate the dispersal of alleles in a finite lattice model and calculate the probability of IIS between sampled individuals. A Bayesian CML approach allows us to make use of prior information and calculate credible intervals for the neighborhood size parameter; it also allows us to incorporate the innate variation and ambiguity of genetic markers into our model.

11.6

Inferring dispersal from spatial patterns of IBD sharing

Harald Ringbauer
IST Austria, Klosterneuburg, Austria

Recently it has become feasible to reliably detect long blocks of pairwise shared genome (>3 cM, say) from human data and will likely also become practical for many other species in the near future. I show that for a range of models of populations occupying a two dimensional habitat with spatially limited migration, the probability of sharing such long blocks falls off approximately exponentially with increasing sample distance and that the rate of this decay only depends on the backward dispersal rate and not on peculiarities of local population structure. Therefore, dispersal can be estimated by fitting this decay rate from observed block sharing data.
Since long shared blocks have predominantly recent common ancestry, using recombination as clock allows to probe relatively recent time scales. Thus, unlike currently used methods based on spatial allelic correlations (which can actually only infer the product of dispersal and effective density), the proposed inference scheme is not subject to ancestral confounding and does not need prohibitively long equilibrium times.
I will discuss the underlying theory and demonstrate the utility and robustness of the novel inference scheme via simulations. Furthermore, I will treat possible applications to existing human data, for instance to extract the scale of dispersal of pre-industrial Europeans.

11.7

Impact of Long-Range Dispersal on Range Expansions, Soft Sweeps and Genetic Hitchhiking

Oskar Hallatschek, Daniel Weissman
University of California, Berkeley, Berkeley, CA, USA

The population genetics of spreading processes is well-understood in the limiting cases of panmixia and short-range migration. The intermediate case of spreading with rare long-range jumps is poorly understood, yet it is highly relevant to the population genetics of modern epidemics or species invasions. For instance, it is not clear under which conditions long-range migration preserves or purges standing genetic diversity, and which patterns of diversity and genealogical trees to expect. We present a simple and intuitive mathematical argument that allows to predict many population genetic consequences of long-range dispersal, including its impact on gene surfing, soft sweeps and hitchhiking. We find that patterns of diversity, as measured by site-frequency spectra and various genealogical tree statistics, exhibit distinct regimes of “patchiness" depending on the tails of the dispersal kernel. Importantly, the sensitivity of the population genetics to the dispersal kernel also enables inference of the latter based on population genomic data.

11.8

Evolution of quantitative traits under a migration--selection balance: when does skew matter?

Florence Debarre [1,4], Sam Yeaman[2], Frederic Guillaume[3]
[1] University of Exeter, Penryn, UK, [2] University of British Columbia, Vancouver, Canada, [3] University of Zurich, Zurich, Switzerland, [4] Wissenschaftskolleg zu Berlin, Berlin, Germany

Quantitative genetic models of differentiation under migration--selection balance often rely on the assumption of normally distributed genotypic and phenotypic values. When a population is subdivided into demes with selection towards different local optima, migration between demes may result in asymmetric, or skewed, local distributions. Using a simplified two-habitat model, we derive formulas without a priori assuming a Gaussian distribution of genotypic values, and we find expressions that naturally incorporate higher moments, such as skew. These formulas yield predictions of the expected divergence under migration--selection balance that are more accurate than models assuming Gaussian distributions, which illustrates the importance of incorporating these higher moments to assess the response to selection in heterogeneous environments. We further show with simulations that traits with loci of large effect display the largest skew in their distribution at migration--selection balance.

11.9

Limits to a species' range in one- and two-dimensional habitats

Jitka Polechová, Nick Barton
IST Austria, Klosterneuburg, Austria

Why a species' range sometimes ends abruptly, even when the environment changes smoothly across space, has interested biologists for decades. Studying one-dimensional habitats, we have found that there is an inherent limit to adaptation arising in any (finite) natural population. Two observable parameters describe the threshold when adaptation fails: i) the loss of fitness due to dispersal to a different environment, and ii) the efficacy of selection relative to stochastic effects in finite populations – the genetic drift [http://dx.doi.org/10.1101/012690]. However, in broad two-dimensional habitats, the effect of genetic drift changes, as stochastic fluctuations of allelic frequencies become only weakly dependent on selection. We show how the limit to adaptation, and hence to a species' range, extends to two-dimensional habitats.

11.10

Simulation of spatial processes shaping genetic diversity across time

Nuno M Silva, Mathias Currat
Anthropology, Genetics and Peopling History Lab, Anthropology Unit – Department of Genetics & Evolution, University of Geneva, Geneva, Switzerland

Neutral genetic diversity in human populations reflects their evolution as a consequence of demographic changes and migrations. Tracing back the history of the human species through the analysis of contemporary genetic data relies on models and indirect evidences. However, recent advances in sequencing technology provide a means to access direct snapshots from the past with DNA data of ancient populations (aDNA). This has opened a larger window of questions about human evolution, but the approaches used so far to analyze ancient molecular data neglect spatial processes - such as gene flow and admixture - shaping the neutral genetic diversity found in human populations. Here we propose to fill this gap by analyzing aDNA using a spatially explicit modeling framework.
Spatially explicit modeling of human history allows to formally testing realistic hypothetical scenarios that could have led to the diversity present in human populations. Here we use a modified version of the program SPLATCHE2 that allows serial sampling during the coalescent process. Contrary to previously existing methods, our approach allows to cope with population structure and migration and with the extensive temporal and geographic heterogeneity commonly found in aDNA datasets, at the sample level.
We demonstrate the importance of spatial processes' effects on the genetic diversity of serial samples, and consequently on their genetic differentiation. We also explore the kind of genetic relationships expected in case of full or partial population continuity in a given area. Finally, we apply our approach to real data from prehistoric Europe.

11.11

The effects of density-dependent range expansion on interspecific genetic introgression

Claudio S. Quilodrán [1], Béatrice Nussberger[2], Mathias Currat[1], Juan I. Montoya-Burgos[1]
[1] Department of Genetics and Evolution. University of Geneva., Geneva, Switzerland, [2] Institute of Evolutionary Biology and Environmental Studies. University of Zurich., Zurich, Switzerland

The dynamics of range expansion need to be taken into account when studying the genetic consequences of an invasive species spreading and interbreeding with native populations. This is also critical when interbreeding results from shifts in the natural distribution of species due to climate change. Models aimed at studying the genetic consequences of species range expansion have been recently developed but usually assume that dispersal is independent from local population densities. However, organisms may disperse because they are attracted by conspecifics, or to the contrary, because they prefer depopulated areas. These behaviours are referred to as positive or negative migratory responses toward conspecifics. Here, through spatially explicit simulations, we assess the effects of various forms of density-dependent dispersal during range expansion on the genetic introgression between two interacting species. We show that massive introgression of neutral genes in the invasive species occurs in all the density-dependent dispersal models (positive and negative), even when hybridization is relatively low. For a given hybridization rate the levels of introgression are lower when dispersal is negatively related to local densities and higher under positive density-dependent dispersal. Our results suggest that organisms that tend to disperse due to conspecific attraction are more affected by genetic introgression. We applied our theoretical framework on a real case of hybridization between European wildcat and domestic cat in Switzerland. We highlight that considering density-dependent dispersal has the potential to improve the predictive power derived from models of species range expansion.

11.12

Insights into British and European population history from ancient DNA sequencing of Iron Age and Anglo-Saxon samples from East England.

Stephan Schiffels [1], Wolfgang Haak[2], Bastien Llamas[2], Elizabeth Popescu[3], Louise Loe[4], Rachel Clarke[3], Alice Lyons[3], Pirita Paajanen[1], Duncan Sayer[5], Richard Mortimer[3], Chris Tyler-Smith[1], Alan Cooper[2], Richard Durbin[1]

[1] Wellcome Trust Sanger Institute, CB10 1SA Hinxton, Cambridge, UK, [2] Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA 5005, Australia, [3] Oxford Archaeology East, CB23 8SQ Cambridge, UK, [4] Oxford Archaeology South, OX2 0ES Oxford, UK, [5] University of Central Lancashire, PR1 2HE Preston, UK

British population history is shaped by a series of immigration periods and associated changes in population structure. It is an open question to what extent these changes affect the genetic composition of the current British population. Here we present whole genome sequences generated from 10 individuals, found in archaeological excavations in Hinxton, Oakington and Linton, close to Cambridge, and ranging from 2,300 years before present (Iron Age) until 1,200 years before present (Anglo-Saxon period). We use modern genetic samples from the 1000 Genomes Project and additional external data from Britain, the Netherlands and Denmark to characterize the relationship of these ancient samples with contemporary British and other European populations. By analyzing the distribution of shared rare variants across ancient and modern individuals, we find that samples from the Anglo-Saxon period are relatively more closely related to central northern Europe, while earlier samples and contemporary British samples are relatively more closely related to Southern European populations. To quantify this series of relationships further, we developed a new method, rarecoal, that fits a demographic model parameterized by split times, population sizes and migration rates to the distribution of shared rare variants across a large number of modern and ancient individuals. We use rarecoal to estimate the history of European population structure within the last 10,000 years and to map our ancient samples onto the European population tree. Our approach provides a unique picture of population history in Europe, and in particular helps characterizing the complex genetic impact of Anglo-Saxon immigrations into Britain.

11.13

GENETIC HISTORY OF LATIN AMERICA: SUBCONTINENTAL ANCESTRY AND REGIONAL POPULATION STRUCTURE

Juan Camilo Chacon-Duque [1], Kaustubh Adhikari[1], Victor Acuna-Alonzo[1,2], Samuel Canizales-Quinteros[3], Gabriel Bedoya[4], Carla Gallo[5], Giovanni Poletti[5], Francisco Rothhammer[6], Maria Catira Bortolini[7], Rolando Gonzalez-Jose[8], CANDELA Consortium[1], Garrett Hellenthal[1], Andres Ruiz-Linares[1]

[1] Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, UK, [2] National Institute of Anthropology and History, Mexico City, Mexico, [3] Facultad de Medicina and Facultad de Química, UNAM; Instituto Nacional de Ciencias Médicas y Nutrición 'Salvador Zubirán', Mexico City, Mexico, [4] Laboratorio de Genética Molecular, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellin, Colombia, [5] Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Perúana Cayetano Heredia, Lima, Peru, [6] Instituto de Alta Investigación Universidad de Tarapacá, Programa de Genética Humana ICBM Facultad de Medicina Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica, Chile, [7] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [8] Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina

Latin America has a complex history of extensive admixture between Native Americans, Europeans and Africans. Thus far, this process has been explored with genetic data to examine patterns of continental ancestry. Here we explore sub-continental ancestry using ~700,000 autosomal SNPs in a sample of ~6,000 Latin Americans from five countries (Brazil, Chile, Colombia, Mexico and Peru) from the CANDELA consortium (Ruiz-Linares *et al* 2014, PLoS Genetics 10(9)).

To do so, we describe a novel statistical approach that exploits patterns of haplotype similarity and gives increased precision over a related approach that was applied successfully to study the ancestral histories and fine-scale patterns of world-wide populations (Hellenthal *et al* 2014, Science 343(6172)) and different regions within Britain (www.peopleofthebritishisles.org). We identify the contributions of geographically precise ancestral components, including those from Africa and Europe, to both regions and single individuals. For Brazil, we highlight how the sources and dates of European Ancestry vary considerably, reflecting influx from Portugal, Spain, Italy and Northern Europe. For the other countries, former Spanish colonies, we identify ancestry from specific regions in Spain (e.g. the Basque Country, Andalusia, Catalonia). Native American ancestry also varies substantially, reflecting major contributions from specific local Native groups. Overall, with this new approach we are able to identify ancestry patterns at high resolution, allowing us a fine-grained analysis of how history shaped the genetic make-up of Latin America.

11.14

Exploring population separation history using physically phased genomes

Shiya Song [1], Elzbieta Sliwerska[2], Sarah Emery[2], Jeffrey Kidd[1]

[1] Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA, [2] Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA

The acquisition and analysis of haplotype resolved genome sequence information is of growing interest due to its applications in both population and medical genetics. We resolved phased haplotypes of five individuals from diverse African populations (including Yoruba, Esan, Gambian, Massai and Mende), using fosmid pool sequencing. We physically phased 98% of heterozygous SNPs into haplotype-resolved blocks, obtaining a block N50 of 1 Mbp. We analyzed population size and separation history using the Pairwise Sequentially Markovian Coalescent model (PSMC) and Multiple Sequentially Markovian Coalescent model (MSMC). We additionally applied PSMC to infer TMRCA distribution from two haplotypes, one from each population. We designed a grid search strategy to compare the simulated TMRCA distributions under different split time and migration rate combinations with the observed TMRCA distribution and found parameters that best explain the observed data. We applied this method to date the separation between Africans and out of African populations, and between western Africans and eastern African (Massai). We accessed the effect of using statistically phased haplotypes and different treatment of unphased sites on such inference. We obtained concordant results by analyzing relative cross coalescence rate using MSMC. We demonstrate the construction of haplotype sequences of sufficient completeness and accuracy for population genetic analysis.

11.15

Inferring selection from a spatially explicit demographic model based on a large number of human populations.

Benjamin Peter, John Novembre
University of Chicago, Chicago, IL, USA

Human genetic diversity is a result of past demographic and selective processes. While most variation is expected to be neutral, some variants have likely been impacted by selective processes. We fit global and continental scale diversity data to a spatially explicit demographic model using a recently developed method (EEMS, Estimation of Effective Migration Surfaces). In this context, the fitted EEMS model serves as a demographic null model of expected genetic differences and similarities between all individuals. As the EEMS posterior is a mixture of Wishart distributions, we can use Bayesian outlier detection methods to detect genetic markers whose global distribution is inconsistent with the fitted demographic model, indicating alternative explanations such as natural selection.

We extensively explore the statistical properties of our approach using simulations and show that it mainly has power to detect loci that have large allele frequency differences between otherwise closely related populations. Applying it to the human origins data set of Lazaridis et. al. reveals several well-known examples of selection such as LCT, RPTOR and the MHC region. This suggests that our approach is suitable to find outliers due to both directional and balancing selection. Overall, these results indicate that using demographic models from a large number of populations may increase the power of outlier detection methods.

11.16

Evidence for a common origin of Blacksmiths and Cultivators in the Ethiopian Ari within the last 4500 years: Lessons for clustering-based inference

Lucy van Dorp [1], David Balding[1], Simon Myers[2], Luca Pagani[3,4], Chris Tyler-Smith[3], Endashaw Bekele[5], Ayele Tarekegn[6], Mark Thomas[1], Neil Bradman[6], Garrett Hellenthal[1]
[1] University College London, London, UK, [2] University of Oxford, Oxford, UK, [3] The Wellcome Trust Sanger Institute, Cambridge, UK, [4] University of Cambridge, Cambridge, UK, [5] Addis Ababa University, Addis Ababa, Ethiopia, [6] Henry Stewart Group, London, UK

Unsupervised clustering algorithms such as ADMIXTURE are commonly used in studies of population structure. However, here we show that in certain situations the results of such algorithms should be interpreted with caution. We present a case study analysing DNA from caste-like occupational groups within the Ari of Ethiopia, specifically comparing Ari Cultivators and socially marginalized Ari Blacksmiths. There are two competing theories regarding the origins of the Blacksmiths as (i) remnants of an ancient hunter-gatherer population that inhabited Ethiopia prior to the arrival of agriculture (e.g. Cultivators) during the Neolithic, versus (ii) relatively recently related to the Cultivators but presently marginalised due to their craft. Two previous studies interpreted ADMIXTURE results as evidence for the Blacksmiths and Cultivators being anciently related, i.e. Hypothesis (i). Applying a novel haplotype-based approach to the same samples, which importantly can distinguish whether genetic differences between groups are likely attributable to relatively recent allelic drift, we provide compelling evidence that the genetic differentiation we observe between the groups today is exclusively due to the social marginalisation of the Blacksmiths over a relatively short time period, i.e. Hypothesis (ii). This finding illustrates how social constructions can directly lead to genetic differentiation between two previously similar groups, as well as providing a cautionary tale of over-interpreting ADMIXTURE plots.

11.17

The evolutionary genomics of a recent invasion: a natural mutation accumulation line of North American Arabidopsis depicted using historic herbarium samples

Moises Exposito-Alonso, Claude Becker, Detlef Weigel, Hernán Burbano
Max Planck Institute for Developmental, Tübingen, Germany

It has been suggested that the small selfer plant *Arabidopsis thaliana* colonized North America around the 19th century and that due to a notable founder effect, one major haplogroup populates the new continent nowadays. By sequencing whole-genomes of stock accessions collected in the US from 1993 to 2006 and herbarium specimens dating from 1863 to 1993, we inferred the ancestral spatio-temporal dynamics of this invasor North American haplogroup. Taking the advantage of several biological properties (i.e. self mating, single haplogroup and low effective recombination rate) that resemble a natural mutation accumulation population, we employed the latest phylogenetic and population genetics methodologies to infer a mutation rate, a genome-wide selection coefficient, and to model population size and migration trajectories. Our results showed that the substitution rate inferred from the field is slightly slower than the extracted in laboratory experiments, and that purifying selection is acting on a genome-wide average. We dated the origin some time back in the early 18th century. We estimated that the arrival was via New York area and the species expanded nearly exponential towards Michigan area, where the population size stabilized and where the main abundance and diversity is currently found. We evidence that the combination of statistical genomics and herbarium ancient genetics techniques dramatically increase the power of ancestral inference in a spatio-temporal explicit framework.

11.18

Identifying selected loci in a flower colour hybrid zone

Nick Barton [1], David Field[1], Hugo Tavares[2], Enrico Coen[2]
[1] *IST Austria, Klosterneuburg, Austria,* [2] *John Innes Institute, Norwich, UK*

Selected loci can be identified by finding regions of genome that are highly differentiated, and that show sharp spatial clines. However, it is not easy to show that high Fst and narrow clines are due to selection rather than random drift, and it is to not easy to show whether gene flow has had a significant effect on divergence. We show how alternative explanations can be distinguished, using data from a hybrid zone in Antirrhinum majus, where linked loci determine differences in flower colour.

11.19

Ancient divergence and repeated adaptive radiation in Macaronesian Islands Arabidopsis

Andrea Fulgione [1], Xavier Pico[3], Carlos Alonso-Blanco[2], Joachim Hermisson[1], Angela Hancock[1]
[1] University of Vienna, Vienna, Austria, [2] CSIC - Spanish National Center for Biotechnology, Madrid, Spain, [3] CSIC - Doñana Biological Station, Madrid, Spain

Continent-island systems represent repeated natural experiments where the details of population histories, adaptive evolution and the early stages of speciation can be revealed in a spatially explicit framework.

The Macaronesian Islands harbor exceptional species diversity and divergence, which is likely a result of ancient pre-glacial colonization events. We collected and sequenced the genomes of a set of 100 *Arabidopsis thaliana* samples from across Macaronesia and North Africa. Contrary to expectations that Arabidopsis was recently introduced in this region, our findings reveal a series of ancient, natural radiations to the archipelagos, followed by more recent admixture into a subset of islands. Overall, diversity is high and on the order of worldwide diversity (theta = 0.3%). The limited set of distinct colonization events combined with low subsequent gene flow allowed us to reproduce the details of demographic history using variant and haplotype information from Macaronesian and diverse mainland population samples.

We also examine the spatial distribution of variation within and across islands to infer adaptive histories. In the Cape Verde Islands, which are situated in a climatic extreme of the species range, Arabidopsis displays striking phenotypic divergence. In this archipelago, we find evidence for rapid parallel adaptation across two different islands based on the geographic distributions of validated functional variants. This study of a model organism in a tractable island system provides a simplified ‘island model’ test case, where we are able to reconstruct the details of demographic and adaptive processes.

11.20

Molecular evolutionary consequences of island colonisation

Jennifer James [1], Robert Lanfear[2,3], Adam Eyre-Walker[1]

[1] University of Sussex, Brighton, UK, [2] Macquarie University, Sydney, Australia, [3] The Australian National University, Canberra, Australia

Island colonisation is expected to have dramatic consequences for the molecular evolution of species. Island species are likely to have experienced a population bottleneck and may have restricted population sizes, resulting in reduced genetic diversity and effective population sizes compared to their mainland counterparts. Additional effects could potentially include increased inbreeding, reduced adaptive potential and inefficient selection. Previous studies comparing the molecular evolution of island and mainland species have focused on summary statistics based on species divergence. We improved upon this approach by using both polymorphism and substitution data to address our predictions, allowing us to explicitly investigate levels of genetic diversity. Surprisingly, we found no evidence that island species experience long-term reductions in effective population size. Island species were found to have significantly less genetic diversity than mainland species; however, this pattern could be attributed to a proportion of island species that had undergone a recent population bottleneck. When these species were excluded, island and mainland species had similar levels of genetic diversity; there was also no difference between island and mainland species in terms of effectiveness of selection or mutation rate. Therefore we have no evidence to suggest that island colonisation has lasting impacts on molecular evolution.

11.21

Adaptive Evolution of a Clinal Inversion Polymorphism in *Drosophila*

Martin Kapun [1], Chloé Schmidt[1], Jérome Goudet[1], Paul Schmidt[2], Thomas Flatt[1]
[1] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland, [2] Department of Biology, University of Pennsylvania, Philadelphia, USA

In(3R)Payne, a common cosmopolitan inversion in *D. melanogaster,* exhibits steep latitudinal clines across multiple continents, but whether this pattern is due to spatially varying selection or demography remains unclear. To address this fundamental question we first estimated inversion frequencies by applying diagnostic marker SNPs to pool-seq data from 10 populations along the North American east coast. By comparing our estimates to published records we find that the inversion cline has remained stable for over thirty years. Consistent with this cline being maintained by selection, Q_{ST}-F_{ST} analysis suggests that *In(3R)P* evolves non-neutrally. In further support of an adaptive scenario, SNP-wise regression reveals numerous inversion-associated alleles differentiated in parallel between the North American and Australian clines. Second, we sequenced pools of isochromosomal lines of *In(3R)P* karyotypes from the endpoints of the North American cline and identify major inversion-specific genetic differentiation among karyotypes. Finally, again paralleling observations from Australia, we find that *In(3R)P* makes a major contribution to the well-known body size cline in North America. Together, these results provide compelling evidence that *In(3R)P* is maintained by spatially varying selection across multiple continental clines.

11.22

Spatially explicit analysis of harbor porpoise (*Phocoena phocoena*) genome-wide SNPs improves population resolution in North and Baltic Seas

Ljerka Lah [1], Daronja Trense[1], Þorvaldur Gunnlaugsson[2], Christina Lockyer[3], Ursula Siebert[4], Harald Benke[5], Anna Roos[6], Iwona Pawliczka[7], Krzysztof Skóra[7], Santiago Lens Lourido[8], Bayram Öztürk[9], Ayaka Öztürk[9], Per Berggren[10], Gísli Víkingsson[2], Ralph Tiedemann[1]
[1] Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [2] Marine Research Institute, Reykjavík, Iceland, [3] North Atlantic Marine Mammal Commission, Tromsø, Norway, [4] Institute for Terrestrial and Aquatic Wildlife Research (ITAW), University of Veterinary Medicine Hannover, Büsum, Germany, [5] Deutsches Meeresmuseum, Stralsund, Germany, [6] Swedish Museum of Natural History, Stockholm, Sweden, [7] Hel Marine Station, University of Gdansk, Hel, Poland, [8] Instituto Espanol de Oceanografia, Centro Oceanográfico de Vigo, Vigo, Spain, [9] Marine Biology Department, Faculty of Fisheries, Istanbul University, Istanbul, Turkey, [10] School of Marine Science and Technology, Newcastle University, Newcastle, UK

The population structure of the highly mobile marine mammal, the harbor porpoise (*Phocoena phocoena*) in the Atlantic shelf waters follows a pattern of significant Isolation-by-distance. Population structure of harbor porpoises from the land-enclosed Baltic Sea, which is connected with the North Sea through a series of basins separated by underwater ridges, however, is more complex. Here, we investigated the population differentiation of harbor porpoises from the populations in the Baltic Sea and adjacent waters, as well as European Atlantic shelf waters and the Black Sea. Our population genomics approach used 1801 SNPs, derived from double digest restriction-site associated DNA sequencing (ddRAD-seq), as well as 13 microsatellite loci from the same set of individuals to conduct spatial principal components analysis (sPCA). Additionally, we expanded the sample set to over 300 porpoises from the Baltic Sea and adjacent regions and analyzed the microsatellite dataset using the spatially explicit sPCA. We observed a distinct separation of the North Sea/Skagerrak population from the other Baltic Sea populations, and identified splits between porpoise populations in the southern Kattegat, the Belt Sea and the inner Baltic Sea. The improved resolution of harbor porpoise population assignments for the Baltic are important for conservation management of this endangered cetacean in threatened habitats, particularly in the Baltic Sea proper. We also show that genome-wide SNPs outperform microsatellite markers, and demonstrate the utility of RAD-tags from a relatively small sample set for population diversity and divergence analysis.

88A

Commonality analyses : Dealing with multicollinearity issues in multivariate regressions

Jérôme G. Prunier [1,5], Marc Colyn[2], Xavier Legendre[3], Kim Nimon[4], Marie-Christine Flamand[1]
[1] Institut des Sciences de la Vie, Université catholique de Louvain, Louvain-la-Neuve, Belgium, [2] CNRS-UMR 6553, Université de Rennes 1, Paimpont, France, [3] Réserve de la Haute Touche, Muséeum National d'Histoire Naturelle (MNHN), Obterre, France, [4] Department of Human Resource Development & Technology, University of Texas at Tyler, Tyler, Texas, USA, [5] CNRS USR 2936 'Station d'écologie expérimentale du CNRS à Moulis', Moulis, France

Correlative analyses in spatial genetics provide unique opportunities to describe the inherent complexity of genetic variation in wildlife species and are the object of many methodological developments. However, multicollinearity among explanatory variables is a systemic issue in multivariate regression analyses and is likely to cause serious difficulties in properly interpreting results of correlative analyses, with the risk of erroneous conclusions, misdirected research and inefficient or counterproductive conservation measures. This is all the more true in spatial genetic studies as predictors are usually derived from landscape characteristics that cannot be experimentally controlled. Using linear and logistic regressions on distance matrices, we illustrate how commonality analysis (CA), a detailed variance-partitioning procedure, can be used to deal with non-independence among spatial predictors. By decomposing model fit indices into unique and common (or shared) variance components, CA allows identifying the location and magnitude of multicollinearity, revealing spurious correlations and thus thoroughly improving the interpretation of multivariate regressions. CA has a great potential to account for complex multicollinearity patterns in spatial genetics and we strongly urge spatial geneticists to systematically investigate commonalities when performing correlative analyses.

89B

Mathematical and evolutionary constraints on F_{ST} for biallelic markers

Nicolas Alcala, Noah Rosenberg
Stanford University, Stanford, California, USA

F_{ST} is one of the most widely used statistics in population genomics, for example, to describe the genetic structure of populations and to test for natural selection. Recent theoretical studies have challenged usual interpretations of F_{ST} as being a measure ranging from 0 (genetically similar populations) to 1 (genetically different populations). For pairs of populations and biallelic markers, the maximum F_{ST} has been shown mathematically to monotonically decline from 1 to 0 as a function of the frequency of the most frequent allele. The mathematical constraints on FST in terms of allele frequencies for more than 2 populations, and the relevance of the constraints under biologically realistic scenarios, however, have not been established. We characterize the mathematical relationship between F_{ST}, the frequency of the most frequent allele M and the number of populations K. For biallelic markers, we generalize the known mathematical constraints of F_{ST} for K=2 to arbitrary K. We show that the constraints disappear when K is large. Next, in each of three migration models (island, 1-dimensional or 2-dimensional stepping stone), we show using coalescent simulations that F_{ST} strongly depends on M when migration is weak and K is small, but not when K is large. Finally, we show that our results explain patterns of human genetic differentiation, such as the 4-fold lower F_{ST} values between certain pairs of continents relative to global F_{ST} values, for high-frequency alleles. We discuss the implications of our results for the interpretation of F_{ST} and its use for tests of local adaptation.

90C

Male-biased dispersal of vampire bat rabies

Daniel Streicker[1,2], Jamie Winternitz [3], Dara Satterfield[1], Edgar Condori[1], William Valderrama[4], Andres Velasco-Villa[5], Sonia Altizer[1], Sergio Recuenco[5]

[1] Odum School of Ecology, University of Georgia, Athens, GA, USA, [2] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK, [3] Institute of Vertebrate Biology, Czech Academy of Sciences, Konesin, Czech Republic, [4] Office of Risk Assessment and Epidemiological Surveillance, National Service of Agrarian Health, Ministry of Agriculture, Lima, Peru, [5] Poxvirus and Rabies Branch, Division of High Consequence Pathogen and Pathology, Centers for Disease Control and Prevention, Atlanta, GA, USA

Identifying key groups associated with the spatial spread of pathogens improves prospects for their control but is commonly limited by scarce data on movements of infected individuals across landscapes. By combining population genetic analyses of maternally and biparentally inherited host markers and rapidly-evolving viral sequences, we demonstrate strong sex bias in the spatial spread of rabies virus in vampire bats (Desmodus rotundus) in Peru, whereby dispersing infected males seed outbreaks in genetically isolated female populations, triggering lethal cross-species infections in humans and domestic animals. This ecological process, in turn, delimits the geographic distribution viral lineages. Female philopatry and male dispersal is the most common life history strategy in wild mammals. Our study provides a simple and generalizable genetic snapshot approach to assess how such heterogeneities influence the spatial spread of pathogens that could be used to more effectively target groups for interventions in a variety of wildlife disease systems.

91D

Landscape scale assessment of genetic structure within a wild population reveals the role of an anthropogenic factor in shaping immune gene variation

Lewis Spurgin, Catalina Gonzalez-Quevedo, Karl Phillips, Richard Davies, David Richardson
University of East Anglia, Norwich, UK

Understanding the spatial scale at which selection acts upon adaptive genetic variation in natural populations is fundamental to our understanding of evolutionary ecology, and has important ramifications for conservation. The environmental factors to which individuals of a population are exposed can vary across short distances, potentially generating localised patterns of adaptation. We used a candidate gene approach to test whether and how selection generates genetic structuring within a population of Berthelot's pipits (*Anthus berthelotii*) on Tenerife, in the Canary Islands. This population is subject to strong, fine-scale variation in climate, disease and the amount of human disturbance, providing an excellent model for adaptive landscape genetics. At a set of immune genes we found localised structuring of genetic variation - a pattern not observed at neutral loci. Importantly, we found direct associations between the prevalence of specific immune gene alleles and disease infection risk and, more surprisingly, with the presence of poultry farms. This study demonstrates how human disturbance can result in fine-scale selection within populations, and highlights the importance of considering small spatial scales when studying adaptive genetic variation.

92A

Spatial ancestry analysis identifies fine-scale population structure in global sample of 3600 exomes

Juan L. Rodriguez-Flores [1], Khalid A. Fakhro[2,3], Monica Ramstetter[1,4], Jacqueline Salit[1], Amal Robay[2], Francisco Agosto-Perez[1,4], Joel A. Malek[2], Charbel A. Khalil[2], Michelle R. Staudt[1], Jason G. Mezey[1,4], Ronald G. Crystal[1]

[1] Weill Cornell Medical College, New York, NY, USA, [2] Weill Cornell Medical College in Qatar, Doha, Qatar, [3] Sidra Medical Research Center, Doha, Qatar, [4] Cornell University, Ithaca, NY, USA

Spatial ancestry analysis (SPA) is a method for inferring the geographic origin of sampled genomes of unknown origin with respect to genomes sampled in known locations. In that sense, SPA is closely related to principal components analysis (PCA) and admixture inference (ADMIXTURE), in that it can be applied to genomes of unclear or unknown origin. In this study, combined use of SPA, PCA and ADMIXTURE analysis was applied to a dataset of over 3600 genomes sampled from Eurasia, Africa and America, including 1000 Genomes Project Phase 3 and 1133 exomes from the Middle East and North Africa (MENA), a region not previously been studied using SPA analysis. Known spatial relationships between European populations were confirmed, and by using European spatial structure as a reference point, previously unknown spatial structure in MENA populations was discovered. An Arabian population previously thought to be “pure” was sub-divided into two sub-clusters, one spatially located in Africa and a second spatially located in the Arabian Peninsula. This discovery was confirmed by PCA and ADMXITURE analysis. In addition, up to 10% MENA ancestry was found in Mediterranean European genomes. In light of recent studies presenting evidence for extensive bi-directional migration between Africa and the Middle East, these results confirm more extensive shared ancestry between African, Mediterranean European, and Middle Eastern populations than expected. The implications for human demographic history and selection of these result are explored.

93B

Phylogenomics method for ranking populations of the endangered Anadromous Atlantic Salmon (*Salmo salar*) for conservation management.

Laurélène Faye, Arne Mooers
SFU, Vancouver, BC, Canada

In an Age of Extinction, wildlife managers need tools to help conserve and manage genetic resources for the future. Wild Atlantic salmon (*Salmo salar*) is at risk across its range, and so genetic prioritization would seem prudent. Recently, Volkmann and his colleagues developed a new tool allowing the ranking of populations by conservation importance, integrating phylogenomic and environmental data (Volkmann *et al.* 2014, doi: 10.1371/journal.pone.0088945). This state of the art approach is to build phylogenetic networks and score each population by its relative distinctiveness or evolutionary isolation. We tested the applicability of such measures for safeguarding species-level genetic variation using landscape genetic simulations using the Nemo platform (Guillamne & Rougemont 2006, doi: 10.1093/bioinformatics/btl415). Then, using *S. salar* genomic-scale >3000 SNPs and 15 microsatellite data we constructed the phylogenetic network of 50 populations representing all the legally designated at-risk Designatable Units (DUs) in Canada and applied the network prioritization approach. We identify some populations that are currently under-evaluated regarding their contribution to present and future *S. salar* genetic diversities. Such populations might be prioritized both for more active research into the genetic basis of local adaptation, and for conservation by managers, immediately.

94C

Genes mirror geography in *Daphnia magna* , too!

Peter Fields [1], Céline Reisser[2,3], Marinela Dukić[1], Christoph Haag[2,3], Dieter Ebert[1]
[1] Universität Basel, Zoologisches Institut, Basel, Switzerland, [2] Centre d'Ecologie Fonctionnelle et Evolutive – UMR 5175, campus CNRS, Montpellier, France, [3] Department of Biology, Ecology and Evolution, University of Fribourg, Chemin du Muśee, Fribourg, Switzerland

Identifying the presence and magnitude of population genetic structure remains a major consideration of evolutionary biology to understand the demographic history of a species as well as make predictions of how the evolutionary process will proceed. A previous phylogeographic study of the crustacean *Daphnia magna* suggested that, despite strong genetic differentiation among populations at a local scale, the species shows only moderate genetic structure across its European range, with a spatially patchy occurrence of individual lineages. We apply RAD sequencing to a sample of *D. magna* collected across the species' Eurasian range and analyze the data using principle component analysis (PCA) of genetic variation and Procrustes rotation to quantify spatial genetic structure. We find remarkable consistency between the first two PCA axes and the geographical coordinates of individual sampling points, suggesting that genetic differentiation is driven primarily by geographic distance. This pattern is consistent with unimpeded (i.e. no barriers, landscape or otherwise) migration at large spatial scales, despite the fragmented nature of favorable habitats at local scales. The pattern closely resembles that identified in human populations, in which genetic distances closely correspond to geographic distances to an extent that a two-dimensional PCA plot of allele frequencies reflects the geographic map of sample origins. The fact that similar patterns were observed in two species with different life histories and ecologies suggests that isolation by distance may be a main driver of genetic structure in many other species, but that the exact pattern structure may only be identifiable with whole genome-scale datasets.

95D

Khoisan hunter-gatherers have been the largest population throughout most of modern human demographic history.

Hie Lim Kim [1], Aakrosh Ratan[2], George H. Perry[3], Alvaro Montenegro[4], Webb Miller[3], Stephan C. Schuster[1]

[1] Nanyang Technological University, Singapore, Singapore, [2] University of Virginia, Charlottesville, USA, [3] Pennsylvania State University, University Park, USA, [4] Ohio State University, Columbus, USA

The Khoisan people from southern Africa maintained ancient lifestyles as hunter-gatherers or pastoralists up to modern times, though little else is known about their early history. We sequenced the complete genomes of five Namibian Khoisan individuals and one Bantu-speaking agriculturist from southern Africa to an average coverage of 27~55-fold. Population genetic analyses using a 420K SNP dataset from 1,462 worldwide individuals demonstrate that two genomes from the Ju/'hoansi (northern Khoisan) population contain exclusive Khoisan ancestry. The two genomes allow us to infer early history of major modern human populations. In order, coalescent analysis was applied to the six southern African genomes we sequenced and eight publicly available whole genomes from various populations. Our analyses reveal that the Khoisan and their ancestors have maintained their large effective population size following their split from non-Khoisan populations ~100-150 kya. In contrast, the ancestor populations of the non-Khoisans, including Bantu speakers and non-Africans, dramatically declined after the split and lost more than half of their genetic diversity. This is in stark contrast to the current census size of the Khoisan hunter-gatherers, which is drastically smaller compared to that of the Bantu-speaking population. Based on paleoclimate records and models, it is predicted that the precipitation in southern Africa increased ~80-100 kya while west-central Africa became drier. We hypothesize that these climate differences in central and southern Africa might be related to the divergent ancestral population histories among modern humans.

96A

A novel analysis of genetic population structure using network communities

Gili Greenbaum
Ben-Gurion University, Midreshet Ben-Gurion, Israel

Understanding genetic structure of populations is a fundamental question in ecology and evolution. Classic methods of analyzing population structure, such as F-statistics and Bayesian clustering, are widely used by population geneticists, however, these methods usually employ equilibrium assumptions such as Hardy-Weinberg or migration-drift equilibria, assumptions that are often violated in real populations. Network theory has recently been used to explore complex structures in genetics at the gene level (e.g., gene-regulation networks) and at the population-subpopulation level (e.g., population graphs). I present a novel population-individual network approach to detect and analyze genetic population structure. First, a network is constructed from individual's genetic data by calculating frequency-weighted allele-sharing relatedness between all dyads. Then I show that the network notion of "community", a group of nodes connected more densely than would be predicted by a random null model of the network, can be used to describe sub-populations in these constructed networks. Under this framework, community detection methods can be used to detect genetic population structure, and further insights, for example regarding interaction patterns between subpopulations, can be gained by evaluating the strength in which each individual belong to its "community". These methods are demonstrated using simulated genetic data, microsatellite data from wild populations and SNP data from human populations (Hapmap project data). This approach not only allows making use of advancements in network theory to address population genetic questions, but it also has the advantage that no equilibrium assumptions are required, making inferences regarding population structure much more robust.

97B

Life after range expansions: heterozygosity-fitness correlations and the evolution of recombination modifiers

Stephan Peischl [1,2], Laurent Excoffier[1,2]
[1] *University of Bern, Bern, Switzerland,* [2] *Swiss Institute for Bioinformatics, Lausanne, Switzerland*

Spatially explicit population-genetics models have shown that range expansions can affect the distribution of neutral genetic diversity across species' ranges through a process known as 'gene surfing'. Positively or negatively selected variants can also surf, which may lead to an excess of deleterious variants in newly colonized regions. While previous work has mainly focused on the evolution of genetic diversity during range expansions, we focus here on the evolutionary processes that occur after a species has successfully colonized its range, and how the demographic history of a population affects subsequent adaptive processes. Using individual based simulations, we find that the intraspecific variation in the mutation load created during the expansion leads to (i) heterozygosity-fitness correlations (HFC) in newly colonized habitats and (ii) facilitates the evolution of recombination modifiers such as inversions or selfing-incompatibility alleles. If deleterious mutations are (partially) recessive, HFC occur immediately after the expansion. In contrast, HFC occur with a lag when mutations are co-dominant. Importantly, HFC are transient phenomena after range expansions that are generally not observed in spatially structured populations at equilibrium. Using a simple analytical model, we investigate how recombination modifier can spread adaptively after range expansions. We predict that the frequency of inversions and the rate of selfing should increase along the expansion axis. Finally, we link our theory to empirical patterns of HFC found in many natural populations that have recently expanded their range, and discuss methods to test our predictions for the evolution of inversions.

98C

Towards testing drivers and limits of radiations using population genomics

Marylaure de La Harpe[1], Margot Paris[1], Oriane Loiseau[2], Ingrid Olivares[3], Anna Weigand[3], Walter Till[4], Thorsten Krömer[5], Leonardo Versieux[6], Gustavo Martinelli[7],Michael Kessler[3], Nicolas Salamin[2], Christian Lexer[1]

[1] University of Fribourg, Dep. of Biology, Unit of Ecology & Evolution, [2] University of Lausanne, Fac. of Biology and Medicine, Dep. of Ecology & Evolution, Le Biophor, [3] University of Zurich, Institut für Systematische Botanik, [4] University of Vienna, Institute of Botany, University of Vienna, [5] Universidad Veracruzana, Centro de Investigaciones Tropicales (CITRO), México, [6] Universidade Federal do Rio Grande do Norte, Centro de Biociências, Dept. de Botânica e Zoologia, Brasil, [7] Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Brazil

The adaptive radiation of bromeliads is one of the most diverse and enigmatic of the Neotropics. The Tillandsioideae subfamily represents a pertinent system for studying molecular signtures of adaptation. Our main research aim is to unravel the drivers and limits of diversification at micro-evolutionary time scales. Tillandsioid taxa were sampled along elevation gradients of various Central and South American mountain regions. Common species (geographic replicates) and local endemics were defined based on their natural distribution ranges and ecological preferences. Geographic replicates were chosen preferentially to disentangle the likely evolutionary mechanisms responsible for speciation and species cohesion. A genomic approach, combining RNA-sequencing and restriction site Associated DNA sequencing (RAD-seq), will generate genetic data for both neutral and non-neutral regions of the genomes of these taxa. The genomic information yielded by this approach will be employed to characterize gene flow and allelic diversity within and among populations occurring along various ecological gradients. Great emphasis will also be given to analyses of population divergence to shed light on the genetic mechanisms responsible for species cohesion and persistence, targeting at putative molecular signature of adaptation. Current results and key points from this research are highlighted, including spatial distributions of the sampled taxa, proof-of-concept data from RAD-seq, and available sampling designs towards reaching our research goals.

99D

Exact calculation of the joint site frequency spectrum for isolation with migration models

Andrew Kern [1], Jody Hey[2]
[1] *Rutgers University, Piscataway, NJ, USA,* [2] *Temple University, Philadephia, PA, USA*

Population genomic datasets collected over the past decade have spurred interest in developing methods that can utilize massive numbers of loci for inference of demographic and selective histories of populations. The site frequency spectrum (SFS) provides a convenient framework for such analysis and accordingly much attention has been paid to predicting theoretical expectations of the SFS under a number of different models. However, to date, exact solutions for the joint SFS of two or more population under models of migration and divergence have not been found. Here we present a novel Markov chain representation of the coalescent on the state space of the joint SFS that allows for rapid, exact calculation of the joint SFS under generalized isolation with migration (IM) models. In turn, we show how our Markov chain method, in the context of composite likelihood estimation, can be used for accurate inference of parameters of the IM model using SNP data. Lastly, we apply our method to recent whole genome datasets from *Drosophila melanogaster*.

100A

Local adaptation by small-effect alleles: persistent phenotypic divergence but transient genetic architecture

Sam Yeaman [1,2]

[1] University of British Columbia, Vancouver, British Columbia, Canada, [2] University of Calgary, Calgary, Alberta, Canada

Population genetic models predict that alleles with small selection coefficients will be swamped by migration and will not contribute to local adaptation. But if most alleles contributing to standing variation are of small effect, how does local adaptation proceed? Here, I review predictions of population and quantitative genetic models and use individual-based simulations to illustrate how the architecture of local adaptation depends on the genetic redundancy of the trait, the maintenance of standing genetic variation (V_G), and the susceptibility of alleles to swamping. Even when population genetic models predict swamping for individual alleles, considerable local adaptation can evolve at the phenotypic level, if there is sufficient V_G. However, in such cases, the underlying architecture of divergence is transient: F_{ST} is low across all loci and no locus makes an important contribution for very long. Because this kind of local adaptation is mainly due to transient frequency changes and allelic covariances, these architectures will be difficult, if not impossible to detect using current approaches to studying the genomic basis of adaptation. Even when alleles are large and resistant to swamping, architectures can be highly transient if genetic redundancy and mutation rates are high. These results suggest that drift plays a critical role in shaping the architecture of local adaptation, both through the erosion of V_G and the turnover of polymorphisms with redundant phenotypic effects.

101B

Contrasting patterns in the high-resolution variation of uniparental markers in European populations highlight very recent male-specific expansions

Chiara Batini [1], Pille Hallast[1], Daniel Zadik[1], Pierpaolo Maisano Delser[1,2], Andrea Benazzo[3], Silvia Ghirotto[3], Eduardo Arroyo-Pardo[4], Gianpiero L. Cavalleri[5], Peter de Knijff[6], Berit Myhre Dupuy[7], Heidi A. Eriksen[8,9], Turi E. King[1], Adolfo López de Munain[10], Jelena Milasin[11], Andrea Novelletto[12], Horolma Pamjav[13], Antti Sajantila[14,15], Aslıhan Tolun[16], Bruce Winney[17], Mark A. Jobling[1]

[1] University of Leicester, Leicester, UK, [2] Museum National d'Histoire Naturelle, Paris, France, [3] University of Ferrara, Ferrara, Italy, [4] Complutense University, Madrid, Spain, [5] The Royal College of Surgeons in Ireland, Dublin, Ireland, [6] Leiden University Medical Centre, Leiden, The Netherlands, [7] Norwegian Institute of Public Health, Oslo, Norway, [8] University of Oulu, Oulu, Finland, [9] Utsjoki Health Care Centre, Utsjoki, Finland, [10] University of the Basque Country, San Sebastián, Spain, [11] University of Belgrade, Belgrade, Serbia, [12] Tor Vergata University, Rome, Italy, [13] Institute of Forensic Medicine, Budapest, Hungary, [14] University of Helsinki, Helsinki, Finland, [15] University of North Texas, Fort Worth, Texas, USA, [16] Bogaziçi University, Istanbul, Turkey, [17] University of Oxford, Oxford, UK

The peopling of Europe has been divided into three main phases encompassing the initial colonization during the Upper Paleolithic, the migrations influenced by the climatic changes of the Last Glacial Maximum and the spread of agriculture during the Neolithic. However, the contribution of events during these different phases to the peopling of the area has been subject of a long-term debate. In particular, special focus has been given to determining the proportions of Europeans descending from Neolithic farmers and Paleolithic hunter-gatherers, colonizing Europe from the Middle East from 10 and 40 thousand years ago (KYA), respectively. Here we analyse a dataset of 340 individuals from 17 European and Middle Eastern populations, resequencing 3.7 Mb of the Y chromosome (MSY) and 16.6 kb of the mitochondrial genome (mtDNA). Comparing these two markers we highlight contrasting patterns in the matrilineal and patrilineal genetic histories. On one hand, the distribution of variation of mtDNA suggests pre-Neolithic expansions throughout Europe, with little influence of subsequent peopling waves. On the other hand, the MSY variation points to a very recent expansion in European populations, dating back only to the Bronze Age. This indicates a widespread male-specific post-Neolithic phenomenon which could be explained by a change in the social structure of human populations at the time.

102C

Ecological distribution and demo-genetic modelling provides novel insights into the evolutionary biogeography of *Amborella trichopoda*

Rémi Tournebize[1], Yves Vigouroux[1], Stéphanie Manel[3], François Munoz[2], Alexandre de Kochko[1], Valérie Poncet [1]

[1] IRD, Montpellier, France, [2] French Institute of Pondicherry, Pondicherry, India, [3] E.P.H.E, Montpellier, France

The role and the impact of paleoclimatic refugia on patterns of tropical biodiversity are still largely debated. This subject has been renewed with the increasing accessibility of ecological niche modelling and of genetic analysis based on massive genomic datasets. We used both approaches to elucidate the diversity pattern of *Amborella trichopoda*, Baill. (Amborellaceae), the sole living taxon of the sister group to all other extant angiosperms and which is endemic to New Caledonia. From paleoclimatic species distribution inference at the last glacial maximum (21,000 years BP) and at the mid-Holocene (6,000 years BP), we propose the existence of different possible refugia in New Caledonia. Using both microsatellites and a large SNP dataset, we performed Approximate Bayesian Computation (ABC) and maximum likelihood analyses to assess whether these putative refugia could explain the observed phylogeographical pattern of *Amborella*. We demonstrated that the strong genetic structure observed among *Amborella* extant populations is compatible with vicariant events across two past refugia. These refugia lasted till the mid-Holocene, longer that previously suggested by sole paleoclimatic niche models. Our results highlight the strong impact that paleo-refugia had on tropical diversity. Moreover, combination of demo-genetic historical inferences based on molecular data and paleoclimatic niche modelling seems a promising approach to address this question more broadly.

103D

planarPart: an efficient and robust approach for genetically-based geographic assignment

Hákon Jónsson [1], Gilles Guillot[3], Rasmus Nielsen[2], Ludovic Orlando[1]
[1] Centre for GeoGenetics, University of Copenhagen, Copenagen, Denmark, [2] Department of Applied Mathematics and Informatics,Technical University of Denmark, Lyngby, Denmark, [3] Department of Integrative Biology, University of California, Berkeley, USA

The geographic separation between populations explains a significant proportion of the human genetic variation. Patterns of genetic variation can be exploited to provide insights about past population histories, illuminating events such as migrations and admixture. Additionally, the spatial structure within genetic datasets can be used to geographically assign individuals based on their genetic profiles, with wide ranging applications from genetic management to forensics. Here, we devised a scoring system measuring the genetic affinity of individuals to different geographical locations and model the score in a kriging framework. This allows us to interpolate the score across the geographical range and find the spatial position which maximizes the genetic affinity. This approach is implemented in a new tool, planarPart, and appears to outperform current alternatives in terms of accuracy and computational performance both at local and global scales. It can position individual in 51 minutes, using a reference set of 818 individuals and, 627,719 genotypes, without restricting assumptions about the shape of the allele frequency surface. Our approach is therefore tailored to large-scale datasets such as those now generated by high-throughput sequencing platforms. Our scoring scheme is applicable on the haplotypes, allowing for estimating the spatial origin of the haplotypes present in the admixed populations, efficiently tapping into the information about our spatial ancestry and history encapsulated in the nuclear genome. Finally, we show the potential of the approach by applying it to an extensive panel of ancient DNA data sets and confirming the population affinities described earlier using other methods.

104A

Balancing selection under cyclic changes of environments that vary in magnitude over space

Davorka Gulisija[1,3], Yuseob Kim [2]

[1] University of Wisconsin, Madison, WI, USA, [2] Ewha Womans University, Seoul, Republic of Korea, [3] University of Pennsylvania, Philadelphia, USA

In classical studies of balancing selection, the plausibility of long-term balanced polymorphism maintained by temporally oscillating selection was debatable. We demonstrate here that balanced polymorphism arises under a highly plausible condition of temporally oscillating selective pressure that varies in magnitude over space. A simple model is built with two subpopulations that exchange migrants, one experiencing large-magnitude oscillation of selection and the other under the same oscillation cycle but of reduced magnitude. Using forward-in-time computer simulations and mathematical analysis, we show that alleles favored during alternative periods can coexist due to negative frequency-dependent selection, even though the basal assignment of allelic fitness in our haploid model is not frequency-dependent. With large difference in the magnitudes of oscillation between subpopulations, frequent migration, and quasi-neutrality, long-term polymorphism that is extremely stable can arise. This mechanism of balancing selection is a special case of storage effect, first proposed by Peter Chesson to describe the ability of heterogeneous environment to maintain species richness. We explain that the storage effect applied to our model results from non-linear combination of allelic fitness over time and space. With the growing list of long-term balanced polymorphism in natural populations discovered by population genomics studies, this mechanism of balancing selection may be responsible for some of those observations.

105B

Population structure in the recently split wild tomatoes S. chilense and S. peruvianum along the western coast of South America

Ian Bedddows, Aparna Reddy, Thorsten Kloesges, Laura Rose
Heinrich Heine University, Duesseldorf, Germany

Wild tomatoes (Solanum L. sect. Lycopersicon (Mill.) Wettst.; Solanaceae) form a monophyletic clade of (currently) 13 recognized species native to western South America. The section shared a common ancestor less than 3 million years ago (Mya), and two highly polymorphic self-incompatible sister species, S. chilense and S. peruvianum, split only around 0.6 Mya. S. chilense and S. peruvianum are both found along the Pacific coast from sea level to 4,000 meters in the Andean mountains, and they share an overlapping range in northern Chile to southern Peru. Both species are found in coastal deserts, and are thereby limited to alluvial deposits of river valleys and fog oases where they meet with periodic moisture. These geographical parameters likely limit pollen flow among subpopulations in both species, but overall population substructure is not well understood. Although substantial prezygotic gene flow barriers exist, previous genetic studies have detected historical gene flow between S. chilense and S. peruvianum. We are currently testing hypotheses of gene flow and population structure using next-generation sequence data. We sequenced 36 sect. Lycopersicon genotypes from a species-wide sampling of S. chilense (18) and S. peruvianum (18), plus two outgroup genotypes from neighboring sections Juglandifolium and Lycopersicoides (paired-end RNA on Illumina HiSeq 2500 platform). More than 2.9x109 ~100 bp paired-end reads were mapped onto the cultivated tomato genome. Following Bayesian SNP calling using ANGSD, we used a subset of fully represented synonymous SNPs to study population structure and gene flow within and between S. chilense and S. peruvianum.

106C

Across Language Families: Genome Diversity Mirrors Linguistic Variation within Europe

Giuseppe Longobardi[1,2], Silvia Ghirotto[3], Cristina Guardiano[4], Francesca Tassi [3], Andrea Benazzo[3], Andrea Ceolin[1], Guido Barbujani[3]

[1] Department of Language and Linguistic Science, University of York, York, UK, [2] Department of Humanities, University of Trieste, Trieste, Italy, [3] Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy, [4] Department of Communication and Economics, University of Modena-Reggio Emilia, Modena-Reggio Emilia, Italy

Comparing genetic and linguistic diversity may cast light on both demographic history and cultural transmission. However, classical studies were hampered, on the linguistic side, by insufficient reliance on quantitative tools and by the impossibility to compare the vocabularies of distantly-related languages. Here, we take advantage of two new tools recently proposed in comparative linguistics: first, a refined list of Indo-European cognate words for quantitative experiments, then a novel method of language comparison based on syntactic features. Since the latter method estimates linguistic diversity from a universal inventory of grammatical polymorphisms, it enables comparison even across different language families. On these grounds, by comparing a broad genome-wide SNP dataset in 15 European populations, we observed significant correlations between genomic and linguistic diversity, the latter inferred from data on both Indo-European and non-Indo-European languages. Contrary to previous observations, on the European scale, language proved a better predictor of genomic differences than geography, and inferred episodes of genetic admixture following the main population splits found convincing correlates also in the syntactic realm, supporting the relevance of a synthesis approach to cultural and biological evolution. These results pave the ground for previously unfeasible cross-disciplinary analyses at the worldwide scale, encompassing populations of different language families.

107D

Linking Species Distribution Models to Landscape Genetics: Comparison of landscape effects on gene flow in three diverging *Thymelicus* skippers

Jan O. Engler [1,2], Dennis Rödder[1], Jan Christian Habel[3], Niko Balkenhol[2]
[1] Zoological Research Museum Alexander Koenig, Bonn, Germany, [2] Department of Wildlife Sciences, University of Göttingen, Göttingen, Germany, [3] Department of Ecology and Ecosystemmanagement, Technical University Munich, Freising-Weihenstephan, Germany

Combining species distribution models (SDMs) with genetic data have revealed several insights in the phylogeographic history of many taxa. More recently, studies also started to incorporate SDMs for the analysis of contemporary gene-flow within a landscape genetic context. In this study, we exemplify this approach in combining SDMs with connectivity models using a comparative genetic set derived from three congeneric Skipper butterflies (Hesperidae, *Thymelicus*) with diverging ecological traits. To understand how landscape characteristics affect gene-flow in species with diverging ecological traits, it is important to analyze taxonomically related sympatric species in the same landscape using identical methods. We based our inferences on two different analytical methods and two metrics of genetic differentiation. Results indicate that land use patterns influence population connectivity in the least mobile specialist *T. acteon*. In contrast, populations of the highly mobile generalist *T. lineola* were panmictic, lacking any landscape related effect on genetic differentiation. In the species with intermediate ecological traits, *T. sylvestris*, climate has a strong impact on inter-population connectivity. Our results show that closely related species with gradually changing ecological traits also show genetic structures and landscape genetic relationships that gradually change from a geographical macro- to micro-scale. Thus, the type and magnitude of landscape effects on gene-flow can differ strongly even among closely related species inhabiting the same landscape, and depend on their relative degree of specialization and mobility.

108A

Examining allele frequency profiles underlying clinal trait variation in *Pinus sylvestris*

Sonja Kujala [1,2], Yongfeng Zhou[1,2], Outi Savolainen[1,2]
[1] *University of Oulu, Genetics and Physiology, Oulu, Finland,* [2] *University of Oulu, Biocenter Oulu, Oulu, Finland*

Allele frequency clines and/or allelic covariation are expected to be found in loci underlying clinal trait variation, such as timing of budset in Scots pine (*Pinus sylvestris*). Population differentiation in the continuous distribution of Scots pine is negligible and allele frequency clines at neutral loci are not expected to be common. We have studied the prevalence of allele frequency clines in a set of SNPs genotyped in 10 European populations of *P. sylvestris* originating from different latitudes. Also within population linkage disequilibrium was examined. Among our data we have 110 SNPs located within candidate genes for timing of budset and stress tolerance. 241 reference SNPs originate from genes not expected to be involved in clinal adaptation. Overall the allele frequency clines were shallow; regression coefficients (frequency versus latitude) varied between -0.013 and 0.041 and were distributed normally around zero. We find no evidence of enrichment of candidate gene SNPs among the steepest regression slopes. Furthermore, the regression coefficients of 22 SNPs found to be associated with timing of budset in our recent analysis had values from the center of the distribution. Linkage disequilibrium (D') was found to decay within less than 1 kb. Within this data set, we see no evidence for steep allele frequency clines in loci underlying clinal trait variation. The prevalence of selection acting on beneficial allelic associations should therefore be studied in more detail in Scots pine, especially in the northern part of the distribution where selection pressure for timing of growth cessation is more intense.

109B

From social to genetic structures: a genome-wide approach in Southeast Asia

Goki Ly [1], Romain Laurent[2], Sophie Lafosse[1], Frederic Bourdier[5], Olivier Evrard[4], Gerard Diffloth[3], Samuel Pavard[1], Raphaelle Chaix[1]

[1] CNRS-MNHN-University Paris Diderot-Sorbonne Paris Cite, UMR7206 Eco-Anthropology and Ethno-Biology, Paris, France, [2] Population Genetics group, Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [3] Retired researcher, Siem Reap, Cambodia, [4] IRD-MNHN, UMR208 Patrimoines Locaux et Gouvernance, Paris, France, [5] INSERM-IRD-University of Aix-Marseille, UMR912 SESSTIM, Marseille, France

Ethnologists have described the impressive complexity of kinship systems in human populations, and in particular the complex rules of alliance (which determine mate choice), rules of descent (which affiliate individuals to kin groups), and rules of residence (which indicate where married couples should settle down). The kinship system of a population should be of particular interest to population geneticists since it conditions when, where and with whom individuals reproduce and raise their children, impacting gene pool evolution. For the past ten years, geneticists have intended to understand how these rules affect the genetic diversity of human populations focusing mainly on uniparentally inherited Y chromosome and mtDNA. In order to assess the influence of kinship systems at the genome-wide level, we sampled ethno-demographic and genetic data (598,764 SNPs) from 535 households belonging to 12 Southeast Asian ethnic groups, which exhibit different kinship systems. Runs of homozygosity (ROH) have been detected using a likelihood ratio approach and classified in three length categories (short, medium and long). We showed that the average size of long ROHs is greater in matrilineal groups than in patrilineal groups (3874kb versus 3056kb). This difference could be explained by the higher level of village endogamy in matrilineal groups in comparison to patrilineal groups, as highlighted by our analysis of ethno-demographic data. The higher level of village endogamy in matrilineal groups could result from the higher will of men in matrilineal ethnic groups to stay in their natal village in order to inherit position of influence inside their village.

110C

Unraveling the microgeographic structure of Southwestern Europe through refined Y-chromosome phylogeny

Neus Solé-Morata, Neus Font, David Comas, Francesc Calafell
Departament de Ciències Experimentals i de la Salut, Institute of Evolutionary Biology (CSIC), Universitat Pompeu Fabra, CEXS-UPF- PRBB, Barcelona, Spain

Y-chromosome haplogroup R1b-P312 is the most frequent patrilineal lineage in Western Europe, and plays an important role in the understanding of historical events within Europe. However, little is known about one of its three subclades, namely DF27 (Rocca et al. 2012), which is supposed to be found at a high frequency in Iberia, as estimated from private gene genealogy studies. Moreover, previous studies have reported a high Y-STR haplotype homogeneity within one of the DF27/Z195 subranches (Z220, Z278 and M153), which has been attributed to the recent, rapid radiation of R-M269 (Larmuseau et al. 2014). We have collected over 2500 male samples from Catalonia, the Balearic Islands, and Valencia (Spain), from volunteers carrying one of 50 different surnames. These samples have been typed for 70 Y-chromosome SNPs with the OpenArray technology (Martinez-Cruz et al. 2011) and 17 Y-STRs (from the AmpFlSTR Yfiler kit) with the aim that the study of new lineages within the R1b-P312 clade will shed some light on the peopling of Western Europe, and particularly regarding the population replacement posited to have taken place in the Bronze Age (Haak et al. 2015).

111D

The genetic impact of the Viking voyages

Rita Rasteiro, Jon Wetton, Mark Jobling
University of Leicester, Leicester, UK

The Viking voyages of exploration and colonization ranged far beyond their Scandinavian homelands but their genetic contribution to the populations of distant lands is still subject to much debate both in academic literature and popular culture. A multi-disciplinary team is investigating a series of questions regarding the impact of these diasporas on the British Isles and beyond. Did they leave a genetic mark? Were these migrations male- or family-biased?
Using data from a variety of historical and archaeological sources, together with SNP chip (Y-chromosomal, mitochondrial) data, we are using model-based simulation approaches to explore the consequences of Scandinavian migrations both in the Viking age and subsequently on today's patterns of genetic diversity, and envisaging possible demographic scenarios using archaeologically-, historically- and linguistically-informed parameters.

112A

Genome-wide identification of highly differentiated polymorphisms in South-American natives

Giordano B Soares-Souza [1], Guilherme Kingma[1], Eduardo Tarazona[1], Maíra R Rodrigues[2,1]
[1] UFMG, Belo Horizonte, MG, Brazil, [2] KU Leuven, Leuven, Belgium

Genetic diversity is associated with the phenotypic differentiation among human populations. Even if most human diversity occurs within populations and the interaction with the environment is crucial to determine the phenotype, the study of variants that differentiated between human populations is essential in evolutionary studies and biomedical research. We identified 18343 polymorphisms (SNPs) from 2926 genes highly differentiated in Native American populations in respect to the West Africa, Europe and East Asia populations. These genetic variants were annotated using the database integration tool MASSA (Multi-Agent System for SNP Annotation). Based in the multi-agent technology, MASSA allows parallel and cooperative execution of tasks, bypassing the problem of the distribution, size and heterogeneity of the biological data. Through the enrichment analysis performed by MASSA, we identified enriched terms in the annotation of differentiated genes in Native Americans. These analyzes confirmed previous knowledge on the genetic structure of natives, such as evidence of positive selection in genes involved in immune response and nervous system activity.

113B

Evidences of local adaptation at clock genes polymorphisms in human populations

Irene Dall'Ara, Silvia Ghirotto, Cristiano Bertolucci, Guido Barbujani
University of Ferrara, Ferrara, Italy

Biological rhythms evolved because of environmental periodic changes, such as daylight cycles. Circadian rhythms give rise to oscillations in behaviour and physiological functions, for anticipating upcoming daily change. Evidences from model organisms showed a relationship between genes of the circadian rhythms and latitude. Using 116,000 SNPs of 25 clock and clock-related candidate genes from the 1000Genomes Project and a reference dataset of putatively neutral polymorphisms, we investigated the genetic structure at these loci. Moreover, we tested for local adaptation searching elevated population differentiation using both the FST-based outlier analysis implemented in BayeScan and the Hierarchical model within Arlequin. We also tested for significant associations between allele frequencies and important ecological variables (including latitude and photoperiod) through the bayesian method of BAYENV. Finally, we checked whether some of the significant SNPs resulting from these analyses fall within a set of 15 candidate SNPs associated with different chronotypes or sleep disorders. The global genetic structure at the candidate genes resembled what expected under neutrality. Among the 116,000 SNPs, BAYESCAN and Arlequin identified respectively 1250 and 1073 SNPs with significant evidences of local adaptations, of which 659 were identified by both methods. The 15% of all the SNPs with signatures of local adaptation are located in RORA gene. Furthermore, one of the positively selected SNPs (rs7209167, CKIδ) has been associated with a diminution of the sleep duration. Correlation with the environmental variables supports the evidence that selection of clock gene alleles based on latitudes or photoperiods could not be ruled out.

114C

Fission and fusion: The population history of a nocturnal primate

Christine Driller [1], Stefan Merker[2], Dyah Perwitasari-Farajallah[3], Hans Zischler[1]
[1] Johannes-Gutenberg University Mainz, Mainz, Germany, [2] State Museum of Natural History Stuttgart, Stuttgart, Germany, [3] Bogor Agricultural University, Bogor, Indonesia

As one of the world´s most remarkable biodiversity hotspots, Wallacea is home to a highly endemic biota. The Indonesian island of Sulawesi is the largest land mass in this biogeographic region. It underwent significant paleogeographic and -climatic change, thus offering a suitable setting for studying the impact of episodic range fragmentation on present faunal distribution. To address this issue, we investigate the phylogeography of an old endemic taxon, the Sulawesi tarsier. These small, nocturnal and arboreal primates presumably colonized the paleo-Sulawesi archipelago during the Miocene. Unraveling their speciation history can therefore reveal considerable insights into driving geological and biological forces that governed terrestrial diversification on Sulawesi over different geological time-scales. As tarsier distribution patterns partly coincide with endemism areas of other Sulawesi taxa, our findings provide support for previously proposed forest refugia on Proto-Sulawesi. Further, we show different levels of speciation in extant tarsier populations emphasizing the decisive role of timing and location of past geographic isolation on this process.

115D

Understanding the origin of diversity in rock-dwelling land snails using *Montenegrina* (Gastropoda, Pulmonata) as a model system

Katharina Jaksch [1,2], Elisabeth Haring[1], Luise Kruckenhauser[1], Helmut Sattmann[3], Zoltán Fehér[1]
[1] Natural History Museum Vienna, Central Research Laboratories, Vienna, Austria, [2] University Vienna, Department of Integrative Zoology, Vienna, Austria, [3] Natural History Museum, 3. Zoological Department, Vienna, Austria

Speciation is one of the crucial keywords in evolutionary biology and until today geographic speciation is the most common and reasonable scenario. This holds especially for organisms with low dispersal activity and a patchy distribution such as rock-dwelling land gastropods.
Here we present first phylogenetic data of the genus *Montenegrina*, which is with 88 described taxa the second largest clausiliid genus. To explore the different stages of speciation in this group we comprehensively sampled *Montenegrina* throughout its whole distribution range in the Balkan Peninsula. As a first step we reconstructed the phylogeny of this genus with multi-locus molecular analyses. Combined with morphological data this will result in a taxonomic revision of *Montenegrina*. As a next step we will assess genetic and morphometric variation within/among populations and thus test ecological niche differences between taxa and reconstruct the phylogeographic history of the genus.
Some *Montenegrina*-populations show a high variation in shell ribbing (strongly ribbed to completely smooth) and a few taxon-pairs differ exclusively in their ribbing. As populations of these taxon-pairs occur sympatrically or parapatrically they could provide indications of cases of non-allopatric speciation or potential hybridization of populations in secondary contact. Several of these taxon-pairs that co-occur in a very small distribution range are currently investigated with microsatellite markers to assess gene flow and to depict population genetic structures. For this task microsatellite markers were established for the first time in clausiliids. Eventually, this extensive study should help to answer fundamental questions of speciation in this highly diverse rock-dwelling land snail.

116A

Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to investigate spatial structures.

Sylvain Piry[1], Marie-Pierre Chapuis[2], Bertrand Gauffre [3 ,4], Astrid Cruaud[1], Karine Berthier[5]
[1] INRA, UMR 1062, Centre de Biologie pour la Gestion des Populations, 34988 Montferrier sur Lez, France, [2] CIRAD, UMR 1062, Centre de Biologie pour la Gestion des Populations, 34988 Montferrier sur Lez, France, [3] INRA, USC1339 (CEBC-CNRS), 79360 Villiers en Bois, France, [4] CNRS, UMR 7372 CEBC – Université de La Rochelle, 79360 Villiers en Bois, France, [5] INRA, UR0407 pathologie végétale, 84143 Montfavet, France

Dedicated analyses in landscape genetics require a priori knowledge on species dispersal abilities. When little is known about the species under study these methods are difficult to apply and, therefore, we rely on exploratory approaches allowing to visualize and confront genetic and environmental variation patterns. We developed a novel exploratory method, free from assumptions, to investigate spatial variations in pairwise genetic metrics computed between georeferenced samples. Graphical outputs can be mapped on landscape layers to further explore potential relationships between genetic and environmental structures. The method relies on both a spatial grid and a network for which the links are materialized by ellipsoidal polygons bearing the pairwise metric values computed between samples. The network is overlayed on the spatial grid and a given cell receives the weighted mean of all ellipses values intersecting a top this cell. MAPI includes a randomization procedure to test the relevance of the spatial structure emerging from the analysis. Using controlled simulated datasets, we showed that MAPI provides reliable and useful information, even in complex situations, by analyzing populations either panmictic or under isolation-by-distance and separated or not by a barrier to gene flow. We also illustrated MAPI using real datasets obtained from biological models with challenging demographic or biologic features.

12 Micro-evo-devo: using natural variation to explain the how and why of phenotypic evolution

12.1

Micro-evolution of nematode vulval precursor cell fates

Marie-Anne Félix, Michalis Barkoulas, Fabrice Besnard
Ecole Normale Supérieure, Paris, France

A key question in micro-evo-devo is the extent to which the genotype-phenotype map biases and constrains phenotypic evolution. A developmental system subjected to random noise or random genetic variation indeed responds in a specific manner at the phenotypic level. Thus, while the phenotypic spectrum produced by random genetic variation is further filtered by natural selection, a bias in phenotypic evolution is already provided by the distribution of available phenotypic variation.

The development of the six competent vulval precursor cells in *C. elegans* provides a model system to study the evolution of developmental processes at different evolutionary scales. Their fates and variational properties can be compared quantitatively.

At the micro-evolutionary scale, we measured the precision of vulval precursor cell fate patterning in different environments and wild *C. elegans* genotypes. The relative frequency ranges of specific variants correspond to measurable properties of the developmental system, such as sensitivity to the dose of the relevant signaling molecules. We also compared the mutational variance of each of the six precursor cells. The P3.p cell fate is the most sensitive in all respects and is that actually evolving fastest in the species and the *Caenorhabditis* genus. Mutational effects may thus produce phenotypic trends, in the absence of selection.

12.2

Sexually dimorphic pigmentation in Drosophila: lineage sorting and independent comparisons in multiple phenotypic transitions

Sarah Signor [1], Artyom Kopp[2]
[1] University of Southern California, Los Angeles, CA, USA, [2] University of California, Davis, Davis, CA, USA

Molecular and genetic analysis of phenotypic variation has revealed many examples of evolutionary transitions involving developmental pathways. Sexually dimorphic pigmentation is a relatively recently evolved phenotype, originating at the base of the D. melanogaster group from known modifications to existing enhancers. Since this time there have been multiple gains and/or losses of this phenotype. Within the D. bipectinata and D. ercepeae species subgroup we have investigated the genetic basis of coloration differences in four species pairs that vary with respect to this phenotype. What we have found is a complex story, involving gene reuse in all or most species for two of the genes involved. This gene reuse is due to both independent evolution and lineage sorting. There are also alleles of different genes that effect pigmentation sorting in various populations of these species, with ubiquitous epistasis. A population genetics analysis of one gene shared between all crosses revealed that five nucleotide differences were responsible for the phenotypic transition in two of the species pairs. These four examples of a shared phenotypic change shed new light on the role of lineage sorting and genetic hotspots in the evolution of phenotypes on short evolutionary time scales.

12.3

Gain and loss of GDF6 expression and the evolution of skeletal traits in fish and humans

Vahan Indjeian [1,2], David Kingsley[2]
[1] MRC Clinical Sciences Centre, London, UK, [2] HHMI and Stanford University, Stanford, CA, USA

Dramatic skeletal changes have evolved within most vertebrate groups. High-resolution mapping experiments with natural populations of sticklebacks identify two distinct, but very closely linked loci that regulate different aspects of armour plate size. The Growth/differentiation factor 6 (GDF6), a secreted bone morphogenetic protein, is the only gene in both intervals. Freshwater fish with reduced armour plates show elevated levels of GDF6 expression and transgenic overexpression of GDF6 in large-plated marine fish results in smaller or missing plates. Gdf6 null mice have skeletal phenotypes, and in humans there is a unique loss of conserved non-coding element in the GDF6 gene. The chimpanzee allele of the enhancer missing in humans drives expression in the hindlimbs but not forelimbs, in anatomic domains that have been specifically modified during the human transition to bipedalism. These results suggest that both upregulation and downregulation of GDF6 contribute to natural variation in skeletal morphology in fish and primates.

12.4

Natural variation at *ath-miR824* modulates the adaptive potential of *Arabidopsisthaliana*

Jinyong Hu [1,2]

[1] Max-Planck-Institute for Plant Breeding Research, Koeln, Germany, [2] Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

MiRNAs (miRNAs) play pivotal roles in post-transcriptional regulation. The production of miRNAs needs a hairpin like structure, in which the miRNA/miRNA* forms the stem. Although sequences variation within the stem significantly impacts the processing of hairpins into functional molecules, polymorphisms in other region and their contribution to adaptive evolution have rarely been investigated especially in natural populations. Here we show the potential adaptive roles of natural variation in the *miR824* precursors of *Arabidopsis thaliana* . Allelic miR824 precursor shows different structure hence distinct potential in the turnover into mature molecules. *MiR824* polymorphism varies following the functionality of *FRIGIDA* (*FRI*), one key regulator in the temperature pathway controlling flowering time variation in *Arabidopsis thaliana* . Interestingly natural variation at *miR824* locus associates significantly with flowering time variation especially in different genetic backgrounds harboring different functionality of *FRI* . Furthermore, allelic variation at *miR824* locus impacts the latitudinal cline of flowering time in *Arabidopsis thaliana* . Our results thus reveal a great potential of the *miR824* variation in mediating the adaptation to differential environmental conditions for *Arabidopsis thaliana* .

12.5

Size evolution in high-altitude Ethiopian Drosophila melanogaster: developmental decanalization and the genetics of a complex adaptive trait

Justin Lack, Amir Yassin, Dylan Braun, Matthew Monette, John Pool
University of Wisconsin-Madison, WI, USA

A primary goal in evolutionary biology is to identify genetic changes underlying adaptation. While sequencing advances have improved this ability, questions remain concerning how adaptive phenotypes evolve at the genetic level. D. melanogaster from the Ethiopian highlands (>3000 m) have evolved wing and body size larger than any previously described population. We pursued the genetic determinants of size evolution using novel genomic methods for mapping trait differences and identifying genes subject to local adaptation. Mapping results indicate the genetic basis of wing size evolution is partially distinct from body size. To refine mapping results, we examined SNPs under QTLs for signals of selection and found multiple candidate loci. These candidates suggest Insulin/Tor signaling and the Epidermal Growth Factor Receptor (EGFR) pathways were both responsible for body and wing size shifts, and subsequent changes in the EGFR pathway augmented wing size and shape. In addition to size variation, large-winged Ethiopian lines possess elevated numbers of wing vein malformations relative to other African populations. This suggests that a consequence of wing-size evolution has been the decanalization of wing developmental pathways. We conducted large-scale mutagenesis of Ethiopian and Zambian lines to test this hypothesis. Mutagenized Ethiopian lines that previously never produced wing vein mutants, produced significantly more vein mutants than identically treated Zambian lines, suggesting wing development in Ethiopian flies is more perturbable relative to typical Sub-Saharan D. melanogaster. This first-ever demonstration of decanalization impacting a recently evolved tissue raises the question of whether developmental instability is a common outcome of morphological evolution.

69A

The genetic architecture of craniofacial shape in the house mouse: lessons from genome wide association studies

Luisa F. Pallares [1], Clarissa C. Parker[2,3], Shyam Gopalakrishnan[2], Peter Carbonetto[2], Abraham A. Palmer[2], Diethard Tautz[1]

[1] Max Planck Institute for Evolutionary Biology, Ploen, Germany, [2] University of Chicago, Chicago, USA, [3] Middlebury College, Middlebury, USA

From an evolutionary perspective, it is more interesting to understand the genetic architecture of a complex trait -number of loci involved, effect size distribution, genetic vs environmental determination, heritability, etc- than to identify the specific genes responsible for the variation of the trait. However if we want to understand the developmental processes involved in the making of such a complex trait, inevitably we will need to know the genes. We use the skull and mandible of the house mouse as models to study within-population morphological variation. We make use of single SNP as well as genome-wide SNP approaches to identify the genetic variants underlying craniofacial shape variation, and to explore the genetic architecture of such traits. We performed a GWAS in an outbred population of mice (CFW). Mandible and skull shape were measured with 3D landmarks and analyzed using geometric morphometrics. Shape information was decomposed in principal component scores. Such scores and a dense SNP coverage (~93,000) were used in the mapping done in GEMMA. 6 and 16 regions were significantly associated with skull shape and mandible shape, respectively. The genome-wide SNP data gave "chip heritability" estimates of ~43% for skull and mandible shape. The results from the single-SNP and genome-wide SNPs estimates suggest a complex genetic architecture for shape traits, and support a polygenic model of inheritance. These results agree with our previously reported estimates using a wild population of mice.

70B

Multivariate analysis of genotype-phenotype maps and their evolutionary potential

Philipp Mitteroecker [1], Jim Cheverud[2], Mihaela Pavlicev[3]
[1] Univ. of Vienna, Vienna, Austria, [2] Loyola University, Chicago, USA, [3] Univ. of Cincinnati, Cincinnati, USA

Multivariate studies of genotype-phenotype association are central to biomedical science as well as evolutionary biology. Most of these studies consist of a vast number of pairwise comparisons between single genetic loci and single phenotypic variables, leading to low fractions of explained variance and a dramatic loss of statistical power. Biologically more important, most complex phenotypes are not determined by single loci, but by the joint effects, both additive and non-additive, of a number of loci. For understanding the evolutionary potential of the complex genotype-phenotype map, it is more useful to investigate the association between certain allele combinations and certain combinations of phenotypic variables that bear biological interpretation. We present a multivariate approach to identify pairs of maximally associated genetic and phenotypic patters (linear combinations). Using QTL data of two mouse populations, we show that the number of such patterns underlying the observed genotype-phenotype association is surprisingly small. We discuss the developmental origin and the evolutionary consequence of such low-dimensional genotype-phenotype maps.

71C

Developmental mechanisms of testis divergence between chimpanzees and humans

Ekin Saglican[1], Rori Rohlfs[2], Ezgi Ozkurt[1], Haiyang Hu[3], Babur Erdem[1], Philipp Khaitovich[3], Mehmet Somel [1]

[1] Department of Biological Sciences, Middle East Technical University, Ankara, Turkey, [2] Department of Integrative Biology, University of California, Berkeley, CA, USA, [3] CAS-MPG Partner Institute for Computational Biology, Shanghai, China

Chimpanzees have ~3 times smaller brains, and ~3 times larger testes compared to humans. Changes in the timing of conserved developmental processes, or heterochrony, could drive large anatomical divergence over short evolutionary timescales. Here we investigated mechanisms of testis divergence between chimpanzees and humans. Because testis developmental series are lacking for both species, we resorted to an indirect approach, using mouse developmental transcriptome series as reference. This analysis revealed that the human testis transcriptome profile most closely resembles those of adolescent mice, whereas the chimpanzee profile resembles those of mature mice. This is unique, as in the brain, both humans and chimpanzees resemble mature mice. The transcriptome data also indicated that chimpanzee testicles harbor a higher proportion of meiotic and post-meiotic germ cells. Other polyandrous mammals, such as macaques and rats, have chimpanzee-like testis profiles, whereas the testis profiles of the monoandrous gorilla and marmoset resemble those of humans. Finally, we performed network analysis to identify potential regulatory factors of differential testis development, and analysis of sequence polymorphism of testis-related genes to compare humans and chimpanzees. Our results indicate that in chimpanzees, evolution of polyandry created pressure to boost sperm quantity, increased purifying selection on testis-related genes, and led to upregulation of key testis development factors and testis overgrowth. Parallel changes in conserved developmental pathways can thus explain convergent testis evolution in polyandrous mammals.

72D

Cutilular origins of the *Thermonectus marmoratus* bifocal lens

Aaron Stahl [1], Mark Charlton-Perkins[2], Tiffany Cook[2], Elke Buschbeck[1]

[1] Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio, USA, [2] Division of Developmental Biology and Department of Pediatric Opthamology, Cincinnati Children's Hospital Center, Cincinnati, Ohio, USA

The development of novel traits has led to many interesting questions such as: how are existing gene networks altered or co-opted to give rise to a new phenotypes, and how does the new structure lead to an innovative function? The unique larval bifocal eyes (stemmata) of the sunburst diving beetle Thermonectus marmoratus are a perfect measure for these questions. The larvae form a functionally novel eye phenotype with a lens that facilitates simultaneous focusing of two images on two retinas. Based on comparative and developmental studies, it is likely that the unorthodox beetle eyes derived from individual units of an ancestral compound eye. Therefore, to begin defining how the bifocal trait may have arisen, we used a comparative approach to investigate how lenses are formed. Specifically we combined transcriptomics and proteomic approaches to identify and compare lens- producing genes from Thermonectus stemmata and Drosophila compound eyes. The results indicate that lenses of Thermonectus and Drosophila are both primarily composed of cuticular proteins (3/4 in Drosophila, and 7/11 in Thermonectus), but also contain other proteins that are not well characterized. In addition, expression analysis reveals that in Thermonectus, some of the lens proteins are expressed most strongly in cells that seemingly contribute to the periphery of the lens, whereas others are more evenly expressed. Accordingly, “bifocalness” may have arisen from a change in expression patterns of specific cuticular proteins in related but distinct cells that contribute to the periphery or center of the lens.

73A

Arguments in favor of Lancelet phylogeny from a predecessor chordate with mesolecithal oocyte, and of the midgut diverticulum origin from a yolk sac

Vladimir Subbotin
Arrowhead Research Corporation, Madison, Wisconsin, USA

Evolutionary morphology implies that any organs of living animals must descend, with great or small modifications, from homologous organs of their common ancestor. Amphioxus possesses the greatest homology puzzle: a midgut diverticulum. All representatives of the subphylum Cephalochordata develop unique morphologic interactions between visceral circulation and a part of intestine. In Lancelet, venous blood from the caudal intestine is collected into an unpaired subintestine vein, which breaks into a capillary network surrounding the midgut diverticulum, and then again is collected into *vena Cardinales posterior*. Such a capillary net between two veins, surrounding a derivative of intestine, is unique to Lancelet and it is homologous to the portal vein/liver pattern in Craniata. However, Lancelet does not possess a liver. What could be a homologous phylogenetic precursor of the midgut diverticulum with unique vasculature? There is no factual answer; however, there is a theoretical solution. Imagine a midgut diverticulum, surrounded by skin with feeding and draining vessels, is protruded down from Lancelet. This midgut diverticulum, with its unique vascular architecture, would be homologous to a yolk sac. But Lancelet has a microlecithal oocyte and no yolk sac. These facts allow one hypothesis – a homologous precursor of the diverticulum is a yolk sac of an ancestral notochordal animal, which became internalized in phylogeny of early Cephalochordates. I hypothesize that: (i) Lancelet descended from a Cephalochordate-like notochordal ancestor with mesolecithal oocyte and yolk sac, which became midgut diverticulum in ontogeny of early Cephalochordates; (ii) internalized yolk sac evolved into liver of vertebrates.

74B

Mapping selection onto embryo development in Drosophila

David Castellano[1], Marta Coronado [1], Irepan Salvador[2], Antonio Barbadilla[1], Isaac Salazar[1,2]
[1] Universitat Autònoma de Barcelona, Department of Genetics and Microbiology/Institut de Biotecnologia i Biomedicina, Cerdanyola del Vallès (Barcelona), Spain, [2] University of Helsinki, Centre of Excellence in Experimental and Computational Developmental Biology Research, Helsinki, Finland

The integration of population genomic and developmental omics data is a necessary step to discover microevolutionary patterns on embryonic development. In this work, we combine available expression data from *Drosophila melanogaster* to categorize coding genes according to their temporal expression patterns and to define 40 spatiotemporal subregions across embryogenesis. We test whether natural selection is affecting these gene categorizations differentially. Applying the DFE-α estimator to *D. melanogaster* polymorphism data and divergence out to *D. yakuba*, the role of natural selection has been inferred. We find a strong negative correlation between developmental timing and the rate of both adaptive and non-adaptive substitutions. Interestingly, the average intron length increases consistently along the development. The intron delay hypothesis, which states that the developmental timing of a gene is governed by the length of their primary transcripts, could account for this observation. We find that the average intron length is positively correlated to the levels of neutral polymorphism (P_s) and negatively correlated to the constraint on segregating sites (P_a/P_s). These results suggest that the gene's intron size could act as a modifier of the Hill-Robertson interference (HRi) and accordingly, early expressed genes would accumulate more non-adaptive substitutions when exons are close together. The fact that early expressed genes exhibit also a higher adaptation rate suggests a complex dynamics for adaptive substitutions. Finally, we show that the embryonic subregions with the highest adaptation are the precursors of the nervous system and the germ band.

75C

Morphogenetic characterization, dating of divergence and evolutionary relationships of malaria vectors Anopheles cruzii and Anopheles homunculus

Camila Lorenz [1,2], José S. L. Patané[1], Lincoln Suesdek[2,3]

[1] Laboratório de Parasitologia - Instituto Butantan - Avenida Vital Brasil, 1500, São Paulo, CEP 05509–300, São Paulo, São Paulo, Brazil, [2] Biologia da Relação Patógeno-Hospedeiro- Instituto de Ciências Biomédicas, University of Sao Paulo. Av. Prof. Lineu Prestes, 2415, São Paulo, CEP 05508-000, São Paulo, São Paulo, Brazil, [3] Instituto de Medicina Tropical – Avenida Dr. Enéas Carvalho de Aguiar, 470, São Paulo, CEP 05403-000, São Paulo, São Paulo, Brazil

Anopheles cruzii and Anopheles homunculus are vectors of etiological agents of malaria in southeastern Brazil, a region known to be a major epidemic spot for autochthonous malaria. Our aims were to (1) enrich the diagnosis of species, considering that the females of the two species are isomorphic, (2) investigate whether there are differences in their genetic composition and morphology, (3) estimate the evolutionary time of divergence between the two species and (4) infer their phylogenetic history altogether with other Anophelinae. We used the wing geometry as morphological markers and used the sequences of the mitochondrial cytochrome oxidase subunit I gene to estimate levels of polymorphism. We also used the genes white, 28S, ITS2, Cytb, and COI in our phylogenetic and dating analyses. The comparison of wing thin-plate splines between the two species showed obvious differences in wing venation. An. cruzii showed higher haplotype diversity, with many rare haplotypes that were displayed by only a few specimens. Phylogenetic analyses revealed that all tree topologies converged and indicated An. bellator as sister taxon to An. homunculus and An. laneanus as sister to An. cruzii. Diversification within the subgenus Kerteszia was 2 to 14.2 million years ago. Despite being morphologically and ecologically related, An. cruzii and An. homunculus show consistent differences. An. cruzii showed greater natural diversification than An. homunculus. Phylogenetic analysis reveals that species are not sister-groups but species that recently diverged within the Kerteszia group, probably concomitantly with the radiation of bromeliads in South America or during the Pleistocene climate change.

76D

Genotype-Phenotype maps: Exploring evolvability and robustness by combining a multilinear framework with Boolean networks

Christine Mayer, Thomas F. Hansen
University of Oslo, Oslo, Norway

Evolvability and robustness help us to understand evolutionary change and the origination of evolutionary innovations and novelties, but the connection between them is a still unanswered question subject to debate. We are using a model based on a multilinear framework combined with the idea of Boolean networks to simulate different genotype-phenotype maps using logical Boolean operations. This combined approach demonstrates that the relationship between evolvability and robustness is dependent on the definition of the underlying genotype-phenotype map. Evolvability and robustness can be positive correlated, as well as negative. Hence, the direction of the correlation is dependent on the complexity and definition of the underlying genotype-phenotype map based on the used logical Boolean operations.

77A

Transcriptome profiling of cave and surface populations of the amphipod Gammarus minus

David Carlini
American University, Washington, DC, USA

The cave realm is an ideal habitat for studying the evolution and genetics of regressive morphological traits. *Gammarus minus*, a freshwater amphipod living in the cave and surface streams in the eastern United States, provides an opportunity to examine the evolution of troglomorphic (= cave specific) traits. In *G. minus*, multiple pairs of genetically related, physically proximate cave and surface populations exist which exhibit an astounding degree of intraspecific morphological divergence. The morphology, ecology, and genetic structure of these populations are well characterized, yet the genetic basis of morphological divergence remains unknown. RNA-Seq data from a pair of morphologically distinct sister populations inhabiting surface and cave habitats was collected (four individuals from each population) and used to identify genes associated with the evolution of troglomorphic traits. Candidate genes were identified as those exhibiting either large difference in expression level, or in their rates of evolution, between cave and surface populations. Of the ~104,000 transcripts identified in the de novo assembly, ~360 were found to be significantly downregulated in the cave population, and ~800 were significantly upregulated in the cave population. Results from sequence analysis of the differentially regulated transcripts provide some evidence for loss of functional constraint in transcripts related to vision and repair of UV-induced DNA damage.

78B

The Origin of Fibromelanosis using Genetic Comparison between Indonesian Cemani Chicken and Other Domesticated Chickens

Anik Budhi Dharmayanthi, Yohei Terai, Yoko Satta
SOKENDAI (The Graduate University for Advaced Studies, Kamiyamaguchi 1560-35, Hayama, Japan

Both Indonesian Cemani and Chinese Silky chicken show the Fibromelanosis(Fm) phenotype, which is typically represented with black pigmentation due to the excess of melanin accumulation in several tissues including skin. Previous studies revealed that the Fm phenotype in Silky is highly associated with an inverted duplication of two segments separated by a 417kb-spacer on chromosome 20, of which one segment contains the Endothelin3 gene (*EDN3*).

This study aims to examine whether this duplication is present in the Cemani and to search for the origin of this duplication in chicken populations. The examination of duplication boundaries and the number of duplication units by PCR and qPCR suggests that the same duplication as in Silky was observed in the Cemani. We also found 11 distinct haplotypes of *EDN3*, among which, haplotypes 2 and 4 were always found in the Silky and Cemani, suggesting that the duplicated *EDN3* is composed of a set of haplotypes 2 and 4. The divergence time of haplotypes 2 and 4 was 1.2 mya, indicating that these haplotypes have emerged before the domestication of chickens and may be present in Red Jungle Fowl.

In addition, further investigation of the 417kb-spacer and *EDN3* sequences in the Silky and Cemani showed a polymorphic pattern in the middle and distal part, compared to a monomorphic pattern in the proximal part. This may indicate that selection has acted on the duplicated region and has reduced the heterozygosity in the proximal part during selection of Fm phenotype in both the Silky and Cemani.

79C

Evolution of sexual dimorphism under genetic polymorphisms of xenobiotic metabolizing enzymes in Medaka fish (*Oryzias latipes*)

Takafumi Katsumura [1,2], Shoji Oda[2], Shigeki Nakagome[3], Tsunehiko Hanihara[1], Hiroshi Kataoka[2], Hiroshi Mitani[2], Shoji Kawamura[2], Hiroki Oota[1]

[1] Kitasato University School of Medicine, Sagamihara, Kanagawa, Japan, [2] University of Tokyo, Kashiwa, Chiba, Japan, [3] The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan

Sexual dimorphisms, which are phenotypic differences between males and females, are driven by sexual selection. Interestingly, sexually selected traits show geographical variations within species despite strong directional selective pressures. This paradox has eluded many evolutionary biologists for some time, and several models have been proposed (e.g. 'indicator model' and 'trade-off model'). However, disentangling which of these theories explains empirical patterns remains difficult, because genetic polymorphisms that cause variation in sexual differences are still unknown. In this study, we revealed a strong association between differences in the activity of the xenobiotic enzyme cytochrome P450 (CYP) 1B1 and differences in the morphology of the anal fin, which is a sexually selected trait, using a genetic model fish, the medaka (*Oryzias latipes*). Our biochemical assays and genetic cross experiments show that high- and low-activity CYP1B1 alleles enhanced and declined sex differences in anal fin shapes, respectively. Behavioral and phylogenetic analyses suggest maintenance of the high-activity allele by sexual selection, whereas the low-activity allele possibly has experienced positive selection due to by-product effects of *CYP1B1* in inferred ancestral populations. Therefore, we provide evolutionary evidence that varying degrees of sexual dimorphism among wild medaka populations reflect successions of sexual and natural selection due to the pleiotropic effect of the *CYP1B1* gene. These findings indicate that the 'trade-off model' between sexual and natural selection can explain why sexually selected traits vary among geographical populations.

80D

The Transcriptomic Analyses of Feather in Chickens.

Chen Siang Ng [1], Chi-Kuan Chen[1], Wen-Lang Fan[1], Ping Wu[2], Siao-Man Wu[1], Jiun-Jie Chen[1], Mei-Yeh Jade Lu[1], Chih-Feng Chen[3], Cheng-Ming Chuong[2], Wen-Hisung Li[1]
[1] Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, [2] Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, [3] Department of Animal Science, National Chung Hsing University, Taichung, Taiwan

Feathers have diverse forms and are an excellent model for studying the development and evolution of morphological traits. The complex structure of feathers allows for various types of potential morphological changes to occur. The genetic basis of the structural differences between different parts of a feather and between different types of feather is a fundamental question in the study of feather diversity, yet there is only limited relevant information for gene expression during feather development. Here we report transcriptomic analysis of five types of feather follicles across different feather types and regions. By using RNA-sequencing we gained a more complete view of the feather morphpgenesis. Comparative analyses of the differentially expressed genes among feather samples revealed a set of genes that are involved in tissue morphogenesis, regulation of developmental process, and tube development. Finally, we also identify numerous potential candidate genes that have contributed to the development of specific structures of various feather types. The expression profiles of these genes are very useful for understanding the evolutionary origin and diversification of feathers.

81A

Evolution of eye size and head morphology between *Drosophila americana* and *D. novamexicana*

Micael Reis [1], Cristina P. Vieira[2], Rodrigo Lata[1], Nico Posnien[3], Jorge Vieira[1]
[1] Molecular Evolution Group, IBMC, Porto, Portugal, [2] Evolutionary Systems Biology Group, IBMC, Porto, Portugal, [3] Department of Developmental Biology, Institute for Zoology and Anthropology, Georg-August-University Göttingen, Göttingen, Germany

The morphology of organs is usually conserved, since it is under intense selection both within and between species. Thus, when changes in the relative size of organs are observed, they likely represent functional adaptations to an ever changing environment. Nevertheless, the molecular basis of such changes is not always the same. For instance, significant differences in eye size between different species of the *melanogaster* group were found to be caused mainly by facet size, while the intraspecific variation in eye size is mainly due to changes in ommatidia number. Therefore, it is imperative to study divergent groups of species to determine if changes in eye size are governed by general or specific mechanisms across the entire *Drosophila* genus.

Preliminary data shows major differences in eye size between species of the *D. virilis* group, which are largest between *D. novamexicana* (smallest eyes) and *D. americana* from the south of the distribution (biggest eyes). These species are diverging from *D. melanogaster* for at least 40 million years, representing an excellent model to find out if natural variation in eye size and head morphology is due to the same changes in the underlying gene regulatory network in divergent *Drosophila* lineages.

In order to reveal the major genomic regions responsible for the observed differences in eye size and head morphology, a genotype-phenotype association approach involving *D. novamexicana* and *D. americana* from the south of the distribution is being performed.

82B

Human *SEMG1* and *SEMG2* variation and its correlation with infertility phenotypes

Filipa Fonseca [1], Patrícia Isabel Marques[1,2], Ana Sofia Carvalho[3], Isabel Damião[4], Vasco Almeida[4,5], Nuno Barros[6], Alberto Barros[6,7], Filipa Carvalho[7], Rune Mathiesen[3], Victor Quesada[2], Susana Seixas[1]

[1] Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal, [2] Department of Biochemistry and Molecular Biology-IUOPA, University of Oviedo, Oviedo, Spain, [3] Human Genetics Department, National Institute of Health Dr Ricardo Jorge (INSA), Lisboa, Portugal, [4] Center of Infertility and Sterility Studies (CEIE), Porto, Portugal, [5] Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal, [6] Center for Reproductive Genetics Alberto Barros, Porto, Portugal, [7] Department of Genetics, Faculty of Medicine, University of Porto, Porto, Portugal

Semenogelins (SEMG1 and SEMG2), the main structural components of the semen coagulum, entrap spermatozoa in a crosslinked matrix until the moment they are hydrolyzed, allowing spermatozoa to regain motility and to fertilize the egg. In primates, the rapid evolution of *SEMG1/2* has been correlated to divergent coagulum features, which may vary from a viscous mass to a rigid copulatory plug in monoandrous and polyandrous species, respectively. To access the impact of *SEMG* sequence variation in human reproductive traits, we performed a screening of *SEMG1/2* coding regions in infertility cases (N=151) with abnormal semen parameters and a control population (N=80). In *SEMG1*, we identified 7 variants, including a CNV, 3 non-synonymous and 3 synonymous variants, where only the CNV and a linked synonymous variant surpassed a MAF>5%, and a single variant was predicted as deleterious. Conversely, in *SEMG2*, we detected 4 non-synonymous variants, in which 3 had a MAF<5% and were predicted as possibly damaging. To empower our analysis, we selected 1 *SEMG1* and 2 *SEMG2* variants for genotyping in extended case (N=102) and control (N=117) samples. However, to circumvent a possible lack of statistical significance, due to the low frequency of candidate variants, we decided to characterize the semen profiles by mass spectrometry based proteomics in a subsample of 40 cases and controls. So far, we found a rare variant in 1% of cases possibly linked to low fertility rates and abnormal semen parameters, but oddly among our cases, the hyperviscosity phenotype appears to be correlated to lower SEMG levels.

83C

Transcriptional regulation of morphological transitions in stigonematales cyanobacteria

Robin Koch, Karina Stucken, Anne Kupczok, Judith Ilhan, Katharina Brandstädter, Tal Dagan
Genomic Microbiology Group, Institute of Microbiology, Christian-Albrechts University, Kiel, Germany

Cyanobacteria constitute a monophyletic group, yet species in this group are highly diverse in their cellular morphology, including unicellular and linear or branching filaments. The mechanisms of cyanobacterial phenotypic diversity and their evolution are yet unclear. *Chlorogloeopsis* and *Fischerella* species, forming multi-seriate and branching filaments respectively, represent the most phenotypically complex cyanobacteria. Interestingly, adding salt or sucrose to their culture induces a morphological transition: *Fischerella* (*F. muscicola* PCC7414 and *F. thermalis* PCC7521) form linear filaments under heterotrophic growth and *Chlorogloeopsis* (*C. fritschii* PCC6912) forms clumped clusters under salt stress. This suggest that such morphological transitions are facillitated by transcriptional regulation rather than differential gene content. Here we quantify the transcriptome conservation among the three species and compare the transcriptional regulation between the natural and transformed morphologies. Using 5'dRNAseq data we annotated transcriptional start sites (TSSs). A comparison of single-copy orthologous genes revealed that 4,456 (30%) TSSs are conserved between the *Fischerella* species and 1,809 (11%) TSSs are conserved between *F. muscicola* and *C. fritschii*. An analysis of *Fischerella* orthologs where the transcriptional changes between the branched and linear filaments were correlated, revealed 78 TSSs that are putatively related to the morphological transition. One upregulated TSS, for example, is located upstream of a putative bicistronic operon that includes a *bolA-like* gene. This gene is homolog to a known morphogene in Escherischia coli (*bolA*) that acts as a repressor of the bacterial actin homolog, *mreB*. Our reults thus suggest that supression of the *mreB* expression may be involved in cyanobacterial morphological transition.

84D

Development and Evolution of Pupfish (genus *Cyprinodon*) Skull Morphology

Ezra Lencer, Richard Harrison, Amy McCune
Cornell University, Ithaca NY, USA

Understanding the origins of novel phenotypic variation is fundamental to the study of biological diversity. Many studies have focused on the role of selection in driving phenotypic change; however, equally important is how phenotypic variation is generated. We investigate the genetic and developmental sources of skull modifications in a geologically recent radiation of three sympatric pupfish species (genus *Cyprinodon*) endemic to the lakes of San Salvador Island, Bahamas. These pupfish differ dramatically in jaw morphology, exhibiting elongate upturned lower jaws used to remove scales from other fish, and robust squat jaws nested under nasal and maxillary extensions associated with the transition to a shelled prey diet. We show how differences in skull morphology emerge over development as a consequence of differential growth of oral jaw elements. Lengthening or shortening of these elements affects the shape of jaw bones as well as the relative placement of bones in the skull leading to derived morphologies unlike those exhibited by any of the ~50 other species in the genus. Whole mount *in situ* hybridization results to date indicate that the spatial and temporal regulation of genes known to affect skull morphology in other taxa are not differentially expressed among pupfish species. We furthermore take an RNA-seq approach to measure regulatory divergence in developing skull tissue among the three San Salvador species at multiple developmental stages that span embryonic development and larval growth. These data will identify novel sources of skull variation, an ecologically important trait that varies dramatically across vertebrates.

85A

The Pattern of Degradation of the Self-Incompatibility Locus Revealed by Species-Wide Genomic Polymorphism Data in *Arabidopsis thaliana*

Pauline Goubet[1], Takashi Tsuchimatsu [2,3], Sophie Gallina[1], Anne-Catherine Holl[1], Isabelle Fobis-Loisy[4], Hélène Bergès[5], William Marande[5], Elisa Prat[5], Dazhe Meng[2], Quan Long[2], Magnus Nordborg[2], Xavier Vekemans[1], Vincent Castric[1]

[1] CNRS-Université de Lille, Villeneuve d'Ascq cedex, France, [2] Gregor Mendel Institute, Vienna, Austria, [3] University of Tokyo, Tokyo, Japan, [4] Université Claude Bernard Lyon I, Lyon, France, [5] Centre National des Ressources Génomiques Végétales, Castanet-Tolosan, France

The evolution of self-fertilization (selfing) from outcrossing via loss of self-incompatibility (SI) is one of the most frequent evolutionary transitions in angiosperms. In the predominant selfing *Arabidopsis thaliana*, three groups of divergent haplotypes (haplogroups) have been maintained at the S-locus, which determines the male-female specificities of SI. Despite the extensive functional and evolutionary studies on the S-locus for decades, its degradation process through the loss of SI is not fully understood mainly due to paucity of large-scale polymorphism data encompassing the whole S-locus region. Here we obtained polymorphism information of the S-locus at a species-wide scale by combining BAC sequencing and genomic re-sequencing. The accessions differed by several major rearrangements including large deletions and inter-haplogroup recombination. We observed a clear geographical structure in the distribution of haplotypes, and a strikingly high frequency of recombinant haplotypes. Importantly, one inversion in the male *SCR* gene, which accounts for loss of SI in the most frequent haplogroup, was shared among all recombinants except derived deletion haplotypes, suggesting that recombination occurred after the loss of SI. Identification of recombination breakpoints suggested multiple origins of these recombinants. These findings are consistent with the scenario that the complete loss of SI in *A. thaliana* involved independent self-compatible mutants fixed in allopatric populations during the last ice age or earlier, and experienced further rearrangements during secondary contact due to post-glacial colonization.

86B

Natural variation in flowering-related traits in perennial *Arabis alpina*

Ulla Kemi [1], Stefan Wötzel[1], Jörg Wunder[1], Per Toräng[2], Jon Ågren[2], George Coupland[1]
[1] Max Planck Institute for Plant Breeding Research, Cologne, Germany, [2] Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Understanding the molecular-genetic basis of adaptive processes is a major goal in modern evolutionary biology. In plants, flowering traits are often associated with fitness. We study the genetic basis of natural variation in flowering time (timing of first flower opening) and other perenniality-associated traits in *Arabis alpina*. In previous field- and greenhouse experiments, we have shown that the timing of flowering varies among and within populations. In this study, we focus on one Scandinavian population, which shows particular within-population variation in flowering time. To study the genetic basis of the observed variation, early and late flowering individuals were crossed to obtain an F2 population for QTL mapping. Flowering time and other perenniality-associated traits were scored in the greenhouse. Interestingly, even though all parental plants made inflorescence buds in the experiment, only one third of the late parents continued to produce open flowers, whereas all early parents completed flower development. In addition, the late parents that flowered showed a much longer time interval between bud emergence and open flowers than the early parents. Flowering time and the time interval between inflorescence bud formation and flowering start segregated within the F2 population providing promising material for QTL mapping. QTL mapping will identify genomic regions governing variation in the studied traits in the population. We will focus on the strongest-effect QTL and use a combination of candidate gene analysis, high-resolution mapping and reverse genetics to identify the causal genes. We will study the adaptive significance of the detected genetic variation in further experiments.

87C

Anthropometric and cardiovascular trait variation among sub-Saharan African populations: the role of gender, subsistence, and genetic ancestry

Matthew Hansen [1], Joseph Lachance[1,2], Sameer Soi[1], Laura Schienfeldt[1,3], Alessia Ranciaro[1], Simon Thompson[1], Jibril Hirbo[1,4], Sarah Tishkoff[1]

[1] University of Pennsylvania, Philadelphia, PA, USA, [2] Georgia Institute of Technology, Atlanta, GA, USA, [3] Coriell Institute for Medical Research, Camden, NJ, USA, [4] Vanderbilt University, Nashville, TN, USA

The African continent is home to a diverse range of indigenous peoples that have adapted to a wide range of ecological environments and subsistence lifestyles. Many complex traits are expected to display variation between populations due to demographic history and/or natural selection to these diverse environments. In an effort to survey phenotypic variation in Africa and begin to understand the genetic and environmental factors that contribute to this variation, we have collected trait measurements on height (N=5,125), BMI (N=5,098), grip strength (N=1,968), systolic and diastolic blood pressure (N=2,002), and pulse (N=2,008) from agricultural, pastoral, and hunter-gatherer communities across eastern and western sub-Saharan Africa. We present the observed variation in these traits between genders, across populations, and across subsistence practices. We find significant differences in trait values among these categories. A subset of 697 individuals were genotyped on the Illumina 1M-Duo SNP array. To assess the impact of genetic ancestry on trait variation, we performed STRUCTURE analysis to determine ancestral cluster proportions, and the trait distributions for each ancestral cluster and subsistence category were inferred by a mixture model. The fraction of variance explained by this model is discussed, as well as the implications for future genotype/phenotype analysis within sub-Saharan Africa for these and related traits.

L 8D

Using Marine Model Organisms and Experimental Evolution to Develop Selective Breeding Tools for Aquaculture

Nathan Churches [1], Sergey Nuzhdin[1], Dennis Hedgecock[2]
[1] University of Southern California, Department of Molecular and Computational Biology, USA, [2] University of Southern California, Department of Marine and Environmental Biology, USA

The aquaculture industry, and specifically bivalve crops, lag in genetically based selective breeding programs when compared to terrestrial commercial crops. Our lab is addressing this this gap by applying high throughput model organism techniques and experimental evolution in two emerging marine model organism systems, the Mediterranean Mussel (*Mytilus galloprovincialis*) and the Pacific Oyster (*Crassostrea gigas*). This poster describes two projects which have similar goals: development of genetics/genomics based selective breeding tools for aquaculture. In collaboration with the Southern California based aquaculture company Catalina Sea Ranch, and with funding from a NOAA SBIR grant, we plan to realize three specific goals for Mediterranean Mussels: 1) to establish genetically distinct family lines of mussel, focusing on desirable commercial characteristics, 2) investigate genetics of viability and selection, and 3) determine mutation rates and their impact on domesticated shellfish cultivars. For Pacific Oysters, we have developed a project proposal (in review phase) for the USDA NIFA, with specific goals to 1) develop 10 family lines 2) phenotype, select, isolate, and grow-out elite individuals from each family line 3) sequence and compare individuals from strategic populations 4) implement 'xQTL Mapping' to create a simple "SNP Typing Assays for Yield" (STAY) for selective breeding protocol and 5) create market-ready US West Coast adapted stocks with our STAY protocol. Each of the proposals described above will produce Southern California warm-water adapted accessions. These will be extremely valuable to academic and industrial efforts alike and, if successful, our projects will help push aquaculture into the modern breeding era.

13 Inferring fitness landscapes from experimental evolution

13.1

Inferring macroscopic epistasis from experimental evolution

Michael Desai
Harvard University, Cambridge, MA, USA

Empirical efforts to characterize patterns of epistasis often take a direct approach: candidate mutations are introduced into a set of genetic backgrounds via crossing or other genetic reconstruction techniques, and the fitnesses of the reconstructed genotypes are measured using competitive fitness assays or related proxies. These data yield a functional relationship between the fitness effect of a mutation and its genetic background. However, evolutionary dynamics typically depend on the entire distribution of fitness effects ("the DFE"), and on how this distribution varies among different genetic backgrounds. This background dependence of the DFE is a "macroscopic" form of epistasis, since it refers to the collective effects of many epistatic effects rather than the background dependence of any individual mutation. I will describe a general framework for quantifying macroscopic episatsis from observations of evolutionary dynamics in laboratory evolution experiments. I will then apply this framework to investigate the role of epistasis in a well-studied laboratory evolution experiment in E. coli, as well as to shorter-term experiments in budding yeast. I will show that fitness measurements alone have little power to discriminate between different models of epistasis, but that addition of genetic information is sufficient to distinguish between several popular models of epistasis.

13.2

The adaptive fitness landscape in an experimentally evolving yeast population

Sandeep Venkataram, Barbara Dunn, Yuping Li, Kerry Samerotte, Lucas Herissant, Jessica Chang, Gavin Sherlock, Dmitri Petrov
Stanford University, Stanford, CA, USA

A central concept in evolutionary biology is the fitness landscape, which describes the fitness effect of all potential mutations on a focal genotype. Recent efforts have begun to characterize the fitness landscape of biological systems, particularly using experimental evolution to characterize adaptive mutations. However, we have yet to identify both the molecular nature and fitness effects of the adaptive fitness landscape in any system at a large scale. Utilizing a novel DNA barcoding system to identify adaptive mutations soon after they occur in an evolving yeast population under glucose limitation, we have successfully isolated and genotyped hundreds of independent adapted lineages, most of which carry a single adaptive mutation and for which we had previously inferred their fitness effects.

We identify two main categories of adaptive mutation in our system, suggesting that there are two separate "peaks" in the local fitness landscape. One class of adaptive mutation appears to be diploidization of our initially haploid yeast, which we estimate to confer a 4% fitness advantage and occurs in over 40% of our sampled lineages. We also observe over 50 lineages with mutations in the RAS-cAMP pathway, conferring a 6-11% growth advantage, with sufficient power to assign mutations in different genes significantly different fitness effects. Our methodology has allowed us to sample both the molecular nature and fitness effects of adaptive mutations in an unbiased manner, giving us unprecedented power to describe the fitness landscape of experimentally evolving yeast.

13.3

An intragenic fitness landscape spanned by beneficial mutations

Claudia Bank [1,2], Ryan T. Hietpas[3], Sebastian Matuszewski[1,2], Daniel N.A. Bolon[3], Jeffrey D. Jensen[1,2]

[1] School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3] Department of Biochemistry & Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA

The architecture of adaptation is dependent on both the shape of the distribution of fitness effects of beneficial mutations, and on epistasis, which describes the interaction between such mutations. One way to picture this information is as a fitness landscape, and the study of theoretical and empirical fitness landscapes is receiving ever-increasing attention. Novel experimental approaches combined with next-generation sequencing methods allow for more accurate and extensive studies of the fitness effects of mutations – allowing us test theoretical predictions and improve our understanding of the shape of the true underlying fitness landscapes. Here, we present fitness data from several hundred engineered mutants that represent combinations of up to six previously identified beneficial single mutations in a small region of the heat-shock protein Hsp90 in *S. cerevisiae*. We discuss the emerging pattern of diminishing-returns epistasis in the framework of theoretical expectations derived from Fisher's geometric model, and describe the emerging picture of the local fitness landscape. By comparing our findings with those derived from previous studies of different design, we discuss the challenges of trying to extrapolate from a local snapshot to the global fitness landscape, and how the information gained will prove important for the future study of adaptive evolution.

13.4

A Gaussian process model of genetic time series data arising in experimental evolution

Jonathan Terhorst, Yun Song
UC Berkeley, Berkeley, CA, USA

Most inference procedures in population genetics are designed to analyze a single random sample from a large population. On the other hand, data sets in experimental evolution often consist of samples drawn repeatedly over time from a small population. Hence, existing models in population genetics are often unsuitable for experimental evolution data; in particular, few methods exist to analyze time series of pooled sequence data.

We have developed a Gaussian process approximation to the discrete-time Wright-Fisher process which models an evolve-and-resequence experiment. The mean and covariance of the process depend on linkage and fitness, allowing one to test for and estimate selection in multi-locus time series data. Using simulated data, we demonstrate that incorporating linkage information enables our method to correctly locate and estimate the fitness of a selected allele from among many linked sites. We study how this power responds to different values of selection strength, initial haplotypic diversity, population size, sequencer effort, experimental duration, and number of replicates. Other population genetic quantities such as effective population size and recombination rates can also be estimated. Our freely available implementation consists of optimized C++ and Python code and is fast enough to perform genome-wide analysis on real data.

13.5

How good are statistical models at approximating high-dimensional fitness landscapes?

Louis du Plessis, Gabriel Leventhal, Sebastian Bonhoeffer
ETH Zürich, Zürich, Switzerland

Fitness landscapes determine the course of evolution by constraining and shaping evolutionary trajectories. Empirical fitness landscapes have so far only offered limited insight into real-world questions, as the high dimensionality of sequence spaces makes it impossible to exhaustively sample all variants of biologically meaningful sequences. In practical terms there is often no other possibility than reverting to statistical descriptions of fitness landscapes, based on sparsely sampled fitness measurements [1,2]. It remains unclear, however, how much data is required for such statistical descriptions to be useful. Here, we assess the ability of regression models accounting for single and pairwise mutations to correctly approximate complex landscapes. We compare approximations based on various sampling schemes of quasi-empirical RNA fitness landscapes and find that the sampling scheme strongly influences quality of the regression. While it is generally impossible to generate sufficient samples to achieve a good fit using randomly sampled sequences, we obtain a remarkably good and unbiased fit to the local landscape when using sequences from a population evolving under strong selection and weak mutation. Thus, while it remains challenging to obtain an accurate statistical description of a complete fitness landscape, current methods can provide a good approximation to the local landscape of naturally evolved populations.

[1] Hinkley, T. et al. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nat Genet, 43, 2011.

[2] Otwinowski, J. and Nemenman, I. Genotype to phenotype mapping and the fitness landscape of the E. coli lac promoter. PLoS ONE, 8, 2013.

13.6

Discreteness and continuity in Adaptive Landscape

Olivier Tenaillon [1,2], Hervé Jacquier[1,2], André Birgy[1,2], Julien Guglielmini[1,2], Alejandro Couce[1,2], Hervé Le Nagard[1,2]

[1] INSERM, IAME UR1137, Paris, France, [2] University Denis Diderot, Paris 7, Paris, France

Throughout the literature, the adaptive landscape is often pictured either as discrete and rugged or as continuous and smooth. The rugged vision suggests that the adaptive landscape results from the sum of pairwise epistatic interactions among mutations while the smooth vision suggests that some higher order interactions or macroscopic epistasis shapes the landscape. Both of these opposing visions have received some experimental support. Using experimental evolution data, high-throughput mutant phenotyping within a single gene, protein lattice models and Fisher Geometric Model, we study the transition between these alternative visions.

13.7

The effects of selection and mutation rate heterogeneity on parallel evolution in *Saccharomyces cervisiae*

Susan Bailey [1], Qianyun Guo[1], Françios Blanquart[2], Thomas Bataillon[1]
[1] *Aarhus University, Aarhus, Denmark,* [2] *Imperial College London, London, UK*

Parallel evolution - similar evolutionary changes arising in independently evolving populations, is often taken to be evidence of strong selection, however theory suggests that heterogeneity in mutation rate can be an equally important driver of parallelism. We explore the contributions of mutation and selection to the degree of parallelism at the gene-level in 40 experimentally evolved *Saccharomyces cervisiae* populations (Lang et al, Nature 2013). We use regression models to identify a number of genomic variables that signifcantly affect variation in mutations per gene (e.g. gene length, dN/dS, and codon bias). Then, first focusing on the synonymous mutations, we use a probabilistic model to generate expectations for the distribution of mutations per gene, with and without heterogeneity in mutation rate. We show that a model incorporating hetergeneous mutation rates best fits these data and that variation in mutation rate estimated from species-level comparisons (dS) captures the required heterogeneity well. We then show that mutation rate heterogeneity is not sufficient to explain variation in non-synonymous mutations per gene, and so unsurprisingly, selection must also play a role. Maximum likelihood techniques are employed to quantify the relative contributions of heterogeneity in mutation rate and selection that best explain these data, first under the assumption that mutations fix sequentially, and then allowing for the effects of clonal competition. We conclude that both processes play important roles in generating the patterns of parallelism observed in this experimental yeast system and discuss the generality of our results.

13.8

Bacterial evolution of antibiotic hypersensitivity

Viktória Lázár, Gajinder Pal Singh, Réka Spohn, István Nagy, Balázs Papp, Csaba Pál
Synthetic and Systems Biology Unit, Biological Research Center, Szeged, Hungary

It has long been known that evolution of resistance towards a given antimicrobial agent frequently increases resistance to multiple other drugs. In sharp contrast, it has remained unclear how frequently evolution of resistance increases sensitivity to other drugs, a phenomenon usually termed as collateral sensitivity. Despite its importance, no systems-level analysis has been performed on such phenomenon leaving open questions about how frequent and how far they are understandable based on accumulated knowledge on individual compounds. Standard tools of microbial evolution experiments provide an ideal model system to systematically investigate this issue under controlled lab conditions. Therefore we initiated large-scale evolution experiments with Escherichia coli populations towards 24 different antibiotics, after that the sensitivity of each evolved line was tested against the entire series of antibiotics. Our results demonstrated that collateral sensitivity occurs frequently during antibiotic adaptation. Specifically, populations adapted to aminoglycosides stood out showing an especially increased sensitivity against multiple other antibiotics. Whole-genome sequencing of the evolved strains revealed a diverse set of parallel mutations with pleiotropic effects. Most notably, adaptation to aminoglycosides is generally achieved by decreasing their uptake through reducing proton-motive force (PMF) across the inner bacterial membrane. As a by-product, these mutations diminish activity of PMF-dependent major efflux pumps, leading to hypersensitivity to several other compounds. More generally, our work offers insights into the mechanistic principles driving the evolution of negative trade-offs under antibiotic selection.

13.9

Evolution of fitness trade-offs in locally adapted populations of *Pseudomonas fluorescens*

Alana Schick [1], Susan Bailey[2], Rees Kassen[1]
[1] *University of Ottawa, Ottawa, Canada,* [2] *Aarhus University, Aarhus, Denmark*

Local adaptation is important for understanding the patterns of diversity we observe in nature. When populations become specialized for growth in certain conditions, trade-offs for growth in other conditions are expected to arise as a result of adaptation being 'blind' to non-selected environments. How often trade-offs occur, how they arise, and the specific loci involved remains poorly understood. In this study, we characterized the trade-offs associated with several beneficial mutations in previously adapted populations of *Pseudomonas fluorescens*. We measured fitness effects of specific mutations in selected and non-selected environments to quantify the degree specialization. We found that beneficial mutations arising later during adaptation were associated with increasingly antagonistic effects in non-selected environments than those arising earlier. We also found that epistatic effects between beneficial mutations were consistently negative in selected environments and unpredictable in non-selected environments. Lastly, we found that mutations in genes associated with motility had similar effects across environments while those associated with regulation varied greatly. Taken together, these results provide a detailed account of the genetics of specialization and suggest that trade-offs arise because beneficial mutations arising later in an adaptive walk become increasingly deleterious in non-selected environments, as opposed to being conditionally neutral.

13.10

Fitness determines adaptability of highly diverse genotypes

Elizabeth Jerison, Sergey Kryazhimskiy, Michael Desai
Harvard University, Cambridge, MA, USA

The fitness landscape can in principle vary in arbitrarily complex ways with the underlying genotype. However, recent work with laboratory microbial populations has suggested a simpler picture: populations adapt at rates determined by their initial fitness, regardless of the specific mutations in the genetic background. While tantalizingly simple, this principle has been demonstrated only for genotypes that differ by a handful of mutations. To determine its applicability to much more highly diverged genotypes, we tested the adaptability of 230 segregants from a cross between a laboratory and a vineyard yeast strain. These segregants typically differ from one another at ~30,00 loci. Surprisingly, we find that adaptability is strongly correlated with initial fitness, while other effects of genotype play a smaller role. In both a permissive and a high temperature environment, populations that begin with higher fitness adapt slower than initially less-fit populations. This suggests that some properties of the fitness landscape may be functions of current fitness, regardless of genotype, even at very diverged points in genotype space. However, there remains a residual effect of the specific genotype. To identify particular genetic loci that influence adaptability over and above fitness, we use the genotyped sergeant panel to map adaptability QTLs. Finally, the correlation between initial fitness and adaptability suggests a simple model of the pleiotropic costs in other environments associated with adaptation. We use our measurements to parametrize this model, and test it by measuring fitnesses of evolved lines in another growth condition.

57A

Novel approaches for the analysis of mRNAseq data: the expression profile versus the abundance

Irina Mohorianu, Tamas Dalmay, Tracey Chapman
University of East Anglia, Norwich, UK

Current approaches for the analysis of mRNAseq data often focus on obtaining a single abundance value per gene. This ignores the additional, rich and important information that is available from RNA sequencing. The study of the characteristics of sequencing reads, such as localization and distribution of abundances, can reveal the distortions resulting from comparisons between samples with different total read numbers. Uncorrected, these can bias and reduce resolution of analyses of differential expression (DE). This can be circumvented by a novel bootstrap normalization. Application of this approach revealed the subtle, previously undetectable, transcriptome differences underlying behaviour in a study of the effects of sexual environment in *D. melanogaster* fruit flies.

The improved quantification and characterization of DE transcripts also revealed a novel biological phenomenon - the existence of alternative pathways activated in response to similar biological inputs. These pathways were observed in a wide spectrum of biological functions, from sensory and immune genes to ejaculate components. Hence by replacing crisp interactions with a fuzzy approach we revealed a mechanism that can potentially confer robustness upon highly plastic traits. Additional investigation of the gene expression profiles led to the identification of sequencing blind regions, i.e. short regions, distinct from introns, flanked by high expression, with no incident reads. Preliminary data indicate that these blind regions may be linked to miRNA targeting, offering a potential alternative method for the identification of miRNA-mRNA target pairs in animal systems.

58B

Fitness Trade-offs determine the role of the molecular chaperone GroEL in canalizing genetic variation

Beatriz Sabater-Muñoz [1], Maria Prats-Escriche[1], Roser Montagud-Martínez[2], Adolfo López-Cerdán[2], Christina Toft[3,4], Mario A. Fares[1,2]

[1] Smurfit Institute of Genetics, Trinity College of Dublin, Dublin, Dublin co., Ireland, [2] Instituto de Biología Molecular y Celular de Plantas (C.S.I.C.-U.P.V), Valencia, Spain, [3] Departamento de Biotecnología, Instituto de Agroquímica y Tecnología de los Alimentos (CSIC), Valencia, Spain, [4] Department of Genetics, University of Valencia, Valencia, Spain

Molecular chaperones fold many proteins and their mutated versions in the cell, modulating the relationship between the primary protein sequence (genotype) and it's structure (phenotype). How do chaperones buffer and canalize genetic variation, the fitness trade-offs of their mutational buffering, and their role in expediting evolution remain largely obscure. Answering these questions is fundamental in evolutionary biology to understand the contribution of mechanisms of mutational buffering to increasing genetic variation and ecological diversification. In this study we performed experimental evolution, genome re-sequencing, and computational analyses to determine the trade-offs and evolutionary trajectories of *Escherichia coli* with high levels of the essential chaperone GroEL. GroEL is abundantly present in bacteria, particularly those with large loads of deleterious mutations, suggesting its role in mutational buffering. We show that *groEL* over-expression is taxing in large populations evolving in the laboratory. Conversely, populations under genetic drift, resembling endosymbiotic bacteria of insects, only persist in the presence of abundant GroEL levels. Genomes re-sequenced from cells evolving under genetic drift exhibited significantly more deleterious mutations at high GroEL levels than at constitutive levels. Interestingly, GroEL buffers mutations in a bewildering set of proteins that interact with the environment, suggesting its role in the emergence of ecological diversification. Remarkably, the cost of *groEL* over-expression imposes selective constraints against mutations in other proteins, indirectly modulating the evolution of these proteins. Our results shed light on the fitness trade-offs of mutational buffering and have important implications in the understanding of the origin of large evolutionary leaps.

59C

The fitness landscape of a yeast tRNA gene

Chuan Li [1], Wenfeng Qian[1,2], Jianzhi Zhang[1]
[1] *University of Michigan, Ann Arbor, MI, USA,* [2] *Chinese Academy of Sciences, Beijing, China*

Mutations provide raw materials for evolution. Fitness landscape, the comprehensive description of the fitness effects of all mutations, is of fundamental importance for understanding general properties of epistasis and evolutionary trajectories. Here I describe the fitness landscape of a tRNA gene in the budding yeast *Saccharomyces cerevisiae*. I have collected ~100,000 yeast colonies, each carrying, at its native genomic location, a variant of the tRNA gene that has 0 to multiple (on average 2) point mutations compared with the wild-type. Using the sensitive and high-throughput bar-seq method, I have measured the relative fitnesses of ~20,000 genotypes. The fitness data are being used to describe the fitness landscape of the tRNA gene and quantify epistasis for a large number of mutation pairs. Because the tRNA function is largely determined by its secondary structure, which can be computationally predicted with high accuracy, my study also offers a rare opportunity to understand the biophysical basis of epistasis and fitness landscape. The preliminary results support the role of tRNA stability in its functionality. Further, fitness landscapes and epistasis will be quantified under multiple conditions to understand the genotype by environment interaction. In addition, the expressions of different tRNA variants are being probed to better understand how the internal promoter of the tRNA works and how fitness is affected by the expression level of the tRNA.

60D

Epistasis, pleiotropy, and pathways of adaptive protein evolution

Jay Storz
University of Nebraska, Lincoln, NE, USA

A fundamental question in evolutionary genetics concerns the roles of epistasis and mutational pleiotropy in shaping trajectories of protein evolution. This question can be addressed directly by using site-directed mutagenesis to explore the mutational landscape of protein function in experimentally defined regions of sequence space. Here we report an experimental assessment of how sign epistasis and pleiotropic trade-offs influence the accessibility of alternative mutational pathways during adaptive walks through protein sequence space. Using a combinatorial protein-engineering approach, we examine the effects of sequential mutational steps in all possible pathways that lead to high-fitness genotypes. We evaluate the conditions under which epistasis increases or decreases the likelihood of parallel amino acid substitutions during the replicated evolution of adaptive protein functions. The results have important implications for understanding the repeatability and predictability of molecular adaptation at the sequence level.

61A

Fitness effects of spontaneous mutations in phytoplanktonic eukaryotes

Marc Krasovec [1], Sophie Sanchez-Ferandin[1], Adam Eyre-Walker[3], Gwenael Piganeau[2]
[1] University of pierre and marie curie, Paris 6, France, [2] CNRS, Arago Laboratory, Banyuls sur mer, France, [3] University of Sussex, Brighton, UK

Mutation accumulation experiments enable us to estimate the spontaneous mutation rate and explore the fitness effects of new mutations. Previous estimations have been obtained for model organisms such as *Drosophila* or yeast, and tend to show a fitness decrease of lines during the experiment. We will present our results on mutation accumulation experiments in five species of unicellular haploid green algae; *Ostreococcus tauri*, *O. mediterraneus*, *Micromonas pusilla*, *Bathycoccus prasinos* and *Nannochloris* RCC4223, for which 40 lines have been maintained for 10 000 generations with serial bottlenecks. Their genomes have been completly sequenced and are interesting for mutation accumulation studies because of their size range (13 to 22 Mb) and GC content (46 to 65 GC%). These algae are the smallest free-living eukaryotes and at the basis of food web in coastal oceanic areas. The distribution of fitness effects of mutations is inferred from fitness assays including growth rates measurment of each line during the experiment. First results allow to bring out significant differences in fitness effects of mutations between our five species. At the end of experiment, comparison of phenotypes between mutant and ancestral lines permit us to explore the fitness landscape of spontaneous mutations in these enigmatic eukaryotes that drive the ocean's ecosystem.

62B

Computational topology reveals the structure of the transcriptional regulation landscape

Tiago Paixão, Ulrich Bauer
Institute of Science and Technology Austria, Klosterneuburg, Austria

Fitness landscapes are a metaphor at the heart of population genetics which have been enormously influential at shaping evolutionary thinking. However, the fact that genotype spaces are extremely high-dimensional and display a non-intuitive topology has called into question the value of the intuitions gained from continuous low dimensional landscapes. For example, do fitness peaks really exist or due to the high dimensionality of these spaces there are always neutral ridges connecting them? The advent of cheap sequencing technology and high-throughput measurement promises to resolve this: vast amounts of data mapping these genotype spaces are starting to be produced. However, an intuitive way to describe the landscapes they correspond to is required.
Here we propose merge trees as a method to visualize these high-dimensional spaces. Merge trees are a mathematical tool from computational topology that represent the topological structure of a function as a tree and is guaranteed to preserve the topological information about number of peaks and the valleys separating them. We demonstrate the usefulness of this tool by applying it to the landscapes generated by a thermodynamical model of gene transcription. Using this model, we generate all possible sequences at a promoter of a gene, under different regulatory scenarios, and compute the merge tree describing this landscape. We find that multiple peaks can be persistent in these landscapes, showing that there are fundamental constraints to the evolution of regulatory sequences.

63C

Life finds a way: Adaptation via mutational reversions.

JianRong Yang, Jianzhi Zhang
University of Michigan, Ann Arbor, USA

Darwinian evolution can be described by an adaptive walk on a fitness landscape towards fitter phenotypes. Selectively accessible evolutionary paths to the global peak of the fitness landscape were previously found to be quite limited in fully reconstructed landscapes containing five segregating sites, suggesting high backward predictability but also ubiquity of dead ends in adaptive walks. We hypothesize that mutational reversions, defined by a forward and a backward mutation at a site, can facilitate adaptation by resolving some evolutionary dead ends. We conducted (i) mathematical analysis under the House-of-Cards model (maximal epistasis), (ii) extensive simulation under the NK model (various intermediate levels of epistasis), and (iii) analyses of previously reconstructed fitness landscapes containing up to nine segregating sites. We found that (1) the probability that the global peak could be reached by an adaptive walk containing one mutational reversion has a low upper bound when the number of segregating sites considered is small, but it rapidly increases as the number of segregating sites increases; (2) up to 40% of evolutionary dead-ends can be resolved by a detour containing one mutational reversion; and (3) the structure of the epistasis network determines the time span between the forward and backward mutations and thus the probability with which the reversion is viewed as a segregating site. These results reveal the role of the largely ignored mutational reversion in adaptation, especially in long adaptive walks.

64D

Functional Transitions in Enzyme Evolution: Balancing Stability, Folding and Catalytic Specificity

Bert van Loo [1], Magdalena Heberlein[1], Elias Dohmen[1], Florian Hollfelder[2], Erich Bornberg-Bauer[1]
[1] University of Münster, Münster, Germany, [2] University of Cambridge, Cambridge, UK

Evolutionary pathways by which proteins have evolved in Nature over billions of years have resulted in an impressive diversity of structures that carry out many functions with unrivalled efficiency. Directed protein evolution in the test tube can emulate natural evolution, but is often limited by low hit rates and small improvements during evolutionary cycles. Furthermore, the combination of mutations that is needed for large improvements cannot always be reached by one-by-one mutational steps due to the occurrence of general loss-of-function (e.g. protein unfolding) or epistatic ratchets (i.e. deleterious or permissive mutations that only become beneficial upon occurrence of further mutations). The question then arises how evolutionary dead ends can be avoided. Based on experimental measurements of the dynamics of fitness landscapes of a protein population during evolution cycles we aim to build mathematical models that simulate evolutionary processes at the level of enzyme activities. Important parameters that shape these fitness landscapes are e.g. expression level, stability and catalytic activity/specificity. In addition we aim to probe these parameters for ancestral sequences that are computed from multiple sequence alignments and phylogenetic relationships for members of the catalytically diverse metallo-β-lactamase and alkaline phosphatase superfamilies. The data for these resurrected ancestors can be used to populate fitness landscapes over large genotypic distances and involve transitions in primary function, one of the key processes in evolution of new functions. Both the laboratory and ‘historical' evolution approaches provide a dynamic picture of protein evolution at the level of catalytic function, protein biophysics and population genetics

65A

EVOLVING SYNONYMOUS GENES TO OBTAIN HYPERACTIVE INTEGRASES.

Jose Antonio Escudero, Aleksandra Nivina, Ana Babic, Céline Loot, Didier Mazel
Institut Pasteur, Paris, France

Integrons are gene recruitment platforms that allow for rapid bacterial evolution, playing a major role in the acquisition of antimicrobial resistance genes. The integrase catalyzes the reaction between *attC* and *attI* sites. However, its activity is finely tuned to preferentially process *attI* x *attC* and *attC* x *attC* reactions, as they allow for the acquisition and rearrangement of cassettes, rather than the *attI* x *attI* reaction. Being a less frequent phenomenon, the structural basis of *attI* recognition has yet remained elusive.

We have conducted directed evolution experiments on the integrase to obtain hyperactive mutants for the *attI* x *attI* reaction. In order to explore a broader evolutionary landscape of the integrase we have recoded the protein into two alternative, yet synonymous, alleles of the *intI1* gene.

In a first round of experiments we obtained among the three alleles 8 mutations conferring an increase in recombination rates. We reinserted all mutations into the three wt alleles, obtaining a 100-fold increase in recombination efficiency. Re-evolving these alleles yielded new mutants with recombination rates for *attI* x *attI* comparable to those of the wild type protein for the *attI* x *attC* reaction. Along these experiments we obtained landscape-specific mutations and the re-evolution of a codon to a residue beyond its evolutionary landscape in any of the three starting alleles.

We have successfully explored an enlarged evolutionary landscape of the integrase allowing us to obtain hyperactive molecules while shedding light on the structural features that are important for the recognition of the *attI* site.

66B

Empirical fitness landscape of yeast U3 snoRNA

Olga Puchta, Botond Cseke, David Tollervey, Guido Sanguinetti, Grzegorz Kudla
University of Edinburgh, Edinburgh, UK

Epistasis within and between genes has a large effect on the shape of fitness landscapes. To map the network of epistatic interactions in a model gene, we constructed a synthetic library of more than 100,000 randomly mutated variants of the yeast gene U3, which encodes a noncoding small nucleolar RNA (snoRNA). The variants contain between zero and twenty random substitutions, deletions and insertions, and each variant carries a unique barcode. We measured the fitness of yeast strains carrying these mutations by performing a competitive growth experiment of the entire pool of mutants followed by deep sequencing of barcodes.

We find that most single substitutions have little effect on fitness, but variants with increasing numbers of substitutions become progressively less fit. Substitutions in evolutionarily conserved regions have larger negative effects on fitness, compared to non-conserved regions. Single nucleotide substitutions and deletions are more deleterious than single nucleotide insertions. We will present a mathematical model that partitions the observed variation in fitness into contributions of single mutations and their epistatic interactions.

67C

An experimental investigation of the extent of epistasis between protein orthologues

Victoria Pokusaeva, Fyodor Kondrashov
Centre for Genomic Regulation, Barcelona, Spain

The factors that determine the rate of protein evolution are many, however, a consensus on the interplay of the identified factors or their relative importance is still in the making. A sizable fraction of amino-acid substitutions seem to be fixed by positive selection, but protein evolution is also substantially constrained by epistasis. At the moment the experimental confirmation of the theoretical and computational work on epistasis is the crucial missing piece in our further progress in understanding patterns of long-term protein evolution and maintenance of protein function. Trying to fill this gap we developed the experimental approach to test and verify theoretical predictions of the extent of epistasis in protein evolution. It implies studying of protein orthologues from different species as evolutionary pathways, or trajectories. Studying experimentally yeast proteins the particular question we address is what the probability that an amino acid state accepted in one genotype is deleterious in another. The issue at hand is the extent of epistatic interactions that can allow for a particular amino acid state in one species and prevent the same amino acid state from fixation in another species. By testing the fitness contributions of different allele combinations we estimate the extent to which epistasis plays a role in the course of long-term evolution.

68D

Lineage dynamics in adapting yeast populations.

Julia Piper, Michael Desai
Harvard University, Cambridge, MA, USA

Predictions of evolutionary outcomes based solely on genome sequence data is limited in experimental evolution studies. Gene-level parallelism that results in differing fates across populations and mutational cohorts formed via hitchhiking obscure our ability to assign likelihoods as to which lineages will survive. Furthermore, while initial fitness has been shown to be a strong predictor of final fitness in adapting laboratory populations of microbes, it is unknown how this phenomenon arises from a system's underlying genomic and physiological architecture when significant sequence-level stochasticity is observed. Here we explore the relationship between genotype and fitness by examining phylogenies from a previous long-term evolution study in S. cerevisiae, in which whole-populations were whole-genome sequenced at various time-points over 1000 generations of evolution. We selected a spread of populations representative of observed dynamics in genomic sequence (hard sweeps, clonal interference, etc) and parallelism in mutation identity. We then isolated clones from whole population samples and performed whole genome sequencing and fitness measurements. Using the resulting phylogenies and their associated fitness data, we examine the various population genetics factors resulting in the lineage dynamics of adaptation in this system.

14 Cancer as a Darwinian process

14.1

Cancer, cooperation, and evolution

Natalia Komarova, Dominik Wodarz
University of Califirnia Irvine, Irvine, CA, USA

Complex traits arise from the interactions among multiple gene products; cancer is a prime example of a complex trait. In the case where the complex phenotype is separated from the wild type by a fitness valley or a fitness plateau, the generation of a complex phenotype may take a very long evolutionary time. Interestingly, the rate of evolution depends in nontrivial ways on various properties of the underlying stochastic process, such as the spatial organization of the population and social interactions among cells. Understanding these trends is crucial for understanding cancer initiation and progression. The role of spatial constraints is quite complex: there are realistic cases where spatial constrains can accelerate or delay evolution, or even influence it in a nonmonotonic fashion, where evolution works fastest for intermediate-range constraints. Social interactions among cells can be studied in the context of the division-of-labor games. Under a range of circumstances, cooperation among cells can lead to a relatively fast creation of a complex phenotype as an emerging (distributed) property. If we further assume the presence of cheaters, we observe the emergence of a fully mutated population of cells possessing the complex phenotype. Applications of these ideas to cancer initiation, as well as biofilm formation and virus dynamics will be discussed.

14.2

Sequencing and molecular evolutionary analysis of tumors within patients reveals that metastatic lineages can arise early and exhibit multiple genetic origins within primary tumors

Zi-Ming Zhao[1], Bixiao Zhao[1], Yalai Bai[1], Atila Iamarino[1,2], Stephen Gaffney[1], Joseph Schlessinger[1], Richard Lifton[1], David Rimm[1], Jeffrey Townsend [1]

[1] Yale University, New Haven, CT, USA, [2] University of Sao Paulo, Sao Paulo, Brazil

It has long been understood that tumorigenesis is an evolutionary process associated with the accumulation of somatic mutations. However, many aspects of that molecular evolutionary process that are fundamental to cancer biology and targeted treatment have been challenging to reveal, such as the divergence times and genetic clonality of metastatic lineages. To identify the roots of metastasis, we gathered a uniquely informative set of normal, primary and multiple matched metastatic tumor tissues from 40 subjects. We sequenced exomes from all tissues and reconstructed their phylogenies and ancestral states, yielding three conclusions. First, in contrast with the linear model of cancer progression, metastases can have multiple genetic origins within primary tumors. Evolved genetic changes in cancer lineages likely only affect the proclivity toward metastasis, and single genetic changes are unlikely to be necessary or sufficient for metastasis. Second, metastatic lineages can arise early in tumor development, sometimes long before diagnosis. Therapeutic practice should be guided by the potential early emergence of metastases in tumor development. Lastly, we use molecular sequence data and inferred phylogenetic trees, tumor-type specific data on tumor cell division times, and clinical data on the timings of diagnosis, biopsy, resection, and autopsy, to reveal the temporal order of occurrence of driver mutations, guiding development of targeted therapeutics effective against primary tumors and metastases.

14.3

Negative selection in the cancer genome

Donate Weghorn [1,2], Shamil Sunyaev[1,2]

[1] Brigham and Women's Hospital, Boston, MA, USA, [2] Harvard Medical School, Boston, MA, USA

Seen from an evolutionary perspective, cancer is a highly complex system that evolves asexually under high mutation rates and strong selective pressures. At the onset of cancer, so-called driver mutations enable cells to grow into a tumor and are hence under positive selection. While the search for new driver mutations is crucial to understanding cancer development, little is known about how much cancer relies on a certain subset of genes that is needed for basic cellular processes and survival. These genes are expected to evolve under negative selection, which acts to maintain them functional much like in healthy tissue cells. We study the opposing selective forces in the evolution of cancer tumors to characterize the contribution of each to the observed pattern of substitutions in grown tumors of various types. Our approach takes into account tumor-specific mutational signatures, which are known to be variable across tumor types and patients. Despite the genome-wide average signal of negative selection being largely weak, we find that at the gene level most tumors exhibit detectable negative selection. This enables a probabilistic description of genes in terms of the evolutionary pressure under which they evolve, which may inform targets of cancer therapy among the set of genes whose preserved functionality cancer cannot do without.

14.4

Cancer across the tree of life: Cooperation and cheating in multicellularity

Athena Aktipis[1,6], Amy Boddy [1,6], Gunther Jansen[3,6], Urszula Hibner[7,6], Michael Hochberg[4,6], Carlo Maley[2,6], Gerald Wilkinson[5,6]

[1] Arizona State University, Tempe, AZ, USA, [2] University of California, San Francisco, San Francisco, CA, USA, [3] University of Kiel, Kiel, Germany, [4] CNRS UMR5554, Université Montpellier 2, Montpellier, France, [5] University of Maryland, College Park, MD, USA, [6] Wissenschaftskolleg zu Berlin, Berlin, Germany, [7] CNRS, Université de Montpellier, Montpellier, France

Multicellularity is characterized by cooperation among cells for the development, maintenance and reproduction of the multicellular organism. Cancer can be viewed as cheating within this cooperative multicellular system. Complex multicellularity, and the cooperation underlying it, has evolved independently multiple times. We review the existing literature on cancer and cancer-like phenomena across life, focusing on complex multicellularity but also reviewing cancer like-phenomena across the tree of life more broadly. We find that cancer is characterized by a breakdown of the central features of cooperation that characterize multicellularity, including cheating in proliferation inhibition, cell death, division of labor, resource allocation and extracellular environment maintenance (which we term the five foundations of multicellular cooperation). Cheating on division of labor, exhibited by a lack of differentiation and disorganized cell masses, has been observed in all forms of multicellularity. This suggests that deregulation of differentiation is a fundamental and universal aspect of carcinogenesis that may be underappreciated in cancer biology. Understanding cancer as a breakdown of multicellular cooperation provides novel insights into cancer hallmarks.

14.5

Experimental evolution of motility in cancer cells

Tiffany Taylor, Anastasia Wass, Phil Dash, Louise Johnson
University of Reading, Reading, UK

Experimental evolution has enabled biologists to tackle fundamental questions relating to adaptation, diversification, and evolutionary constraints and trade-offs. Typically, experimental evolution uses microbes, because of their ease of culturing and because their fast replication times and large population sizes allow the evolutionary process to occur quickly. However, cancer cell lines grown in vitro share these convenient characteristics.

Taking an evolutionary view of cancer reveals its similarities with other ecological and evolutionary processes. One example of this is the relatively well-explored parallel between metastasis in cancer and dispersal in whole-organism ecology: a risky behaviour that evolves because it enables the discovery and exploitation of new resources. Theory suggests that competition for resources can be important in driving the evolution of dispersal, and this is supported by data from whole-organism field studies and microbial experimental evolution.

We present results of experimental evolution in cancer cells subjected to novel environments that impose varying levels of competition for resources. In particular, our evolved cancer cell lines show stable, heritable changes in motility, a key trait in metastasis.

Cancer experimental evolution shows promise both as a means of determining the selective forces behind the progress of cancers, and as a model system to further our understanding of the evolution of co-operation, cheating and other complex cell behaviours.

495A

CloneTree: a novel method to infer clone phylogenies based on multi-region sequencing data.

Karen Gomez, Sayaka Miura, Oscar Murillo, Sudhir Kumar
iGEM@Temple University, Philadelphia, PA, USA

Tumor genomes within individuals undergo significant changes over time, a process that can be described by molecular evolutionary principles. Today, scientists determine genetic heterogeneity of tumors by profiling the frequencies of mutations found in different tumors in a person and different parts of the same tumor. We present a new method that uses these data to infer genotypes of cells that comprise tumors (clonotypes), the evolutionary history of clonotypes, and the relative frequencies of different clonotypes in different tumors (or regions). Our method uses molecular phylogeny inference with ancestral state reconstruction to simultaneously infer clonotypes, their phylogeny, and their frequency in tumors. In empirical and computer simulated data analysis, we found that our evolutionary approach performs better than current methods that are based primarily on statistical clustering of mutation frequencies and/or joint treatment of SNV clustering and evolutionary inferences. Our method produced the largest number of correct clonotypes, which contained fewer errors and the least ambiguity. We also found that it is possible to infer the evolutionary tree of clonotypes as well as the clonal composition of tumors from the mutation frequency data. Therefore, the new method will be useful for studying the evolution of tumors within individuals.

496B

Extensive cellular replacement in interior tumor results in extreme genetic diversity

Yawei Li [1], Zheng Hu[1], Xuemei Lu[1], Chung-I Wu[1,2]

[1] Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, [2] Department of Ecology and Evolution, University of Chicago, Chicago, USA

The genetic diversity within tumors is expected to be very large. In our previous work on a human hepatocellular carcinoma (HCC-15), we took 286 samples from a slice of the tumor for whole-exome sequencing/genotyping. Our data analyzed with evolutionary models revealed extreme diversity within a single tumor. However, the growth mode or cell dynamics is largely unknown, which is very important in determining the level of diversity in a tumor. Here we try to simulate the growth of HCC-15 in silico under different models envisaged based on the shape and cell numbers and compare several statistics with the observed in HCC-15. In particular, we devise three 3D growing models: 1) no cell replacement inside the tumor and no cell migration (basic model), 2) no cell replacement inside the tumor but there is cell migration (migration model), 3) frequent cell replacement inside the tumor and no cell migration (replacement model). Based on the information of site frequency spectrum, mutation number distribution and spatial clonal structure, we found the replacement model is more likely to generate the observations in HCC-15 compared with other two models. The results show that cell dynamics even in the dense tumor mass is striking. Moreover our simulations indicate that cell migration is not, at least in this case, a driving factor shaping the clonal diversity. In summary, our simulations based on evolutionary models have revealed extensive cellular replacement in interior tumor, which results in extreme genetic diversity in a tumor.

497C

BIOLOGICAL TRADEOFFS IN CANCER EVOLUTION：A TEST ON FREE-LIVING MAMMALIAN CELLS

Tao Li, Jianlin Liu, Yuezheng Zhang, Chung-I Wu, Xuemei Lu
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Evolutionary theory proposes that the tradeoffs apply to organisms to determine the phenotypic evolution. In cancer evolution, both rapid cell proliferation and cell survival strategies are balanced to propagate in challenging environments. One of the characteristics of tradeoffs is antagonistic pleiotropy: mutations that increases proliferation may decrease survival or vice versa. We performed selections lasted for hundreds of generations from a monoclonal Hela cell line (A-cells). Then we selected cells, one with high survival rate under high cell density condition (S-cells) and the other with fast growth rate in low cell density environment (F-cells). Both selected cells displayed better fitness than A-cells under corresponding selection conditions. The growth rate of F-cells is significantly higher than S-cells under the low-density condition, nonetheless, both of them grew with lower rate under high-density condition, with the F-cells at higher apoptosis rate and in lower G0 phase proportion than the S-cells. When co-culture the F-cells and S-cells, both apoptosis rate and G0 phase proportion of F-cells were significantly higher than itself which cultured separately. It suggests that the proliferation rate of S-cells may be inhibited by mixing with F-cells. The transcriptome data indicate that this phenotypic observation may be caused by Dlg-mediated Hippo-YAP pathway regulation and associated with other key signaling pathways. Whole genome data indicate that the clonal expansion occurred during selection. This study suggests that cancer cells may be subject to tradeoffs between cell proliferation and survival during somatic evolution driven by density-dependent selection. Evolutionary mechanism studies will be further carried out.

498D

Genomic variations and evolution of early-stage liver tumors of HBV transgenic mouse model

Zuyu Yang [1], Mingming Jia[2], Guojing Liu[1], Huaining Hao[1], Qiang Gong[1], Yawei Li[1], Xin Wu[1], Shengdian Wang[2], Xuemei Lv[1], Chung-I Wu[1]

[1] Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, [2] Institute of Biophysics, Chinese Academy of Sciences, Beijing, China

As the development of high-throughput genomic analysis, sequencing the cancer mouse model provides a new opportunity for cancer gene discovery and to understand fundamental mechanisms of tumourigenesis and progression. Here, we sought to characterize the genomic variations in a hepatocellular carcinoma (HCC) mouse model. 12 tumor sections and adjacent non-tumor tissues from four mice were used for whole-exome and/or whole genome sequencing and verifying genotyping. (1) Zero to 17 single nucleotide variations (SNVs) at coding region were found in each tumor. Except *Hras* Q61L mutation shared by three tumors in one mouse, all the other SNVs were only found the sequencing sample suggesting independent origin of the tumors and the great heterogeneity of tumors in the inbred mice. (2) A single *Hras* mutation sharing by the three tumors from 516M1 indicates that a single mutation can drive cell to start proliferation and migration within liver. (3) Diploid karyotype of tumor cells indicates that low frequency mutations found in almost every tumor are results of partial cells mutated, implying that multi-clones cooperation are needed for early stage of tumor cells expansion.

499A

Palindromes as cis-acting mutagenic agents in cancer genes

Lilach Friedman, Einat Hazkani-Covo
Open University of Israel, Ra'anana, Israel

Quasi-palindromes, or imperfect inverted repeats, undergo spontaneous mutation to perfect palindromes. These mutations have been observed in many organisms and in human they are associated with mutations leading to disease. This mechanism occurs via template switching, in which one arm of the repeat serves as a template for synthesis of the second arm and is responsible to at least 5% of the mutations that affect p53 in human cancers.

We analyzed the mutation signature of yeast cell lacking RAD27/FEN1, an important factor of Okazaki fragment maturation and we found that quasi-palindrome to palindrome formation is one of the hallmark of the strain. Since Okazaki fragment maturation is an important process with high mutagenic potential and therefore can lead to cancer associated mutations, we further analyze quasi-palindromic to palindromic conversion in cancer cell.

We used the Catalog of Somatic Mutation in Cancer (COSMIC) database to study the contribution of mutations that are the result of quasi-palindrome to palindrome correction to the overall known mutations in cancer genes. There are ~1500 quasi-palindromes that become perfect in mutated cancer genes. When the number of changes among palindrome arms is two or more there is statistical enrichment in theses events compared to a random set composed of the same mutations randomly sorted to the same samples. The similarity and differences between COSMIC genes and variations in human populations would be further discussed.

500B

A New Mutation-Profile-Based Method for Understanding the Evolution of Cancer Somatic Mutations

Zhan Zhou[1,3], Yangyun Zou[1], Gangbiao Liu[1], Jingqi Zhou[1], Shiming Zhao[1], Zhixi Su [1], Xun Gu[1,2]

[1] State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China, [2] Department of Genetics, Development and Cell Biology, Program of Bioinformatics and Computational Biology, Iowa State University, Ames, USA, [3] Institute of Pharmacology, Toxicology and Biochemical Pharmaceutics, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China

Human genes may perform different functions and exhibit different effects on fitness in cancer and normal cell populations. Here, we present an evolutionary approach to measuring the selective pressure on human genes in cancer and normal cell genomes using the well-known dN/dS (nonsynonymous to synonymous substitution rate) ratio. We develop a new method called the mutation-profile-based Nei-Gojobori (mpNG) method, which applies sample-specific nucleotide substitution profiles instead of conventional substitution models to calculating dN/dS ratios in cancer and normal populations. Using 7,042 exome sequences from tumor-normal pairs, and germline variations from 6,500 exome sequences (ESP6500) as references, we found a significant increase in dN/dS values for human genes from cancer somatic mutations relative to germline substitutions. However, in spite of our strict criteria, we were still able to predict 260 genes that underwent purifying selection in tumor cells but not in germline populations. The lack of significant correlation between dN/dS values of tumor cells and human-mouse orthologous genes implies that the purifying selection in our predicted gene sets may be cancer specific, suggesting that these genes are particularly essential for tumor cells. Moreover, cancer gene expression analysis revealed that of the 260 predicted cancer genes subject to purifying selection, 39 were frequently highly expressed in cancer cells, most of which have been experimentally confirmed to be essential for the survival and proliferation of cancer cells. Therefore, our computation pipeline used to identify cancer purifying selection genes in cancer may provide useful information for detecting potential drug targets or prognostic biomarkers.

501C

Cancer evolution is associated with pervasive positiveselection on globally expressed genes

Sheli Ostrow - Galili [1], Ruth Hershberg[1], Ruth Barshir[2], James DeGregori[3], Esti Yeger-Lotem[2]
[1] Department of Genetics, the Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel, [2] Department of Clinical Biochemistry & Pharmacology, Ben-Gurion University of the Negev, Beer-Sheva, Israel, [3] Department of Biochemistry and Molecular Genetics, Program in Molecular Biology, Integrated Department of Immunology, University of Colorado School of Medicine, Aurora, USA

Cancer is an evolutionary process in which cells acquire new transformative, proliferative and metastatic capabilities. A full understanding of cancer requires learning the dynamics of the cancer evolutionary process. We present here a large-scale analysis of the dynamics of this evolutionary process within tumors, with a focus on breast cancer. We show that the cancer evolutionary process differs greatly from organismal (germline) evolution. Organismal evolution is dominated by purifying selection (that removes mutations that are harmful to fitness). In contrast, in the cancer evolutionary process the dominance of purifying selection is much reduced, allowing for a much easier detection of the signals of positive selection (adaptation). We further show that genes that are globally expressed across human tissues show a very strong signal of positive selection within tumors. Indeed, known cancer genes are enriched for global expression patterns. Yet, positive selection is prevalent even on globally expressed genes that have not yet been associated with cancer, suggesting that globally expressed genes are enriched for yet undiscovered cancer related functions. We find that the increased positive selection on globally expressed genes within tumors is not due to their expression in the tissue relevant to the cancer. Rather, such increased adaptation is likely due to globally expressed genes being enriched in important housekeeping and essential functions. Thus, our results suggest that tumor adaptation is most often mediated through somatic changes to those genes that are important for the most basic cellular functions. Together, our analysis reveals the uniqueness of the cancer evolutionary process and the particular importance of globally expressed genes in driving cancer initiation and progression.

502D

Patterns of single cell phylogenies under different evolutionary models of cancer

Yuantong Ding, Andreas Aristotelous, Edward Patz, Rick Durrett, Allen Rodrigo
Duke University, Durham, NC, USA

Current staging system of tumor based on radiology and pathology is inadequate to predict the tumor progression and patient survival time. However, if we consider tumor as a population of single cells, evolutionary analysis may give insight into the underline process of tumor progression. In this project, we use simulations to test whether phylogenetic structure may help classify the tumor stage more accurately. We built an agent-base tumor growth model, which allows lineage tracking of single cells, and simulated whole genome sequences based on cell lineages with infinite-site model, then preformed phylogenetic and population genetic analysis and evaluate their correlation with the metastasis stage of the tumor. Our results showed that phylogenetic patterns differ under different evolutionary models of cancer, and this approach may have potential application in tumor diagnosis and provide suggestions for experiment design for tumor single cell study.

503A

On cancer risk and the number of stem cell divisions

Imre Derényi, Gergely J Szöllősi
ELTE-MTA "Lendulet" Biophysics Research Group, Budapest, Hungary

Tomasetti and Vogelstein (Science, 2 January 2015, p. 78) demonstrate that cancer risk in different tissues is correlated with the total number of stem cell divisions, and conclude that the majority of cancer risk is attributable to “bad luck”. Here we show that their interpretation of the observed correlation is fundamentally flawed, and a thousand fold variation in risk remains after accounting for differences in the total number of stem cell divisions. We also explore possible explanations for the observed slower then linear increase of cancer risk with the total number of stem cell divisions and argue that this is the result of "Peto's paradox among somatic tissues".

504B

Estimating Trio Model Parameters to Improve Detection of De Novo Mutations

Melissa Ip [1,3], Reed Cartwright[1,2]

[1] The Biodesign Institute, Arizona State University, Tempe, AZ, USA, [2] School of Life Sciences, Arizona State University, Tempe, AZ, USA, [3] Barrett the Honors College, Arizona State University, Tempe, AZ, USA

De novo mutations are unique to an individual and play important roles in morbidity and mortality. Here we develop a trio model for statistically estimating de novo mutations from high-throughput short-read sequencing data of an individual and its parents. The multinomial distribution is used to calculate somatic genotype likelihoods from per-site pileup data. By combining the genotype likelihoods with a Jukes-Cantor-derived model for somatic and germline mutations and an infinite-sites model of segregating variation, we are able to calculate the probability that any site contains a de novo point mutation. Mutation rates and associated parameters for the trio model can be estimated directly from the data using an expectation-maximization algorithm. By using parameters optimized on each dataset, we can produce high quality estimates of sites of de novo mutation, which can be empirically validated. Furthermore, we tested our model on both simulated data and existing genomic datasets to demonstrate its robustness for both parameter estimation and mutation prediction. Our software is useful for personal genomics. For instance, if an individual suffers from a rare metabolic disease, our tool can predict whether they have a de novo mutation in a gene. It can also be used to study cancer mutations by comparing normal and tumor tissue. Beyond medical applications, our software can estimate germline mutation rates and patterns that are important for the study of evolution and population genetics.

15 The evolution of alternative splicing

15.1

Origins and impacts of new mammalian exons

Jason Merkin[1], Ping Chen[2], Maria Alexis[1], Sampsa Hautaniemi[2], Chris Burge [1]
[1] MIT, Cambridge, MA, USA, [2] University of Helsinki, Helsinki, Finland

Mammalian genes are composed of exons, but the evolutionary origins and functions of new internal exons are poorly understood. Here, we analyzed patterns of exon gain using deep cDNA sequencing data from five mammals and one bird, identifying thousands of species- and lineage-specific exons. Most new exons derived from unique rather than repetitive intronic sequence. Unlike exons conserved across mammals, species-specific internal exons were mostly located in 5' untranslated regions and alternatively spliced. They were associated with upstream intronic deletions, increased nucleosome occupancy, and RNA polymerase II pausing. Genes containing new internal exons had increased gene expression, but only in tissues where the exon was included. Increased expression correlated with level of exon inclusion, promoter proximity, and signatures of cotranscriptional splicing. Together these findings suggest that splicing at the 5' ends of genes enhances expression and that changes in 5' end splicing alter gene expression between tissues and between species.

15.2

The importance of homology in alternative splicing

Iakes Ezkurdia[2], Federico Abascal[1], Jose Manuel Rodriguez[1], Juan Rodriguez[1], Enrique Carillo[1], Angela del Pozo[3], Jesus Vazquez[2], Alfonso Valencia[1], Michael Tress [1]
[1] Spanish National Cancer Research Centre (CNIO), Madrid, Spain, [2] Centro National de Investigaciones Cardiovasculares (CNIC), Madrid, Spain, [3] Instituto de Genetica Medica y Molecular, Madrid, Spain

Although eukaryotic cells can express a wide range of alternatively spliced transcripts (the Ensembl version of the human genome currently houses over 70,000 alternative coding variants) it is not clear whether these variants can be defined as dominant or alternative, or whether genes can express a range of transcripts simultaneously across cells. To date large-scale investigations into the pattern of transcript dominance across distinct tissues have produced contradictory results.
In order to shed light on this dilemma we interrogated 8 large-scale human proteomics experiments using a rigorous peptide identification strategy. We carried out an analysis of alternative splicing at the protein level to determine whether genes have a dominant splice variant. While we identified peptides for 64% of protein coding genes, we found peptide for just 280 alternative splicing events, suggesting that the vast majority of protein-coding genes may have a single dominant protein isoform.
The dominant isoforms we identified in the experimental proteomics analyses were overwhelmingly supported by reference isoforms from two completely orthogonal sources, CCDS consensus variants, agreed upon by manual genome curation teams, and APPRIS principal isoforms, predicted automatically from patterns of protein conservation, structure and function.
Our results also suggest that alternative splicing may be controlled at the level of translation. Few of the alternative splicing events we detected disrupt protein functional domains, and over 20% of the identified splice events were generated from highly conserved homologous exon substitutions. The combination of proteomics evidence and ancient origin indicates the importance of homologous exons in alternative splicing.

15.3

The evolution of untranslated regions of mRNAs in primates

Iris Finci [1,3], Anamaria Necsulea[2], Henrik Kaessmann[1,3]
[1] University of Lausanne, Lausanne, Switzerland, [2] École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [3] Swiss Institute of Bioinformatics, Lausanne, Switzerland

Untranslated regions (UTRs) of protein-coding mRNAs are important in many aspects of post-transcriptional regulation, such as translational control, mRNA stability and mRNA subcellular localisation. UTR sequence variation can be achieved through alternative splicing, alternative polyadenylation and alternative promoter usage. However, functional roles and evolutionary dynamics of UTRs remain poorly understood, largely due to incomplete UTR annotations in non-model organisms.

We use extensive RNA-sequencing (RNA-seq) data from 8 tissues across 7 primates (human, chimpanzee, bonobo, gorilla, orangutan, macaque and marmoset) to annotate UTRs and assess their evolutionary patterns. We generated refined annotations for each species using RNA-seq data, adding thousands of new exons and extending known exon boundaries. We then classified exonic sequences as UTR or protein-coding based on their nucleotide substitution patterns, open reading frames and similarity to known proteins. We also included splicing information to define the exon-intron structure of UTRs across tissues and species.

Using this comprehensive dataset, we screened for differential UTR usage across tissues and species. Interestingly, we observed higher differential exon usage in UTRs compared to coding exons in all primates, with many cases of tissue-specific and lineage-specific UTRs. Moreover, we observe that, unlike gene expression, alternative UTR usage evolves rapidly, similarly to alternative splicing. Currently, we are assessing the functional relevance of lineage-specific UTRs by analysing changes in their miRNA binding potential and nonsense-mediated decay susceptibility.

Besides providing new insights into the functional evolution of UTRs, we believe our annotations provide a highly valuable resource for future investigations of post-transcriptional gene regulation.

15.4

Appearance and fixation of alternative splicing events during evolution and their impact on the protein structures

Adel Ait-hamlat, Laine Elodie, Hugues Richard
UPMC, Paris, France

Alternative splicing (AS) is the mechanism by which several distinct proteins, or isoforms, are produced from the same gene. AS is common in higher eukaryotes and is recognized to be essential for the functional diversification of the proteome. Although AS mechanisms have been widely studied at the genomic level, its role in the evolution of protein structures and functions remain largely uncharted.

We have developed a computational method to infer evolutionary scenarios that can explain transcripts observed within a set of species. The method combines a transcripts phylogeny reconstruction algorithm with a protein structure prediction routine. It is implemented in an automated tool called transPhyl. By applying transPhyl to the c-Jun N-terminal kinases (JNKs) family, we were able to propose a plausible evolutionary scenario explaining the transcripts observed in seven species, ranging from nematode to human. In particular, we could date the appearance and fixation of an ASE that resulted in the production of isoforms differing by two mutually exclusive exons. These isoforms were shown to have differential binding affinity to their substrates. By mapping this ASE onto the predicted structural models we could identify and characterize at atomic level the molecular determinants of the selectivity of JNKs isoforms to their targets.

15.5

Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible

Liana Lareau, Steven Brenner
University of California, Berkeley, Berkeley, CA, USA

Ultraconserved elements, unusually long regions of perfect sequence identity, are found in genes encoding numerous RNA-binding proteins including SR splicing factors. Expression of these genes is regulated via alternative splicing of the ultraconserved regions to yield mRNAs that are degraded by nonsense-mediated mRNA decay (NMD), a process termed unproductive splicing (Lareau et al. 2007; Ni et al. 2007). As all human SR genes are affected by alternative splicing and NMD, one might expect this regulation to have originated in an early SR gene and persisted as duplications expanded the SR family. But in fact, unproductive splicing of most human SR genes arose independently. This paradox led us to investigate the origin and proliferation of unproductive splicing in SR genes. We demonstrate that unproductive splicing of the splicing factor SRSF5 (SRp40) is conserved among all animals and even observed in fungi; this is a rare example of alternative splicing conserved between kingdoms, yet its effect is to trigger mRNA degradation. As the gene duplicated, the ancient unproductive splicing was lost in paralogs, and distinct unproductive splicing evolved rapidly and repeatedly to take its place. SR genes have consistently employed unproductive splicing, and while it is exceptionally conserved in some of these genes, turnover in specific events among paralogs shows flexible means to the same regulatory end.

637A

The roles of splicing, expression, protein structure, and DNA methylation in exonic sequence evolution

Feng-chi Chen [1], Trees-Juen Chuang[2], Ben-Yang Liao[1], Chia-Lin Pan[1], Hsuan-Yu Lin[1], Andrew Ying-Fei Chang[1], Gideon Wu[1]

[1] National Health Research Institutes, Zhunan, Miaoli County, Taiwan, [2] Academia Sinica, Taipei, Taiwan

What determines the rate of sequence evolution is a fundamental issue in evolutionary biology. Evolutionary rates differ not only between genes but also between exons of the same genes. A number of factors are associated with such intragenic variations in evolutionary rate, including but not limited to whether the exon is alternatively spliced, exonic expression level, structural-functional constraints, and the level of DNA methylation. How these factors individually and jointly affect exonic evolutionary rates is an issue of great interest. We demonstrate that firstly, splicing and protein structural disorderness affect the evolutionary rates of mammalian exons independently although disordered protein regions tend to be encoded by alternative exons. Secondly, splicing and structural-functional constraints appear to be more influential than exonic expression level in determining exonic evolutionary rates in mammals. But in plants splicing plays a less important role in this regard. Thirdly, CpG methylation affects the evolution of exons in a context-dependent manner. In mammals, the mutagenic effect is the strongest at first exons but selection seems to suppress substitutions at heavily methylated internal exons. In plants, by contrast, the strongest mutagenic effect of CpG methylation occurs in internal exons despite the stringent selective constraints on this exon group. Overall, our results suggest that exon evolution is simultaneously affected by a multiplicity of biological factors that work either independently or interactively. Interestingly, the evolutionary effects of these factors differ considerably between animals and plants, possibly reflecting the divergences in biological significance of these factors between the two lineages.

638B

Extreme developmental temperatures trigger decanalization of alternative splicing

Ana Marija Jakšić [1,2], Christian Schlötterer[1]

[1] Institut für Populationsgenetik, Vetmeduni Vienna, Wien, Austria, [2] Viennna Graduate School of Population Genetics, Wien, Austria

Canalization describes the robustness of phenotypes to either genetic or environmental perturbations. It ensures the stability of developmental phenotype and can affect different levels of phenotype, from gene expression to final phenotypic levels. In our study, we investigate genetic canalization of alternative splicing contransting two genetically different laboratory strains of *Drosophila melanogaster* (Oregon-R and Samarkand) at four different developmental temperatures (13ºC, 18ºC, 23ºC and 29ºC) and environmental canalization across these four developmental temperatures. We find that genetic canalization is highly dependent on the environmental conditions. While at 18ºC only minor differences in alternative splicing are noted, more extreme temperatures resulted in strong genetic decanalization. Similarly, environmental decanalization was the strongest at extreme temperatures, but differed markedly among genotpyes. The differences in alternative splicing between the two strains follow largely a dominant mode of inheritance. Since we previously found the same pattern of genetic canalization and dominance for gene expression, but for completely different sets of genes, we will discuss the possible differences in underlying genetic basis of the canalization of alternative splicing and canalization of gene expression.

639C

U1 snRNP Guides Splice Site Recognition in the Intragenic Territories of *Drosophila melanogaster*

Gildas Lepennetier, Francesco Catania
Institute for Evolution and Biodiversity, University of Muenster, Muenster, Germany

How the spliceosome selects splice sites in eukaryotic genes is still a topical subject in RNA splicing research. At present, it is widely accepted that splice sites in eukaryotes are recognized via the mechanisms of intron definition or exon definition. Briefly, in intron definition, the spliceosome targets short introns, while in exon definition it focuses on short exons. Recently, an additional model for splice site recognition – U1-dependent definition – has been proposed. This model integrates previous observations in support of intron and exon definition with the notion that splicing factors and cleavage/polyadenylation factors interact antagonistically during the process of transcription. We investigated the well-annotated *D. melanogaster* genome aiming to validate and extend the U1-dependent definition model. We observe distinct – and as of yet unreported – patterns of coevolution between intron and next exon sizes, variations in splice site strength and in the occurrence of cryptic poly(A) sites along genes that suggest the existence of distinct *intragenic territories* which are subject to different molecular environments and selective pressures. These observations support the U1-dependent definition model for splice site recognition. Our results bring a new vision on the evolution of gene structure, revealing an important role for non-adaptational intracellular forces in shaping gene architecture. While considerable work needs to be done before the 'splicing code' is totally unraveled, our observations and the theoretical model that we propose offer a basis for future studies on other organisms.

640D

Evolution of gene structural complexity: An alternative-splicing based model accounts for intron-containing retrogenes

Chengjun Zhang [1], Andrea R. Gschwend[1], Yidan Ouyang[1], Manyuan Long[1]
[1] Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA, [2] National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, HuBei, China

The structure of eukaryotic genes evolves extensively by intron loss or gain. Previous studies have revealed two models for gene structure evolution through the loss of introns: RNA based gene conversion, dubbed the Fink model and retroposition model. However, retrogenes that experienced both intron loss and intron retaining events have been ignored; evolutionary processes responsible for the variation in complexity exon-intron structure were unknown. We detected hundreds of retro-duplication derived genes in the human, Drosophila, rice, and Arabidopsis genomes and categorized them either as duplicated genes which have all introns lost (IL-type) or as duplicated genes which have at least lost one and retained one intron compared to the parental copy (IR-type). Our new model attributes intron retention alternative splicing (AS) to the generation of these IR-type gene pairs. We presented 25 parental genes that have an intron-retention isoform and have retained introns in the same locations in the IR-type duplicate genes, which directly support our hypothesis. Our AS-based model in conjunction with the retroposition and Fink models, can explain the IR-type gene observed. We discovered a greater percentage of IR-type genes in plants than in animals, which may be due to the abundance of intron retention cases in plants. AS-based model can explain the evolution of complex gene structures in eukaryotic genomes and complements previous models. Given the prevalence of alternative splicing in eukaryotes, this new model adds great explanatory power to the understanding of the evolutionary mechanisms behind the exon-intron structures of genes.

641A

Analysis of the doublesex genes of Bactrocera tua

Thanaset Thongsaiklaing [1,2], Lertluk Ngernsiri[3]

[1] Center for Agricultureal Biotechnology, Kasetsart University, Kampaengsaen Campus, Nakhon Pathom, Thailand, [2] Center for Excellence on Agriculturral Biotechnology: (AG-BIO/PERDO-CHE), Bangkok, Thailand, [3] Faculty of Science, Kasetsart University, Bangkok, Thailand

The doublesex (dsx) is the bottom gene in the sex determination cascade. The gene is highly conserved among insects. Here, the Btdsx gene of Bactrocera tua is identified. The Btdsx pre-mRNA is alternatively spliced in sex specific manner, resulting in male and female Dsx protein isoforms. Interestingly, Btdsx is spliced into three transcripts which encode two female-specific (BtdsxF1 and BtdsxF2) and one male-specific (BtdsxM) isoforms. The complete cDNA sequences of BtdsxF1, BtdsxF2 and BtdsxM are 3546, 3651 and 3575 bp, respectively. Even the difference in nucleotide sequence, both BtdsxF1 and BtdsxF2 have the same length of open reading frames (942 bp) encoding 313 amino acid residues. While the 1206 bp BtdsxM ORF encods 401 amino acid residues. Both the nucleotide and the amino acid identities between the BtdsxF1 and BtdsxF2 are of the same percentage, 92 %. Knock down the expression of Btdsx using RNAi experiment by injecting early emergent adult flies with dsRNA of Btdsx affect the size of ovary but has no any effect on testis.

642B

Transcriptome-wide N6-methyladenosine profiling of rice callus and leaf reveals the presence of tissue specific competitors involved in selective mRNA modification

Yuli Li, Xiliang Wang, Cuiping Li, Songnian Hu, Jun Yu, Shuhui Song
Beijing Institute of Genomics, CAS, Beijing, China

N6-methyladenosine (m^6A) is the most prevalent internal modification present in mRNAs of all higher eukaryotes. With the development of MeRIP-seq technique, in-depth identification of mRNAs with m6A modification becomes feasible. Here we present a transcriptome-wide m^6A modification profiling effort for rice transcriptomes of differentiated callus and leaf, which yields 8,138 and 14,253 m^6A-modified genes, respectively. The m^6A peak (m^6A modified nucleotide position on mRNAs) distribution exhibits preference toward both translation termination and initiation sites. The m^6A peak enrichment is negatively correlated with gene expression and weakly positively correlated with certain gene features, such as exon length and number. By comparing m^6A-modified genes between the 2 samples, we define 1,792 and 6,508 tissue-specific m^6A-modified genes (TSMGs) in callus and leaf, respectively. Among which, 626 and 5,509 TSMGs are actively expressed in both tissues but are selectively m^6A-modified (SMGs) only in one of the 2 tissues. Further analyses reveal characteristics of SMGs: (1) Most SMGs are differentially expressed between callus and leaf. (2) Two conserved RNA-binding motifs, predicted to be recognized by PUM and RNP4F, are significantly over-represented in SMGs. (3) GO enrichment analysis shows that SMGs in callus mainly participate in transcription regulator/factor activity whereas SMGs in leaf are mainly involved in plastid and thylakoid. Our results suggest the presence of tissue-specific competitors involved in SMGs. These findings provide a resource for plant RNA epitranscriptomic studies and further enlarge our knowledge on the function of RNA m^6A modification.

643C

Induced transcription and stability of CELF2 mRNA drives widespread alternative splicing

Samuel Allon, Michael Mallory, Kristen Lynch
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Diverse patterns of mRNA processing allows for dramatic differences in phenotype across cell types and within cell lineages at different stages of development. A deeper understanding of how this diversity evolved will require detailed knowledge of the mechanisms of mRNA processing and the pathways that regulate them. Here we demonstrate that expression of the splicing factor CELF2 is regulated in response to T cell signaling through combined increases in transcription and mRNA stability. Transcriptional induction occurs shortly after stimulation and is dependent on activation of signal-dependent transcription factors. Subsequently there is an increase in the stability of the CELF2 mRNA that correlates with a change in CELF2 3'UTR length and contributes to the total signal-induced enhancement of CELF2 expression. Importantly, we uncover dozens of splicing events in cultured T cells whose changes upon stimulation are dependent on CELF2 expression, and provide evidence that CELF2 controls a similar proportion of splicing events during human T cell development. Together these findings uncover novel mechanisms of regulation for a prominent splicing factor, quantify the impact that this splicing factor has in T cell development, contrast the roles a single splicing factor can have across multiple cell types, and present a model of evolution operating on multiple layers of gene regulation to achieve a precise expression level.

644D

Evolutionary dynamics of mammalian circular RNAs

Franziska Gruhl [1,2], Peggy Janich[1], David Gatfield[1], Henrik Kaessmann[1,2]
[1] Center for Integrative Genomics, University of Lausanne, Lausanne, Vaud, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Vaud, Switzerland

The transcriptional landscape of the mammalian genome comprises a variety of different well-known RNA types, such as protein-coding mRNAs, long noncoding RNAs or microRNAs. Spatial and temporal changes in expression of these RNAs are thought to have contributed to the evolution of species- or lineage-specific phenotypic features. Circular RNAs (circRNAs) were recently discovered to represent a rather abundant class of transcripts. However, given that previous studies assessed circRNAs mainly in cell lines from a few individual species, the evolutionary dynamics of circRNAs remain poorly understood. To study the functional and evolutionary relevance of circRNAs, we generated comprehensive RNA sequencing datasets (total RNA; enzymatically-enriched for circRNAs) for three organs (cerebellum, liver, testis) across five species (human, rhesus macaque, mouse, rat, opossum) that represent three mammalian lineages (primates, rodents, marsupials). We combined experimental and computational approaches to thoroughly predict and annotate circRNAs on a genome-wide scale across these species on the basis of these data. I will present initial insights obtained from detailed comparative analyses of these extensive novel catalogs of mammalian circRNAs.

645A

No evidence for avoidance of exonic splice enhancers in intronless genes

Rosina Savisaar, Laurence Hurst
University of Bath, Bath, UK

Exonic splice enhancers (ESEs) are short nucleotide motifs that occur in clusters at exon ends. Their best-characterized function is to bind splice factors that attract the spliceosome to an exon, thereby promoting its inclusion in the mature transcript. We hypothesized that the occurrence of ESEs within intronless transcripts could lead to the inappropriate recruitment of the splicing machinery and, ultimately, to the production of an aberrant mRNA. Intronless genes should therefore be under selection to avoid ESEs. To test the latter hypothesis, a set of human intronless protein-coding genes, pruned of paralogs, was examined. Surprisingly, intronless genes were significantly enriched in ESEs when compared to random expectations ($p \approx 0.0003$). Moreover, controlling for differences in nucleotide composition, intronless coding regions were as dense in ESEs as intron-containing sequences ($p \approx 0.78$ for difference). In order to determine whether the motifs found in intronless genes were evolving particularly fast, suggesting avoidance, K4 from alignment to macaque was calculated. K4 was found to be lower in actual ESE regions than in nucleotide-controlled simulations, with a p-value approaching significance ($p \approx 0.055$). When sample size differences were controlled for, the rate of evolution of ESEs in intronless genes could not be distinguished from that in intron-containing genes ($p \approx 0.95$ for difference), suggesting selection for maintenance rather than avoidance. Our results are consistent with previous work showing SR-proteins, which bind ESEs, to have splicing-independent roles. We find no support for the hypothesis that intronless genes are under selection to avoid ESE motifs.

646B

The evolution of alternative splicing among closely related Nicotiana species

Zhihao Ling, Thomas Brockmüller, Ian Baldwin, Shuqing Xu
The Max Planck Institute for Chemical Ecology, Jena, Germany

Alternative splicing (AS) is widespread in higher eukaryotes and contributes to different adaptation processes. However, the evolution of AS remains unclear, especially in plants. The genus Nicotiana (Solanaceae) is widely distributed and adapted to different local environment. In this study, we took the advantage of recently sequenced Nicotiana attenuata genome, investigated the evolution of AS in leaves of six closely Nicotiana species that showed diverged leave phenotypes using mRNA sequencing. We specifically ask two questions: 1) among closely related species, does AS evolve faster than gene expression (GE)? 2) Which genes are rapidly evolving at AS level and which genes have conserved AS among different species? Furthermore, because Nicotiana genus experienced a whole genome triplication event shared with other solanaceous species, we are also investigating the impact of gene duplication to the evolution of AS in Nicotiana. Our preliminary data show that among different species, the divergence at both GE and AS level increases with phylogenetic distance, but the evolutionary changes at AS level is much more conserved than at GE level. Our study will provide new insights on the evolution of AS in plants.

647C

Alternative splicing of the *timeless* gene in *Drosophila montana*

Maaria Kankare [1], Riikka Tapanainen[1], Kalevi Trontti[2]
[1] *University of Jyvaskyla, Jyvaskyla, Finland,* [2] *University of Helsinki, Helsinki, Finland*

Alternative splicing, where one gene can produce a variety of different transcripts and eventually proteins, is known be a very common event in all eukaryotes. This process is known to promote evolutionary potential of the organisms but the knowledge on its role in adaptation is still scarce. This research aims to clarify how alternative splicing of an important circadian clock gene, *timeless* (*tim*) is connected to adaption of a northern malt fly species to seasonally varying environments. In *Drosophila melanogaster* light is known to stimulate the expression of tim with thermosensitive splicing at cold but not at warm temperature suggesting a seasonal adaptive response.

We determined the splicing patterns of tim and studied the frequency of different transcripts in *D. montana* at different temperatures and light conditions. We used flies originally collected from two extremes of a cline in USA (Fairbanks, Alaska; 64°55'N and Azalea, Oregon; 42°48'N) and in Finland (Pyhätunturi; 67°06'N and Lahti; 60°59'N). Flies were maintained in two different temperatures (16°C and 19°C) and light dark cycles (24:0 and 22:2) from egg to adults. We determined different splicing forms with molecular cloning and Sanger sequencing and measured the expression level differences of the transcripts with qPCR. We discuss the differences in splicing patterns and the effect of temperature and light-dark cycles between and among populations from the same and different continents.

648D

Analysis of alternative splicing in *Drosophila* using RNA-sequence data

Lauren Gibilisco, Qi Zhou, Shivani Mahajan, Doris Bachtrog
UC Berkeley, Berkeley, CA, USA

Alternative splicing creates transcriptome and proteome diversity and has been widely studied in mammals and Drosophila melanogaster, but much less is known about alternative splicing in other Drosophila species. Our current work investigates alternative splicing in D. pseudoobscura, D. miranda, D. albomicans, and D. nasuta. We aim to form a better understanding of the magnitude and evolution of alternative splicing in Drosophila by using RNA-sequence data from different tissues, stages, and sexes to characterize alternative splicing within and between species.

649A

Evolution of Alternative Splicing in Ubiquitin Specific Protease 4

Caitlyn Vlasschaert [1,2], Douglas Gray[1,2], Xuhua Xia[1,3]

[1] University of Ottawa, Ottawa, Canada, [2] Ottawa Hospital Research Institute, Ottawa, Canada, [3] Ottawa Institute of Systems Biology, Ottawa, Canada

Background: Ubiquitin specific protease 4 (USP4) is a highly networked deubiquitinating enzyme with reported roles in cancer, innate immunity and RNA splicing. In mammals it has two dominant isoforms arising from inclusion or exclusion (skipping) of exon 7 (E_7).

Objective: We studied two plausible mechanisms for the generation of these isoforms: (I) E_7 skipping due to the long upstream (I_6) intron of mammalian USP4 and (II) E_7 skipping due to inefficient 5' splice sites (5'SS) and/or branchpoint sites (BPS).

Methods: Employing a combinatorial in silico and in vitro approach, we first derived a predictive framework by characterizing relative intron lengths and splice site strengths bioinformatically. We then generated a series of minigene constructs to pinpoint the mechanism of exon skipping in mammals.

Results: Both transcript variants were generated from a USP4-E_7 minigene construct with short flanking introns, an observation consistent with the second mechanism whereby differential splice signal strengths are the basis of E_7 skipping. This mechanism was confirmed by optimization of sequence elements through site-directed mutagenesis. Bioinformatic analysis predicted that exon skipping would not occur in two vertebrate species. This prediction was confirmed experimentally. Optimization of 5'SS_7 by site directed mutagenesis eliminates E_7 skipping.

Conclusions: Our study illustrates the power of combining bioinformatic prediction and experimental verification in elucidating alternative splicing mechanisms and highlights the potential of quantitative approaches in modelling and predicting the occurrence of alternative splicing.

650B

Studying the effect of alternative splicing variations in eukaryotes

Emilie Chautard [1,2], Clara Benoit-Pilven[1], Léon-Charles Tranchevent[1], Camille Marchet[2], Vincent Lacroix[2], Didier Auboeuf[1]

[1] Centre de Recherche en Cancérologie de Lyon, Inserm U1052, CNRS UMR5286, Centre Léon Bérard, Université de Lyon, F-69008 Lyon, France, [2] UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, INRIA Erable, Université Claude Bernard, F-69100 Villeurbanne, France

Recent genome-wide analyses reveal that more than 90% of all human genes are regulated by alternative splicing. Interestingly, alternative splicing has been recently demonstrated in unicellular organisms such as *Capsaspora owczarzaki*. We have previously shown than human epithelial, endothelial and fibroblast cells exhibit specific splicing programs independently of their tissues of origin. Genes regulated by alternative splicing show an enrichment in cytoskeleton, cell adhesion, and motion programs, which are the main features distinguishing these cell types. We wish to extend this analysis to other muti- and uni-cellular eukaryotes to better understand what are the effect of alternative splicing on cell type-specific programs.
To meet our objectives, we developed a bioinformatics pipeline to annotate, quantify and visualize ASEs from RNA-Seq data. Our pipeline combines two different in-house approaches that produce complementary results. The first approach is based on read alignments to a reference genome, which can be produced from various spliced-aware aligners. The second approach is based on KisSplice, a local transcriptome assembler developed in the LBBE team. Our pipeline is able to identify new or known ASEs, even if a reference genome and its annotations are missing or incomplete. In order to identify the possible impact of an alternative splicing variation, we have developed FasterDB, a web based interface that can be used to investigate alternative splicing events (i) by predicting the functional domains that are impacted by each variation and (ii) by analyzing them globally at the pathway level to identify the highlevel functions that may be impacted.

16 Exploring the consequences of ancient and contemporary gene flow

16.1

Excavating archaic hominin DNA from the genomes of modern humans

Joshua Akey
Department of Genome Sciences, University of Washington, Seattle, USA

Anatomically modern humans overlapped and mated with Neanderthals, Denisovans, and possibly unknown archaic hominins. We developed an approach to identify archaic hominin lineages that persist in the DNA of modern humans, and applied it to whole-genome sequences from 379 European and 286 Asian individuals. In total, we recovered over 15 Gb of introgressed sequence that spans ~20% of the Neanderthal genome (FDR = 5%). Analyses of surviving archaic lineages suggests that there were fitness costs to hybridization, admixture occurred both before and subsequent to divergence of non-African modern humans, and Neanderthals were a source of adaptive variation for loci involved in skin phenotypes, including regulatory sequences of the BNC2 and POU2F3 genes. Furthermore, we have recently sequenced 35 individuals of Melanesian ancestry and have constructed a map of surviving Denisovan lineages. Strikingly, many of the deserts of Neanderthal sequence are also depleted of Denisovan sequence, refining genomic regions that may harbor genetic changes that contribute to uniquely modern human traits. Our results provide a new avenue for paleogenomics studies, allowing substantial amounts of population-level DNA sequence information to be obtained from extinct groups, even in the absence of fossilized remains.

16.2

Introgressed Neandertal alleles contribute to gene expressiondifferences among present-day humans.

Michael Dannemann [1], Aida Andrés[1], Janet Kelso[1], Philipp Khaitovich[2], Ekaterina Khrameeva[2], Kay Prüfer[1]

[1] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2] CAS-MPG Partner Institute for Computational Biology, Shanghai, China

Recent genomic studies have shown that Neanderthals admixed with modern humans, and that as a result approximately 2% of the genomes of all present-day non-Africans are of Neandertal origin.

Evidence that introgression from Neanderthals may have contributed to human phenotypic diversity has come from studies of specific immune loci, and from a genome-wide map of Neandertal ancestry in present-day non-Africans which showed that skin and hair -associated genes are enriched in genomic regions with high Neandertal ancestry. However, beyond these studies, the functional implications of Neanderthal introgressed alleles, particularly the contribution of Neandertal alleles in non-coding regions, remains largely unknown.

We have explored the extent to which Neandertal alleles contribute to differences in gene expression between present-day non-African individuals by correlating Neanderthal alleles in non-Africans with changes in gene expression.

To determine those expression changes that have functional relevance we analysed the subset of changes for which the linked Neanderthal haplotype has been associated with a particular phenotype via genome-wide association studies in humans.

Our results suggest a substantial contribution of Neanderthal haplotypes to differences in human gene expression that are associated with immune and metabolic function.

I will present specific candidate loci for which there is strong evidence for the contribution of multiple archaic haplotypes to GWAS-validated phenotypic differences, and argue that these may be best explained by the influence of Neanderthal alleles on gene regulation.

16.3

Comprehensive phenome-wide association analysis of Neanderthal introgression supports its relevance to disease risk in modern humans

Corinne Simonti [1], Benjamin Vernot[2], Joshua Denny[1], Joshua Akey[2], John A. Capra[1]
[1] Vanderbilt University, Nashville, TN, USA, [2] University of Washington, Seattle, WA, USA

Gene flow from archaic hominins—such as Neanderthals and Denisovans—into the predecessors of modern day human populations is theorized to have helped those populations adapt to new environments. However, resolving the impact of archaic hominin introgression has been challenging due to the difficulty of identifying introgressed alleles and the expense of testing if these alleles are associated with phenotypes in modern human populations. To overcome these limitations, we identified a high confidence set of Neanderthal introgressed alleles in a cohort of ~28,000 adult individuals of European ancestry with genotype data linked to electronic medical records (EMRs). We defined cases and controls for more than 1,500 clinical phenotypes derived from the use of ICD-9 billing codes in the EMR. We tested ~1,500 Neanderthal SNPs for associations with these phenotypes and found associations with bone mineral density, body mass index, abnormal pap smears, visual field defects, hypercoagulable state, and several others. In addition to these individual SNP associations, we used genome-wide complex trait analysis (GCTA) to determine whether Neanderthal ancestry overall was associated with variation in relevant phenotypes. Neanderthal ancestry explains a significant fraction of risk for myocardial infarction and depression ($p < 0.01$) beyond overall genomic risk. We are now replicating these results in independent cohorts. Overall, our results support current theories about the importance of archaic hominin introgression in adaptation to new environments and suggest that their legacy impacts disease risk in modern humans.

16.4

Genetic consequences of hybridization of diverged stocks of endangered Australian freshwater fish after translocations early in the 20th century

Alexandra Pavlova [1], Stephen Amish[8], Luciano B Beheregaray[6], Rhys Coleman[7], Dean Gilligan[5], Brett A Ingram[3], Joanne Kearns[2], Annika M Lamb[1], Mark Lintermans[4], Gordon Luikart[8], Jarod P Lyon[2], Shawn Narum[9], Minami Sasaki[6], Zeb Tonkin[2], Paul Sunnucks[1]

[1] Monash University, Melbourne, VIC, Australia, [2] Arthur Rylah Institute, Land, Fire and Environment, Department of Environment and Primary Industries, Heidelberg, VIC, Australia, [3] Fisheries Victoria, Regulation and Compliance, Department of Environment and Primary Industries, Alexandra, VIC, Australia, [4] University of Canberra, Canberra, ACT, Australia, [5] NSW Department of Primary Industries – Fisheries NSW, Batemans Bay, NSW, Australia, [6] Flinders University, Adelaide, SA, Australia, [7] Melbourne Water, Docklands, VIC, Australia, [8] University of Montana, Missoula, MT, USA, [9] CRITFC, Hagerman, ID, USA

Understanding how selectively neutral and adaptive genetic variation is distributed across the species range could help management to develop targeted approach for maximizing species persistence into the future. The Australian freshwater fish Macquarie perch Macquaria australasica is an endangered species whose range shrank to small and isolated headwater populations after rivers of its native range became dammed, regulated and invaded by non-native fish.
To mitigate population declines and reduce extinctions, translocations of mature Macquarie perch into natural and novel habitats were performed by fisheries through the 20th century. Some of these translocations involved movement of fish across major river basins, so that individuals from populations previously historically isolated and potentially diverged under drift and selection for contrasting environments (e.g. summer temperatures) were mixed. This practice is now ceased and replaced by supplementation of non-native populations by fingerlings bred in hatcheries to support recreational fisheries.
Here, using a range of molecular markers (mtDNA, nuclear microsatellites, RAD tags) we explore genetic consequences of these translocations, with reference to levels of neutral and potentially adaptive genetic diversity and population structure in the native range of Macquarie perch. We found that fish translocated to a novel environment have contributed genetic variation to the local population, contrary to expectations of outbreeding depression, but in agreement with observations of a non-declining population. This hybridization is not sex-biased and extends beyond first generation. One of the major management implications of this finding is a positive role translocations played in elevating genetic diversity and potentially improving population health.

16.5

The roles of shared ancestry and shared genes in niche adaptation in *Aeromonas*

Johann Peter Gogarten [1,2], Matthew S. Fullmer[1], Sophie M. Colston[1], Joerg Graf[1,2]
[1] *Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA,*
[2] *Institute for Systems Genomics, University of Connecticut, Storrs, Connecticut, USA*

Aeromonads are found in a number of fresh and brackish water habitats and in association with various animals, ranging from beneficial symbionts of leeches and zebrafish to pathogens of amphibians, fish, and humans. We previously found high rates of HGT between *Aeromonas* species, and mosaic 16S rDNA due to recombination between divergent copies. Nevertheless, whole draft genome sequences from 56 *Aeromonas* genomes analyzed with ANI and *in silico* DNA-DNA hybridization, MLSA, and a phylogeny calculated from 2,710 expanded core genes of this genus revealed that the recent evolutionary history (the assignment of strains to species and most sister species relationships) could be unambiguously inferred using the different approaches.

Against this backdrop we examine over 100 Aeromonas genomes for genes responsible niche specialization. We identify genes encoding sialic acid utilization that are shared between strains of *Aeromonas veronii, A. hydrophila, A. fluvialis, A. allosaccharophila* and *A. jandaei* isolated from leech, wound infections following leech therapy, and which are largely absent in environmental isolates. Phylogenetic reconstruction of the sialic acid utilization operon shows that *A. fluvialis* and *A. jandaei* recently acquired it through HGT from different *A. veronii* strains. These genes appear to play a role in adaptation to the leech gut, which is rich in sialic acid containing complex oligosaccharides (leech mucin and, after a blood meal, glycoproteins in the erythrocyte plasma membrane). The absence of the sialidase encoding gene in several of the sialic acid utilization operons also illustrates the dependence of some strains on common goods produced by others.

16.6

Probabilities and patterns of adaptive introgression

Joachim Hermisson [1], Derek Setter[1], Hildegard Uecker[1,2], Jeffrey Jensen[3]
[1] University of Vienna, Vienna, Austria, [2] IST Austria, Klosterneuburg, Austria, [3] EPFL Lausanne, Lausanne, Switzerland

Gene flow across emerging reproductive barriers from a strongly diverged sister population or "species" is increasingly recognized as an important source for adaptation. However, adaptive introgression alleles will typically enter the recipient population with linked alleles that are deleterious in the new environment or genomic background. Successful adaptive introgression thus requires that deleterious material is removed from the introgression haplotype by recombination. Alternatively, weaker deleterious alleles can also hitch-hike to fixation along with the adaptive type.

In a first part, we present analytical results based on branching process theory to show how the introgression probability and the probability for deleterious hitch-hiking depends on the selection coefficients and the linkage pattern among the positively and negatively selected loci. In the second part, we employ coalescent theory and simulations to study the expected footprint of adaptive introgression that results from these events. We find significant differences from classical selective sweeps.

16.7

Detecting adaptive introgression in humans

Emilia Huerta-Sanchez [1], Fernando Racimo[2], Rasmus Nielsen[2]
[1] University of California Merced, Merced, CA, USA, [2] University of California Berkeley, Berkeley CA, USA

Comparisons of DNA from archaic and modern humans show that modern and archaic humans interbred, and in some cases received an evolutionary advantage from doing so. This process, adaptive introgression (AI), may lead to a faster rate of adaptation than is predicted from models with mutation and selection alone. Within the last couple of years, a series of studies have identified regions of the genome that are likely examples of AI. In many cases, once a region was ascertained as being introgressed, commonly used statistics based on both haplotype as well as frequency information were employed to test for positive selection. Introgression by itself, however, changes both the haplotype structure and the distribution of allele frequencies, and these are the patterns that many methods use to detect selection. Therefore, patterns generated by introgression alone may lead to false inferences of positive selection. Here we use simulations to investigate the false positive rate of these statistics under null models that include introgression and examine the data to discover new statistics that can specifically differentiate between introgression and adaptive introgression. We then examine the 1000 Genomes data and the HGDP dataset to identify regions under adaptive introgression.

16.8

Evaluating genomewide introgression in *Drosophila yakuba* and *D. santomea*

Ana Llopart
University of Iowa, Iowa City, IA, USA

Interspecific hybridization provides the unique opportunity for species to tap into the genetic variation of another closely related species. A previous analysis of complete mitochondrial genomes in *Drosophila yakuba* and *Drosophila santomea*, two sister species that started diverging 400,000 years ago and are known to produce hybrids (ca. 3%) in nature today, indicated that mitochondrial introgression is recurrent in this system. To evaluate gene flow in nuclear genes, I obtained several sequences of *D. yakuba* and *D. santomea* complete genomes. I uncovered discrete areas of the *D. santomea* genome that have trespassed species boundaries and produced multiple waves of introgression. Here I present a study of the temporal dynamics of the accumulation of genic incompatibilities within a single lineage and discuss its implications on genomewide models of speciation.

16.9

450 diverse high coverage whole genome sequences reveal ancient population admixture in modern human populations

Luca Pagani
on behalf of all authors, Tartu, Estonia

Complete high-coverage individual genome sequences carry the maximum amount of information for reconstructing the evolutionary past of a species in the interplay between random genetic drift and natural selection. Here we present a novel dataset of 450 human genomes from 156 populations that represent a dense geographic coverage of Eurasia. The genomes, chosen to be representative of each population based on SNP-chip data, were sequenced at 40X on the same platform (Complete Genomics) and processed on a uniform bioinformatics pipeline.

Our dataset has an unprecedented combination of high spatial and genomic coverage. This enabled us to refine current knowledge on continent-wide patterns of heterozygosity, long and short range gene flow, archaic admixture, and changes in effective population size over time.
In particular, we have clarified the admixture dynamics of Eurasian populations during the last 3000 years, confirmed and further resolved the genetic relationship between recently sequenced ancient human remains and modern populations, and highlighted significantly higher amounts of Neanderthal gene flows in Island South East Asian and Oceanian populations. We have also assembled an extensive catalogue of genes under positive selection in various human groups.

Finally, ChromoPainter (Lawson et al. 2012) and MSMC (Schiffels and Durbin 2014) have cemented genetic evidence of an early African origin for the people currently inhabiting Papua New Guinea. Our results are compatible with a first migration out of Africa of these Oceanian populations, which subsequently experienced 80% of gene flow from populations coming from the second, main Eurasian out of Africa.

16.10

Genomic signatures and mechanisms of introgression between domestic and wild animals: the case of a gene important for immunity

Christine Grossen [1], Heidi Lischer[1], Daniel Croll[2], Lukas Keller[1]
[1] University of Zürich, Zürich, Switzerland, [2] ETHZ, Zürich, Switzerland

Gene flow has wide-ranging effects on the evolution of populations and species. Gene flow substantially impacts human welfare, e.g. by creating highly transmissible viruses. High rates of gene flow into endangered species are also a major concern for conservation efforts and can increase the extinction risk. Gene flow can be a creative force driving rapid adaptation by introducing highly adapted, complex genetic variants into new genomic backgrounds. In contrast, gene flow can be a constraining force counteracting local adaptation by swamping local populations with maladapted genes (due to ecological and genetic incompatibilities). We found evidence for introgression from domestic goat into populations of Alpine ibex, a wild goat species. After near extinction Alpine ibex have successfully been reintroduced across their historic species range. We showed that critical genetic diversity at the major histocompatibility complex (MHC) was recovered through introgression from domestic goats. The region surrounding the introgressed haplotype showed strong linkage disequilibria, strong sequence clustering and low diversity among haplotypes carrying the goat-type allele. We are using a combination of population based RAD-seq and whole-genome re-sequencing of specific individuals of interest, including long-read PacBio sequencing to investigate genomic signatures and mechanisms of introgression from domestic goat into Alpine ibex. Introgression at the MHC may have contributed to the genetic rescue of Alpine ibex. However, hybridization and subsequent introgression between species likely generates genetic and ecological incompatibilities, hence Alpine ibex populations may also have suffered from negative consequences of mating with a distinct species.

29A

Early Interbreeding between Ancestors of Humans, Neandertals, and Denisovans within Africa

Gundula Povysil, Sepp Hochreiter
Institute of Bioinformatics, Johannes Kepler University, Linz, Austria

Since the release of high-coverage whole genome sequences of a Neandertal and a Denisovan, the interbreeding of ancestors of humans and other hominins outside of Africa has been intensively studied. In contrast to that, their interbreeding within Africa still lacks proper attention although hominins lived longer side by side within than outside of Africa and therefore had plentiful opportunities for admixture.

We investigate the genetic relationships of humans, Neandertals, and Denisovans by identifying very short DNA segments that these hominins share identical by descent (IBD). By focusing on rare variants, our method HapFABIA reliably identifies very short IBD segments that reveal events from a very distant past because shorter segments are presumably older than longer ones. Using the 1000 Genomes Phase 3 whole genome sequencing data, we extracted two types of very old IBD segments that are shared with Neandertals/Denisovans: (1) longer segments primarily found in Asians and Europeans that indicate introgression events outside of Africa; (2) shorter segments mainly shared by Africans that indicate interbreeding of ancestors of humans and other ancient hominins within Africa. These segments are not restricted to a single population of African descent, but are common in all African populations of the 1000 Genomes Project.

Our findings can help to shed light on selection processes that were initiated by interbreeding between hominin groups at different points in time. Furthermore, our results suggest that interbreeding with other hominins was a common feature of human evolution starting already long before ancestors of modern humans left Africa.

30B

A prediction of the hybridisation potential between Hominin species using mitochondrial DNA

Greger Larson, Richard Allen, Hannah Ryan, Anna Linderholm, Ross Barnett, Laurent Frantz, Liisa Loog, Charlotte King, Mark White, Ophelie Lebrasseur, Simon Ho
University of Oxford, Oxford, UK

Claims for hybridisation between humans and Neanderthals began surfacing not long after the discovery of our closest hominin ancestor. The lack of clear-cut characteristics however, meant that archaeological claims for the existence of human x Neanderthal hybrids failed to gain universal acceptance. The unique mitochondrial DNA sequences typed in fossil Neanderthal remains led several studies to conclude that Neanderthals and humans never produced fertile offspring. The generation of nuclear sequences from Neanderthals, however, clearly demonstrated a history of gene flow between the two species, and subsequent ancient hominin genomes have revealed a complex pattern of admixture across numerous human lineages (including Denisovans).

To our knowledge, no study has yet addressed whether or not these instances of admixture should have been expected given the genetic distance between species pairs. In order to test how unusual the ability of humans to hybridise with closely related ancestors is within the context of known mammalian hybrids, we aligned mitochondrial sequences from multiple individuals of a wide variety of wild species pairs. These pairs represented both instances of fertile offspring and those pairs that produced offspring, but with limited fertility. By comparing several genetic distance measures, our results demonstrate not only that all Neanderthal and humans could have produced fertile offspring, but also that the same is likely true of all hominin lineages that diverged after the split between humans and chimpanzees. Lastly, our results suggest that though perhaps not impossible, a mating between a human and a chimpanzee would be unlikely to produce offspring.

31C

A maximum likelihood implementation of an isolation-with-migration model for three species

Daniel Dalquen [1], Tianqi Zhu[2], Ziheng Yang[1,2]

[1] University College London, London, UK, [2] Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Gene flow is an important process in the evolution of populations and the emergence of new species that affects the patterns we observe in genetic sequence data from both within and between species. The isolation-with-migration (IM) models are convenient for studying migration between populations as well as their phylogenetic structure. They incorporate the population or species phylogeny in a model of migration. However, the need to average over the entire genealogical history at every locus, including a theoretically unlimited number of migration events, poses a computational challenge for likelihood-based implementations of these models when analysing large datasets.

Here we present an extension to our previous maximum likelihood implementation of the symmetrical IM model for three species (SIM3s) that accepts arbitrary data configurations for loci with two or three sequences (loci being loosely linked genomic segments such that recombination within each locus is negligible). By limiting the number of sequences per locus to three, the genealogical process of coalescence and migration can be described by continuous-time Markov chains. This characterisation integrates over the migration histories analytically, enabling the analysis of genome-scale sequence data with tens of thousands of loci.

We show in a simulation study that adding data from a third species increases the power of the likelihood ratio test for gene flow between two in-group species, while inclusion of population data substantially improves parameter estimation. Finally, we apply the new method to a Drosophila dataset, detecting small but significant amounts of gene flow between D. melanogaster and D. simulans.

32D

Divergence history of the Carpathian and smooth newts.

Piotr Zielinski [1], Krystyna Nadachowska-Brzyska[2], Wieslaw Babik[1]

[1] Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland, [2] Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden

Information about the time of initial split, current and ancestral population sizes, and history of post-divergence gene flow between diverging species is essential for the understanding of the process of speciation. Gene genealogies are useful for testing various models of divergence and gene flow and estimate their parameters. We investigated historical demography and gene exchange between the Carpathian (*Lissotriton montandoni*) and smooth (*L. vulgaris*) newts. These sister species are morphologically, ecologically, behaviorally and genetically differentiated but hybridize in nature and introgression of some parts of the genome has been reported. Using Approximate Bayesian Computation (ABC) framework we evaluated several alternative hypotheses regarding the extent, direction and timing of genome-wide gene flow between these species. Inferences were based on 66 nuclear, non-coding markers (ca. 32kb) collected from 58 populations. The model of recent (last glacial period) interspecific gene flow, allowing for demographic change during the Pleistocene was highly supported and favored over alternative models of i) isolation, ii) constant and iii) old gene flow. The estimated divergence time of ca. 4mya suggests pre-Pleistocene species divergence, which is consistent with the fossil record. Our study indicates that despite the apparent long-term evolution in isolation the species not only retained the ability to hybridize but also that recent hybridization has led to genome-wide introgression, demonstrating that at the genomic level reproductive isolation is far from complete.

33A

Tracing Genetic History of Populations in Myanmar

Min-Sheng Peng [1,2], Quan-Kuan Shen[1,2], Shi-Fang Wu[1], Wen-Zhi Wang[1], Ya-Ping Zhang[1,3]
[1] Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China, [2] Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan, China, [3] Yunnan University, Kunming, Yunnan, China

According to the southern coastal route hypothesis, Myanmar is the entrance from South Asia to Southeast and East Asia. Nowadays, Myanmar is the homeland for various ethnic groups. All make it as a crucial region to study human evolutionary history. Herein, we genotyped ~900,000 genome-wide SNPs in 175 unrelated individuals of eight populations representing five ethnic groups: Bamar, Chin, Naga, and Rakhine from Myanmar as well as Jingpo (Kachin) from Yunnan, Southwest China. Incorporating the published data, populations of Myanmar display as basal branches to other Southeast and East Asian populations in the tree of global populations, which is in accordance with the southern route migration hypothesis. Admixture analyses detected gene flows from South Asia in Bamar and Rakhine populations from southern Myanmar rather than Chin, Naga, and Jingpo populations from northern Myanmar. Genomic runs of homozygosity and decay of linkage disequilibrium results indicate a distinctive record of the demographic history between populations from northern and southern Myanmar. Serious genetic drifts likely occurred in Chin and Naga populations. The existence of population structure is further revealed by analyses of haplotypes by using ChromoPainter and FineStructure. Interestingly, the genetic eivdence are largely in agreement with the model of Sino-Tibetan languages, suggesting certain population expansions from north to south. Thus, our analyses of genome-wide SNPs uncovered a complex history in Myanmar populations.

34B

The Spatial Mixing of Genomes in Secondary Contact Zones

Alisa Sedghifar [1], Yaniv Brandvain[2], Peter Ralph[3], Graham Coop[1]
[1] University of California, Davis, Davis, CA, USA, [2] University of Minnesota, St. Paul, St. Paul, MN, USA, [3] University of Southern California, Los Angeles, CA, USA

Populations frequently undergo periods of relative isolation that are followed by secondary contact and interbreeding. The resulting mixing of differentiated genomes generates signatures of linkage disequilibrium (LD) in admixed populations. The extent to which this LD is broken down by recombination over time can provide information about the timing of admixture. Commonly, this timing is inferred for single admixed populations, without considering the spatial distribution of neighboring populations and gene flow between them. This can mislead inference of admixture times, as expected patterns of LD can be influenced by the assumed demographic model. Here, we introduce a theoretical framework for modeling neutral patterns of LD in a geographically explicit setting, where two differentiated populations are diffusing back together in space to form a contact zone. We derive an expression for expected LD across geographic space as a function of the age of the contact zone and the dispersal distance of individuals. We have fit the expected decay of LD with recombination distance to genomic data from sets of spatially sampled human populations in Indonesia, India and Central Asia to infer dispersal distances and the timing of secondary contact. In addition to using information from the geographic distribution of LD to infer contact zone age, our framework provides a suitable null model for investigating patterns of admixture and introgression.

35C

The genetic history of early maize in the Americas

Jazmín Ramos Madrigal [1], Nathan Wales[1], Bruce D. Smith[2], Thomas P. Gilbert[1]
[1] Centre for GeoGenetics, University of Copenhagen, Copenhagen, Denmark, [2] Program in Human Ecology and Archaeobiology, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC, USA

Approximately ten thousand years ago, a wild population of teosinte (*Zea mays* spp. *parviglumis*) began being cultivated by the inhabitants of what is today central Mexico. While is not yet clear whether people domesticated the plant for its floury seeds or the sweet flavor, thanks to this event, a huge diversity of maize landraces are grown today around the world. After the domestication of maize, humans dispersed it throughout the Americas, requiring adaptation to diverse habitats, in part assisted by continuous crossing with wild teosinte conspecies. In addition, early farmers selected phenotypic and biochemical traits now present in all maize landraces.

At the moment, only limited details are known about maize's migration routes from central Mexico and the environmental factors that initially shaped the teosinte populations. Recent genetic evidence suggests maize diffused through a highland route from central Mexico to the southwestern USA.; however, additional northeastern and southern routes remain to be investigated. We generated high-throughput sequencing data from ancient maize cobs from two different archaeological sites and time periods in Mexico: 2,300-year-old specimens from northeastern Mexico, and 4,000-year-old maize cobs from dry caves close to the presumed center of domestication. Using these data, we explore the genetic characteristics of early maize domesticates, the extent and timing of the introgression with wild teosinte populations, and the possibility of additional migration routes.

36D

Migration-Selection Balance and the Unlikely Evolution of Blindness in Cavefish

Reed Cartwright, Rachel Schwartz, Alexandra Merry, Megan Howell
Arizona State University, Tempe, AZ, USA

The evolution of blindness in cavefish (and other cave dwelling species) has been a topic of research in biology since the 19th century. Several hypotheses have been put forth about the about the origin of blindness, including adaptation, mutation pressure, and random genetic drift. Using population-genetic theory, we address the origin of blindness in a cave population of fish that experiences gene flow from the surface. We demonstrate that the constant immigration of seeing alleles reduces the probability that blindness will evolve, such that the only way it will evolve is for seeing alleles to be strongly selected against in the cave population. For instance, if the immigration rate is low (0.001) and the blindness allele is rare on the surface (q=0.001), then selection would need to be uncharacteristically strong (above 10%) for blindness to evolve in the cave. Viability or fertility selection against eyes in the darkness of a cave is unlikely to produce such levels of selection. Thus a more likely explanation, which was first proposed in 1925 and mostly forgotten, is phototaxis: fish with sight preferentially leave the cave. While it would be surprising if in the darkness of a cave functioning eyes caused a 10% increase in mortality, it would not be surprising if they lead to a 10% increased chance of emigration.

37A

ORIGIN AND DYNAMICS OF ADMIXTURE AND DELETERIOUS MUTATIONS IN LATIN AMERICA (BRAZIL)

Fernanda Kehdy[1], Mateus Gouveia[1], Moara Machado[1], Andrea Horimoto[3], Bernardo Lessa-Horta[4], Mauricio Barreto[5], Maria Fernanda Lima-Costa[2], Alexandre C Pereira[3], Maira R Rodrigues[1], Wagner CS Magalhaes[1], Eduardo Tarazona-Santos [1], The Brazilian EPIGEN Consortium[6]
[1] Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, [2] Centro de Pesquisa Rene Rachou, FIOCRUZ, Belo Horizonte, MG, Brazil, [3] Instituto do Coracao, Universidade de Sao Paulo, Sao Paulo, Brazil, [4] Universidade Federal de Pelotas, Pelotas, RS, Brazil, [5] Universidade Federal da Bahia, Salvador, BA, Brazil, [6] EPIGEN-Brazil, Brazil, Brazil

South Americans are underrepresented in human genomic diversity studies. We present results of the EPIGEN-Brazil initiative, the most comprehensive project to study the genomic diversity in a Latin American population. We performed a population-based genome-wide (around 2.2 M SNPs) analysis of 6,487 individuals, to elucidate how ancestry, kinship and inbreeding interact in three populations with different history from Northeast (African ancestry: 53%), Southeast and South (both European ancestry >70%) of Brazil. We showed that ancestry-positive assortative mating permeated Brazilian history. We developed an Approximate Bayesian Computation framework to infer admixture dynamics across the last five centuries, and traced European ancestry in Southeast/South to a wider European/Middle Eastern region, inferring more recent European immigration, respect to Northeast (where ancestry seems restricted to Iberia). We broadened our understanding of the African diaspora, whose major destination was Brazil, by revealing that Brazilians display two within-Africa ancestry components: one associated with non-Bantu/Western Africans, more evident in Northeast and African-Americans; the other associated with Bantu/Eastern Africans, more present in Southeast/South. Whole-genome analysis (42X) of 30 individuals shows that continental admixture, rather than local Post-Columbian history, is the main determinant of the individual amount of deleterious genotypes, although in a more complex fashion than previously thought.

38B

Inference of Admixture Time and Proportions

Amy Ko, Rasmus Nielsen
University of California, Berkeley, Berkeley, CA, USA

Admixture between previously diverged populations leaves genetic traces whose information can help us address various questions in population genetics. For example, identification of migrant chromosomes, or admixed tracts, plays an important role in inferring the admixture time between the populations, mapping genetic loci of diseases that show differential risks between populations, and studying the role of introgression in adaptation.

Here we present an HMM-based method to identify admixed tracts in a given genome, which in turn allows us to estimate the admixture time and proportion. We apply our method to estimate the admixture time between the Neanderthals and the Kostenki 14 individual, one of the oldest Anatomically Modern Humans from European Russia dating back at least 36,200 years. We estimate the admixture time to be approximately 54,000 years before present, consistent with previous studies

39C

Patterns of Denisovan and Neanderthal ancestry in 35 Papuan individuals

Benjamin Vernot [1], Kay Prufer[2], Janet Kelso[2], Svante Pääbo[2], Joshua Akey[1]
[1] University of Washington, Seattle, WA, USA, [2] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

As anatomically modern humans dispersed out of Africa, they encountered Neanderthals in Eurasia, and low levels of hybridization occurred. The ancestors of some Oceanic and Australian populations additionally hybridized with a second archaic species, Denisovans. Recently, we developed an approach to identify surviving Neanderthal lineages in contemporary individuals, and recovered over 600 Mb of the Neanderthal genome present in modern Europeans and East Asians. Here, we have sequenced 35 individuals from Papua New Guinea, and extended our methods to identify both Neanderthal and Denisovan ancestry in the same individual. We present a detailed characterization of Neanderthal and Denisovan archaic ancestry in these individuals, as well as in new 1000 Genomes populations. The map of surviving archaic sequences shows marked heterogeneity across the genome, and we identify many "deserts of archaic sequence" that are significantly depleted for Denisovan and Neanderthal sequence in all populations, suggesting purifying selection against archaic sequence in these regions. These genomic regions are of particular interest because they delimit sequences that may confer uniquely human characteristics. Additionally, we identify several genes that appear to have been adaptively introgressed from Denisovans into Papuans. We also show that archaic sequences likely have been acquired multiple times in the histories of these populations, suggesting that the history of admixture is likely more complex than previously thought. Collectively, these results provide new insights into how hybridization shaped the evolution of modern humans.

40D

Introgressive hybridization and gene flow among Elymus species(Triticeae: Poaceae)

Genlou Sun [1], Hongwei Zuo[2], Dexiang Wu[2]
[1] *Saint Mary's University, Halifax, NS, Canada,* [2] *Anhui Agricultural University, Hefei, China*

Introgression through hybridization is a creative evolutionary process, which can produce and accumulate genetic novelties faster than through mutation alone, and plays critical roles in driving speciation. Recurrent hybridization and introgression resulting in gene flow between species could overturn the genetic differentiation among populations required for speciation and even breakdown the previously established boundaries. A better understanding of the frequency of introgression and the dynamic evolutionary consequences of hybridization is, therefore, of widespread evolutionary interest. Comparison of population structure between sympatric and allopatric populations can reveal specific introgression and determine if introgression occurs in a unidirectional or bidirectional manner. Simple sequence repeat markers were used to characterize sympatric and allopatric population structure of plant species, E. alaskanus, E. caninus, E. fibrosus, and E. mutabilis. Comparison of gene flow between species that occurs within the same geographic locations vs. gene flow between populations within species might provide evidence of introgression. Our results indicated that gene flow between species that occur within the same geographic location is higher than that between populations within species, suggesting that gene flow resulting from introgressive hybridization might have occurred among the sympatric populations of these species, and may play an important role in partitioning of genetic diversity among and within populations. Migration rate from E. fibrosus to E. mutabilis is highest (0.2631) among the four species studied. Asymmetrical rates of gene flow among four species were also observed. The findings highlight the complex evolution of these four Elymus species.

41A

Impact of enrichment conditions on cross-species capture of fresh and degraded DNA

Johanna Paijmans[1], Joerns Fickel[1,2], Alexandre Courtiol[2], Michael Hofreiter[1], Daniel Foerster [2]
[1] Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [2] Leibniz-Institute for Zoo and Wildlife Research, Berlin, Germany

By combining high-throughput sequencing with target-enrichment, investigators are able to obtain molecular data of genomic regions of interest for research that is otherwise constrained by sample quality (e.g. degraded and contamination-rich samples) or a lack of *a priori* sequence information (e.g. studies on non-model species). Despite the use of hybridization capture in various fields of research for many years, the impact of enrichment conditions on capture success have been poorly investigated. We evaluated the impact of a key parameter - hybridization temperature - on the capture success of mitochondrial genomes across the Carnivoran family Felidae. Capture was carried out for a range of samples types (fresh, archival, ancient) and levels of sequence divergence between bait and target (i.e. across a range of species). Capture performance was assessed by comparing post-capture results to shotgun data from the same samples. Results are discussed in the context of parameters of interest for investigators planning future studies employing capture techniques.

42B

Assessing the impact of population demography on the ability to replicate association signals across populationsAssessing the impact of population demography on the ability to replicate association signals across populations

Michelle Kim, Michael DeGiorgio
Pennsylvania State University, University Park, PA, USA

In recent years, genome-wide association studies have been used to identify genetic variants associated with complex traits. In humans, these studies have often been in the context of complex diseases, with individuals sampled from a limited set of geographic regions. However, when performing such an association for complex traits, it is important to know the likelihood that an identified variant would replicate in a study from another population. To investigate the influence of demographic parameters on the ability to replicate association signals in diverse populations, we employed coalescent simulations to model genetic variation from a pair of populations under a divergence model. A number of demographic parameters may influence the ability to replicate signals. We considered parameters for divergence time, bottleneck strength, population growth, gene flow, and ancient admixture. We initially explored the impact of divergence time and bottlenecks on the ability to replicate signals. As expected, increasing divergence time decreased sharing between the populations. Increasing bottleneck strength resulted in a greater proportion of significant signals found in the bottleneck population that were also found in the population without the bottleneck. In contrast, a smaller proportion of significant signals were replicated in the population that experienced the bottleneck. Our study highlights how demographic features affect the sharing of significant hits from association studies, and can be used to determine a strategy for sampling populations that maximizes the replicability of trait associations.

43C

Recent genetic history of Denmark

Georgios Athanasiadis [1], Jade Cheng[1], Blas Benito[2], Bjarni Johann Vilhjálmsson[1], Peter Kjærgaard[1], Frank Grønlund Jørgensen[1], Mikkel Heide Schierup[1], Thomas Mailund[1]
[1] Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark, [2] Department of Bioscience, Aarhus University, Aarhus, Denmark

In this work we explore the extent to which recent and more distant historical events left their mark on the genetic structure of the current Danish population. Our ultimate goal is to gain insights into the historical complexity of Denmark from the genetic perspective.
For this purpose, we ran an extensive genetic analysis on the "Where Are You From?" data set, which consisted of approximately 600 high school students from across Denmark. Each student provided a saliva sample for DNA analysis and completed an anonymised online questionnaire on geographic and biometrical data. All participants gave their informed consent and the Ethical Committee of the University of Aarhus approved the study.
Genotyping was carried out with the 23andMe DNA analysis service based on a custom HumanOmniExpress-24 BeadChip from Illumina®. More than 500,000 SNPs were available for the analysis. Genetic structure was interrogated through PCA, ancestral component analysis (ADMIXTURE) and ChromoPainter. We compared PCA performance with an IBD-based procedure that infers geographic origin of a given sample as the weighted mean of the geographic coordinates of its closest relatives. For a more general context, publicly available data from adjacent populations were included in the analysis. Finally, we studied the effects of urbanization on the genetic structure of the sample through the geographical distribution of total genomic length of ROH.

44D

Persistent molecular associations in eukaryotes by fusion of genes from ancient prokaryote operons

Raphaël Méheust
Université Pierre et Marie Curie, Paris, France

Eukaryogenesis (i.e the origin of eukaryotic cells) has been accompanied with major changes. Size population decreased, transcription/translation have been decoupled because of nucleus innovation, intron invaded genes… All these changes deeply impacted genome architecture. Notably, assuming that eukaryotes arose from the merging of one archaea and at least one eubacteria (i.e the ancestor of mitochondria), one could expect operons, a group of co-transcribed genes, a central structure in prokaryotic genomes, could be present in eukaryotic genomes, at least, as part of the DNA transferred from prokaryote ancestors. However, operons of prokaryotic origin are absents in eukaryotes. In this study, we found that some eukaryotic specific genes are fusions of genes colocalized in operon in extant prokaryotic species. Some of these fusions are widely distributed in eukaryotic supergroups and thus seem ancient, some others are lineage specific. Interestingly, a few number of these fusions have frozen, in the sense that the genes that composed the fusion are never found alone in the genome, suggesting strong functional constraints maintaining these genes fused.

45A

Crop-to-wild gene flow influenced by human activities: insights from the cultivated apple and its wild European relative

Alice Feurtey [1], Amandine Cornille[1], Maud Tenaillon[2], Pierre Gladieux[1], Antoine Branca[1], Tatiana Giraud[1]

[1] Laboratoire Écologie Systématique et Évolution UMR8079, Orsay, France, [2] UMR Génétique Quantitative et Évolution - Le Moulon, Gif-sur-Yvette, France

Gene flow is an essential component of population adaptation and species evolution. Understanding of the natural and anthropogenic factors affecting gene flow is also critical for the development of appropriate management, breeding, and conservation programs. Here, we explored the natural and anthropogenic factors impacting crop-to-wild and within wild gene flow in apples in Europe using an unprecedented dense sampling of 1889 wild apple (Malus sylvestris) from European forests and 339 apple cultivars (Malus domestica). We made use of genetic, environmental, and ecological data (microsatellite markers, apple production across landscapes and records of apple flower visitors, respectively). We provide the first evidence that both human activities, through apple production, and human disturbance, through modifications of apple flower visitor diversity, have had a significant impact on crop-to-wild interspecific introgression rates. Our analysis also revealed the impact of previous natural climate change on historical gene flow in the nonintrogressed wild apple M. sylvestris, by identifying five distinct genetic groups in Europe and a north-south gradient of genetic diversity. These findings identify human activities and climate as key drivers of gene flow in a wild temperate fruit tree and provide a practical basis for conservation, agroforestry, and breeding programs for apples in Europe.

46B

Early modern human dispersal from Africa: genomic evidence for multiple waves of migration

Silvia Ghirotto [1], Francesca Tassi[1], Massimo Mezzavilla[2], Sibelle Torres Vilaca[3], Lisa De Santi[1], Guido Barbujani[1]

[1] Department of Life Sciences and Biotechnologies, University of Ferrara, Ferrara, Italy, [2] Institute for Maternal and Child Health- IRCCS "BurloGarofolo"- Trieste, Trieste, Italy, [3] Leibniz Institute for Zoo and Wildlife Research, Berlin, Berlin, Germany

It is unclear whether early modern humans left Africa through a single, major process, dispersing simultaneously over Asia and Europe, or in two main waves, first through the Arab peninsula into Southern Asia and Oceania, and later through a Northern route crossing the Levant. Here we show that accurate estimates of the divergence times between European and non-African populations are more recent than those between Australo-Melanesia and Africa, and incompatible with the effects of a single dispersal. This difference cannot possibly be accounted for by the effects of hybridization with archaic human forms in Australo-Melanesia. Furthermore, in several populations of Asia we found evidence for relatively recent genetic admixture events, which could have obscured the signatures of the earliest processes. We conclude that the hypothesis of a single major human dispersal from Africa appears hardly compatible with the observed historical and geographical patterns of genome diversity.

47C

Genetic heterogeneity of Berber peoples as a result of differential migration and admixture patterns

Lara R.Araúna [1], Javier Mendoza-Revilla[1,2], Hassan Izaabel[3], Asmahan Bekada[4], Soraya Benhamamouch[4], Karima Fadhlaoui-Zid[5], Pierre Zalloua[6], Garret Hellenthal[2], David Comas[1]
[1] Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain, [2] Genetics Institute, University College London, London, UK, [3] Laboratoire de Biologie Cellulaire & Génétique Moléculaire (LBCGM), Université IBNZOHR, Agadir, Morocco, [4] Department of Biotechnology, Faculty of Sciences, University of Oran, Oran, Algeria, [5] University El Manar, Tunis, Tunisia, [6] The Lebanese American University, Chouran, Beirut, Lebanon

The genome-wide structure of North African populations has been described as an amalgam of autochthonous, Middle Eastern, European and Sub-Saharan ancestries. The indigenous North African genetic ancestry has been estimated to be derived from a "back to Africa" migration dated in pre-Holocean times, and suggested to be in highest frequency in Berber populations, which are considered the autochthonous populations of North Africa.

The lack of genetic knowledge on Berber populations in the current datasets prevents the understanding of the complex population processes occurred in North Africa. We have genotyped ~900,000 genome-wide SNPs in four different Berber groups and compared the results with other North African and neighboring groups. Our research reflects the complex structure of North African populations due to multiple admixture processes. Berber groups are very heterogeneous, showing different patterns of admixture and demographic histories, such as relevant endogamy levels within groups. The analysis of haplotype structure and the ancestral relationship between individuals has allowed us to estimate historical dates for some migrations coming from the Middle East and Sub-Saharan Africa into North Africa, affecting in an heterogeneous way the populations of the region. In addition, the comparison of autosomal and sexual chromosomes has shown a sexual bias in the admixture between North African populations and their surrounding neighbors, being the North African autochthonous component mainly derived from males and the European admixture driven by women. This study brings light to the origin of Berbers and the complex admixture processes in the population of North Africa.

48D

Peculiar pigment patterns and possible progenitors of a poisonous pufferfish

Seita Miyazawa
Osaka University, Osaka, Japan

Complex and camouflaged animal markings can be formed by the 'blending' of simple color patterns. A mathematical model predicts that crossing between animals having inverted spot patterns (for example, 'light spots on a dark background' and 'dark spots on a light background') will necessarily result in hybrid offspring that have peculiar labyrinthine patterns as 'blended' intermediate phenotypes. We are exploring the possibility that at least a part of animal species with labyrinthine patterns may have evolved through hybridization and/or introgression between spotted animals. Our current progress of comparative genomic approach using pufferfish will be presented.

49A

Demographic inferences from a diverse panel of African human genomes

Gerard Serra-Vidal [1], Belen Lorente-Galdos[1], Gabriel Sampere[1], Lukas Kuderna[1], Himla Soodyall[2], Pierre Zalloua[3], Karima Fadhlaoui-Zid[4], David Reich[5], Tomas Marques-Bonet[1,6], David Comas[1]

[1] Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciencies Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain, [2] University of the Witwatersrand and National Health Laboratory Service, Faculty of Health Sciences, Johannesburg, South Africa, [3] The Lebanese American University, Chouran, School of Medicine, Beirut, Lebanon, [4] University Tunis El Manar, Department of Genetics, Immunology and human Pathology, Tunis, Tunisia, [5] road Institute and Harvard Medical School, Department of Genetics, Boston, USA, [6] Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Catalonia, Spain

Understanding genetic variation across ethnically and geographically diverse extant African populations is of great importance for reconstructing human complex demographic history. Here, we study the recent history and relationships among 15 different African populations, by analyzing the whole-genome sequence data of 21 individuals sequenced at deep coverage covering all major continental linguistic groups, ecosystems and life-styles within Africa. We detected 12 million single nucleotide substitutions, providing a rich picture of the genome diversity and population history in Africa. We observe a remarkable correlation among genetic diversity and geographic origins and recent demographic history of the individuals studied. While different hunter-gatherer groups show more differentiation compared with the rest of samples, Bantu individuals are genetically more homogeneous and present evidence of admixture with neighboring hunter-gatherer groups, depending on the geographic area. Northern African individuals are closely related to non-African populations, in agreement with a recent split of both groups and continuous gene flow. To gain insight into the deepest split of our species, we explore if recent admixture of Pygmies and Khoesan with other populations may cover up their real diversity, becoming the human most diverse groups.

50B

The Footprint of Adaptive Introgression After Secondary Contact

Derek Setter [1,3], Joachim Hermisson[1,3], Jeffrey Jensen[2]
[1] University of Vienna, Vienna, Austria, [2] École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [3] Vienna Graduate School of Population Genetics, Vienna, Austria

Increasing evidence of gene flow despite reproductive barriers between divergent populations or emerging species indicts introgressive variation as an important but often overlooked source for adaptation. Introgressive adaptive alleles will typically be linked to other foreign alleles that are locally maladapted or incompatible within the genomic background of the recipient population. Successful introgression requires that the adaptive allele escapes from linked and strongly deleterious alleles through recombination. On the other hand, weakly deleterious alleles may hitchhike to high frequency or even fixation when tightly linked to the adaptive variant. Exploiting branching process theory, coalescent theory, and simulations, we study the expected genomic signature of positive selection that results from an adaptive introgression event. We find that due to divergence between populations and linkage to deleterious alleles, adaptive introgression generates footprints of positive selection that differ significantly from those of classical selective sweeps.

51C

Low genetic diversity of sharks: natural patterns or induced by exploitation?

Fernando Mendonça [1,3], Rui Coelho[2], Bruno De-Franco[1], Bruno Ferrette[1,3], Millke Morales[3], Juliana De-Biasi[1,3], Sâmia Camargo[1,3], Camila Carmo[3], Carolina Magalhaes[3], Claudio Oliveira[3], Miguel Santos[2], Fausto Foresti[3]

[1] Instituto do Mar, Universidade Federal de São Paulo - UNIFESP, Santos, São Paulo, Brazil, [2] Instituto Português do Mar e da Atmosfera, IPMA, Olhão, Algarve, Portugal, [3] Instituto de Biociencias de Botucatu, Universidade Estadual Paulista, UNESP, Botucatu, São Paulo, Brazil

Genetic diversity patterns are extremely relevant components in management plans on threatened species. However, such evaluations typically determine the current variation of this diversity without being able to characterize the magnitude of loss over the time. Considering the increasing number of sharks that have been included in the IUCN Red List of Threatened Species, without prior reviews about the molecular diversity could have been made, we present a comparison of these parameters, using sequences of the control region mitochondrial DNA of 1.016 sharks belonging to three endangered pelagic species (*Carcharhinus longimanus*, *Alopias superciliosus* and *Sphyrna zygaena*) and three not currently threatened (*Prionace glauca*, *Galeocerdo cuvier* and *Carcharhinus leucas*) from the Atlantic Ocean, of regions without signs of population genetic structure. As a result, among the endangered species was observed average nucleotide diversity π=0.00144 and haplotype Hd=0.415. Among the species not threatened such indexes were π=0.00306 and Hd=0753, 53% higher for nucleotide diversity and 45% higher for the haplotype diversity with respect to endangered species. Assessing the diversity loss of a population would imply the use of historical samples collected before population declines. However, this finding of different levels of genetic diversity in endangered species may mean that in sharks the genetic variability might not be naturally low as previously assumed, but may also be the effect of population declines due to human exploitation. Thus, the evolutionary potential of many species of sharks may become compromised even before detecting situations of population collapses.
Financial support: BIOTA FAPESP 2011/23787-0 and FCT SFRH/BPD/93936/2013

52D

Genetic analyses for the restoration of the Beluga sturgeon (*Huso huso*) in the Adriatic region.

Chiara Caruso [1], Anna Barmintseva[2], Dalia Onara[3], Radu Suciu[3], Nikolai Mugue[2], Leonardo Congiu[1]

[1] Università degli Studi di Padova, Padova, Italy, [2] Russian Institute for Fishery and Oceanography (VNIRO), Moscow, Russia, [3] Danube Delta National Institute, Tulcea, Romania

The Beluga sturgeon (*Huso huso*) is a locally-extinct species in the Adriatic region, mainly due to anthropic factors. In order to successfully re-introduce this species, strong genetic bases are needed to set up a proper reintroduction plan that will lead to achieve a self-sustaining population as genetically similar as possible to those previously living in the Adriatic basin. On this basis, we performed a phylogeographic investigation at a mitochondrial level (for subsequent comparison with museum samples of the Adriatic population) of 286 Beluga specimens from 5 non-extinct populations from Black Sea and Caspian Sea basins. Preliminary results show that the pattern of mitochondrial genetic differentiation is not related to hydrological-connections between different basins. The analysis of molecular variance shows that 84.89% of the variance is justified by intra-population diversity while 15.11% is among populations ($p<0.001$). Nonetheless, the value of haplotype diversity is high ($h= 0.95 \pm 0.03$) and shared haplotypes between basins (5.2%) are scarce. The present pattern of mitochondrial variability observed in non-extinct populations seems to be the result of genetic drift acting independently on recently separated populations that were probably remixed by post-glacial flooding. Further analyses, also at nuclear level using microsatellite markers, are needed to confirm this phylogeographic pattern. Finally, ancient Beluga specimens collected in Italian museums will be characterized in order to find haplotypes ascribable to at least one contemporary population.

53A

Estimation of genetic connectivity among species from marine protected areas through genome-wide next generation sequencing (NGS) approaches.

Marta Paterno [1], Elisa Boscari[1], Leonardo Congiu[1], Niccolò Forin[1], Ilaria Anna Maria Marino[1], Chiara Papetti[1], Gabriele Procaccini[2], Lorenzo Zane[1]

[1] Dipartimento di Biologia, Università di Padova, Padova, Italy, [2] Stazione Zoologica A Dohrn, Napoli, Italy

Studies of genetic isolation patterns provide a powerful source of information for the inference of connectivity among populations of marine organisms. These studies are especially relevant for the identification of network of sites to be protected: networked Marine Protected Areas (MPAs) can operate cooperatively in order to reduce the alteration of marine ecosystem caused by the anthropic impact and ensure the long-term preservation of biodiversity and habitats on various spatial scales and with a wide-range of protection levels. We obtained genetic connectivity measures among existing and candidate MPAs in the Adriatic and Black seas by evaluating the genetic structure of 2 invertebrates species: the Mediterranean mussel *Mytilus galloprovincialis* and the sea urchin *Paracentrotus lividus*. The innovative 2bRAD technology was applied to simultaneously identify and genotype hundreds of SNPs at genome-wide scale. Specifically, we used about 500 and 800 SNPs to investigate the genetic differentiation among populations of *M.galloprovincialis* and *P.lividus*, respectively. *P.lividus* (not present in the Black Sea) showed a general pattern of genetic homogeneity in the Adriatic Sea, probably the result of the high larval dispersal potential of this species. However also *M.galloprovincialis* presents planktonic-larval stages, in the Black Sea populations are well-connected while a signal of genetic differentiation has been detected in the Adriatic Sea. Our results provide connectivity data regarding protected sites and an interesting comparison between dispersal capacity of the two study species throughout the same geographical area; furthermore this is the first time that genetic variability distribution of *M.galloprovincialis* was evaluated using nuclear *loci*.

54B

Eurasiatic Wild Ass Phylogeny from the Pleistocene to the Present

E. Andrew Bennett, Thierry Grange, Eva-Maria Geigl
Institut Jacques Monod CNRS, Paris, France

Nearly all Eurasiatic Wild Ass populations are on the verge of extinction or already extinct. Thus there is an urgent need to shed light on their phylogeography and evolution in order to devise appropriate conservation strategies. However, these animals remain poorly described genetically, and their past and present global population structure has not been systematically explored. Indeed, the osteological determination of bone and teeth remains of the genus Equus has proven extremely difficult leading to a taxonomic oversplitting of ancient equid taxa, thus confusing the link between ancient, recently extinct and extant isolated populations. We present a phylogeographic analysis of more than 200 samples from 54 ancient and 16 historical/modern sites in Europe and Asia yielding nearly 150 sequences of wild animals, half of which come from ancient archaeological and paleontological specimens, providing a time depth of 100,000 years. We established the connection between several ancient lineages with those from the end of the 19th century and provide novel analyses of modern wild populations from Iran, Mongolia and Israel. Our study addresses long-standing debates in paleontology, zooarchaeology and zoology concerning the systematic and phylogenetic position of certain equid species, which previously stood on shaky ground. Our study assigned phylogenetic positions to the enigmatic *E. hydruntinus* and the *E. hemippus* and clarifies the systematics of several, presumably well established extant "species". In particular, we question the assignment as an independent species of *E. kiang* by demonstrating that continuous gene flow occurred until recently between *E. kiang* and *E. hemionus kulan*.

55C

Native North American genomic diversity and post-European contact admixture

Aida T. Miró-Herrans [1], Jennifer A. Raff[1], Ripan S. Malhi[2], Michael H. Crawford[3], M. Geoffrey Hayes[4,5], Deborah A. Bolnick[1,6]

[1] Department of Anthropology, University of Texas at Austin, Austin, TX, USA, [2] Department of Anthropology, University of Illinois, Urbana-Champaign, Champaign, IL, USA, [3] Department of Anthropology, University of Kansas, Lawrence, KS, USA, [4] Department of Anthropology, Northwestern University, Evanston, IL, USA, [5] Center for Genomic Medicine, Feinberg School of Medicine, Northwestern University, Evanston, IL, USA, [6] Population Research Center, University of Texas at Austin, Austin, TX, USA

Although many genetic studies have investigated the origins and early history of Native Americans, less attention has been given to recent history and the impact of post-European contact on Native American genomic diversity. Because the British, Spanish, Russian, and other European powers established colonies in different regions of North America, Native Americans may have experienced varying demographic histories in each region, resulting in genetic contributions from different sources. In this study, we investigated possible effects of European contact on Native North American genomic diversity. We assessed admixture patterns in three regions of North America (the southern US, the Pacific Northwest, and Alaska) by analyzing newly collected whole-genome sequences from six contemporary Native Americans. We compared these sequences to new whole-genome sequences from three Siberians and to published whole-genome sequences from populations with Native American ancestry and non-native populations. Our analyses showed varying degrees of admixture in the different geographic regions sampled. Individuals from the Pacific Northwest, Alaska, and Siberia exhibited European genetic contribution. They also showed shared Asian and Native American ancestry. This shared ancestry suggests more recent admixture events may have occurred across the Bering Strait. We performed demographic modeling to identify when these different admixture events may have occurred. Our results demonstrate the complexity of post-European contact admixture and the variability of these processes in different regions of North America. These results highlight the need for further study of the effects of European contact on Native American genetic diversity.

56D

Hybridisation and introgression in bluebells (Hyacinthoides) – a natural process in Spain

Jeannine Marquardt [1,2], Andrew Leitch[1], Richard Nichols[1], Harald Schneider[2]
[1] *Queen Mary University of London, London, UK,* [2] *Natural History Museum, London, UK*

The native British bluebell, *Hyacinthoides non-scripta*, is a spring flowering lily, which is well-known in natural, old forests across the British Isles. In the 17th century another bluebell taxon from the Iberian Peninsula was introduced as an ornamental plant. Since the beginning of the last century there have been reports of it in the wild, and it has recently been found to form fully fertile hybrids with the native bluebell. Both hybrids and the alien taxa are spreading, yet that usually occurs close to urban areas probably because of garden escapes. Several surveys and studies looking into the distribution of these taxa have suggested they may be putting the native British bluebell at risk, by outcompeting and replacing them in their natural habitats. However, our understanding of environmental drivers influencing alien invasion is confused by human impact, including on-going plantings and changing land use.
To better understand the dynamic between alien and native taxa, I study a natural hybrid zone between the British bluebell and its sister species, *Hyacinthoides hispanica*, in Spain, where there is minimal human influence. In this natural hybrid zone we learned from the hybrid's flower morphologies and additional chloroplast data that the parents contribute symmetrically to the intermediate forms, and the hybrids appear without evident hybrid deficiencies. Gene flow is mainly mediated through pollen exchange, because clonal buds and seed dispersal are only important on an imminent range.
We will use genome-wide markers to get an understanding of patterns of introgression across hundreds of loci.

17 Genomics of sex bias: Addressing questions with or without genomes

17.1

Recombination suppression in the sex chromosome system of the plant Silene latifolia

Deborah Charlesworth
University of Edinburgh, Edinburgh, UK

The special properties of sex chromosomes (genetic degeneration of Y and W chromosomes, and accumulation of repetitive sequences) are consequences of the absence of genetic recombination. Some plant systems are excellent for studying recombination suppression in detail, because closely related species sometimes exist that do not have separate sexes, or that have homologous sex chromosomes that can be used to test whether recombination suppression is still evolving. I will describe results from the plant Silene latifolia that suggest that several recombination suppression events have recently occurred, that have increased the extent of the Y chromosome's non-recombining region at the expense of the pseudo-autosomal region.

17.2

Comparative analysis of X chromosome gene regulation in nematodes

Sarah Albritton, Maxwell Kramer, Anna-Lena Kranz, Sevinc Ercan
New York University, New York, NY, USA

Gene expression from the X chromosome and autosomes differ with respect to sex-biased gene expression and dosage compensation. To understand evolution of X chromosome gene regulation, we performed mRNA-seq analysis of males (XO) and females (XX) in five nematode species, including *C. elegans*. First, in those species where X is unannotated, we assigned sequence contigs to the X chromosome by Illumina sequencing of DNA isolated from males and females. mRNA-seq analyses in young adult males and females suggest that nematode X chromosomes are enriched for genes with high female-biased expression and depleted of genes with high male-biased expression. Genes with low sex-biased expression do not show the same trend of X chromosome enrichment and depletion. Highly sex-biased genes are primarily expressed in the gonad, and low sex-biased genes are enriched for those expressed in the soma. The differential enrichment of high and low-sex biased genes on the X reflects the tissue-specific regulation of X chromosome transcription. Interestingly, in both flies and worms, evolution of gene expression on the X chromosome is faster in the sex where X chromosome dosage compensation occurs. We are now studying if dosage compensation contributes to expression divergence.

17.3

Numerous transitions of sex chromosomes in Diptera

Beatriz Vicoso [1], Doris Bachtrog[2]
[1] IST Austria, Lower Austria, Austria, [2] UC Berkeley, California, USA

In many species, including many mammals and insects, sex is determined by the presence of sex-chromosomes. Dipteran insects, such as Drosophila melanogaster, generally have XY sex chromosomes and a conserved karyotype, consisting of six chromosomal arms (5 large rods and a small dot). This XY pair of Drosophila was assumed to be ancestral and shared by most Dipteran insects, an assumption supported by the apparent homology of the mosquito and Drosophila X-chromosomes. Here, we analyze the genomes of 37 species belonging to 22 different families of Diptera, and uncover tremendous hidden diversity in sex chromosome karyotypes. The most frequently sex-linked chromosome is the small dot chromosome, presumably reflecting the ancestral karyotype of higher Diptera. However, we also identify species with undifferentiated sex chromosomes, and others where a different chromosome replaced the dot as a sex chromosome, or where up to three chromosomes became incorporated into the sex chromosomes. While surprising in itself, this diversity also allows us to test theories of sex chromosome evolution more systematically than was previously possible. Our multi-species transcriptome analysis shows that dosage compensation has evolved multiple times in flies, consistently through upregulation of the single X in males. However, X chromosomes generally show a deficiency of genes with male-biased expression, possibly reflecting sex-specific selective pressures. These species thus provide a rich resource to study sex chromosome biology in a comparative manner, and show that similar selective forces have shaped the unique evolution of sex chromosomes in diverse fly taxa.

17.4

A novel fine-scale human recombination map reveals sex differences

Chenling Xu, Graham Coop
University of California Davis, Davis, USA

Many aspects of meiosis differ between human males and females. In particular, there are striking contrasts both in the process of recombination and in broad-scale recombination rates between the two sexes. Much less is known about fine-scale sex variation in recombination rates and recombination hotspot usage. These differences could reflect sex-specific biological processes such as sex-biased germline gene expression and chromatin states, or sexually dimorphic recombination regulation mechanisms. We developed a novel Bayesian method to build fine-scale recombination maps from phased trio genotypes. Our approach simultaneously integrates over the uncertainty in the recombination rate and the exact position of the recombination events. We verified our method with simulations, and then obtained a map combining several existing datasets including the hutterite and the deCODE genotype data. Using our maps, we characterized the relationship between sex-specific recombination rate and genome features. We found that females have a much higher rate in the 5'UTR. We also identified recombination hotspots and found, somewhat surprisingly, that many hotspots are shared between males and females. In addition to the above analysis, this fine-scale sex-specific recombination map will also be a useful resource for many other studies.

17.5

Multifactorial Male Determination in the House Fly Primarily affects the Expression of Male-Biased Genes

Richard Meisel [1], Jeffrey Scott[2], Andrew Clark[2]
[1] *University of Houston, Houston, TX, USA,* [2] *Cornell University, Ithaca, NY, USA*

Sex determination (SD) pathways evolve fast, possibly because of sex-specific selection pressures acting on alleles linked to SD genes. The housefly has a multifactorial SD system, making it an ideal model to test this hypothesis. The ancestral housefly male-determining factor, M, is located on the Y chromosome, but M has been found on autosomes and the X chromosome. M on the third chromosome (III^M) has reached high frequencies in multiple populations throughout the world and forms stable latitudinal clines. If the invasion of III^M was driven by sex-specific selection pressures, we expect male phenotypes to be affected by alleles on the III^M chromosome. To test this hypothesis, we used mRNA-seq to determine if the expression of male-biased genes differs between III^M and canonical Y^M males. The gene expression differences we identified between Y^M and III^M males are the result of both *cis* and *trans* effects. More genes are differentially expressed between Y^M and III^M males in testis than head, and genes with male-biased expression are most likely to be differentially expressed between Y^M and III^M males. We additionally found that III^M males have a "masculinization" expression profile, suggesting that the III^M chromosome has accumulated an excess of male-beneficial alleles. Finally, we identified genes that are differentially expressed between Y^M and III^M males that are suggestive of possible male phenotypes that are affected by the polymorphic sex chromosomes. These results support the hypothesis that sex-specific selection on male phenotypes drives evolution turnover in SD pathways.

17.6

Evolution and turnover of vertebrate sex chromosomes

Jennifer A. Marshall Graves [1,2]

[1] LaTrobe University, Melbourne, Victoria, Australia, [2] University of Canberra, Canberra, aCT, Australia, [3] Australian National University, Canberra, aCT, Australia, [4] University of Melbourne, Melbourne, Victoria, Australia

Genome sequencing is now providing molecular characterization of sex chromosomes in many vertebrates, revealing details of XY and ZW differentiation. Therian mammals, including humans, have a conserved X chromosome, and a degenerate Y defined by the male-dominant *SRY* gene that evolved from the X-borne *SOX3*. Birds have a highly conserved Z chromosome common to both sexes and a female-specific W that has degenerated to different extents in different lineages. Sex appears to be determined by differential dosage of a Z-borne gene *DMRT1* in ZZ males and ZW females. Remarkably, monotreme memmals have a sex chromosome complex with homology, not the mammal XY, but to the bird Z. Reptiles and fish show a huge variety of sex determining mechanisms. Some are genetic, including XY and ZW systems, some highly differentiated like the human XY and chicken ZW, and some cytologically homomorphic. Many reptiles and some fish have no sex chromosomes, but determine sex by environmental cues such as temperature of incubation. Some species, such as the Australian dragon lizard, do both.

Sex chromosome turnover has been rare in mammals, but three changes can be identified that coincide with major mammal divergences, and there are lineages of rodents with recently evolved systems. Turnover may be quite common in some reptile and fish lineages, involving physical rearrangement of sex chromosomes with autosomes, or acquisition of a novel sex determining gene that defines new sex chromosomes. This can occur via environmental sex determination, which can be interchange surprisingly easily with genetic sex determination.

17.7

Genetic basis of genital evolution between Drosophila species

Kentaro Tanaka[1], Corinna Hopfen[2], Matthew Herbert[1], Christian Schlötterer[5], David Stern[3], John Masly[4], Alistair McGregor[1], Maria Daniela Santos Nunes [1]

[1] Oxford Brookes University, Oxford, UK, [2] Max Planck Institute for Biology of Ageing, Cologne, Germany, [3] Janelia Research Campus, Ashburn, USA, [4] University of Oklahoma, Norman, USA, [5] Vetmeduni, Vienna, Austria

In *Drosophila* male genitalia display faster divergence between closely related species than other morphological traits. Phylogenetic and experimental selection studies indicate that a large component of this divergence is driven by sexual selection, however the precise mechanisms leading to differential male reproductive fitness are still a matter of dispute and probably differ among species comparisons. In the *Drosophila simulans* complex, the morphology of several genital and anal structures differs significantly among the three species of the group and they have also been shown to exhibit post-mating prezygotic reproductive isolation. Using QTL and introgression mapping we detected several loci on chromosome 3L and 3R that underlie differences in clasper, posterior lobe and anal plate morphology between *D. mauritiana* and *D. simulans*. Most of these loci affect the trait in the same direction and act additively. However, we also found evidence for epistasis, in particular between the 3rd and the X chromosome. We conducted an RNAi screen in *D. melanogaster* against positional candidate genes located in four regions mapped to high resolution on chromosome 3L and that are differentially expressed either between males and females or between *D. mauritiana* and *D. sechellia.* We found that only seven of these genes have a regulatory role in the development of genitalia. Ongoing genome-editing experiments in *D. mauritiana* and *D. simulans* using the CRISPR-Cas9 system will allow us to identify the evolved genes and possibly the nucleotide changes responsible for the genitalia differences between these two species.

17.8

Variable autosomal and X divergence estimates near and far from genes in great apes

Pooja Narang [1], Melissa Wilson Sayres[1,2]

[1] School of Life Sciences, Arizona State University, Tempe, AZ, USA, [2] Biodesign Institute, Arizona State University, Tempe, AZ, USA

Rates of mutation are higher in males than in females across mammals, a phenomenon called male mutation bias. Changes in life history correlate with changes in male mutation bias. Specifically, generation time is a significant positive predictor of the magnitude of male mutation bias across species. However, variation on short evolutionary time scales is unknown. An added complication is that male mutation bias measured among closely related species will be biased by ancestral polymorphism. With the availability of whole genome sequences for a wide variety of species, and for many individuals within a species, we can study how recent changes in life history correlate with changes in male mutation bias. In the present work, we computed male mutation bias across the Great Apes using the genomic data belonging to 5 species, 7 subspecies and 79 individuals (Prado-Martinez etal, 2013). We looked at the X-chromosome and autosomal (X/A) divergence across the subspecies as a function of distance from genes and computed the estimates of male mutation bias for all the subspecies. Intriguingly, although diversity is known to increase with distance from genes, we observed that rate of X/A divergence varied with distance from genes. This trend is observed in all the subspecies. The estimates of male mutation bias close to the genes were twice as high as estimates farthest from genes, suggesting a role of selection in shaping patterns of divergence.

17.9

Modeling population size changes leads to accurate inference of sex-biased demographic events

Shaila Musharoff [1], Suyash Shringarpure[1], Carlos D. Bustamante[1], Sohini Ramachandran[2]
[1] Stanford University, Stanford, CA, USA, [2] Brown University, Providence, Rhode Island, USA

Sex-biased demographic events ("sex-bias"), e.g. male-biased migration or female-biased mating, can be inferred by comparing X chromosomal to autosomal genetic variation. The conventional sex-bias estimator Q assumes constant population size. However, due to effective size differences, X chromosomes and autosomes recover genetic variation at different rates after size changes, so Q-based inference on a population of non-constant size biases conclusions about sex-bias and the female fraction of the effective population size (pF).

Our novel sex-bias framework models population size changes and estimates pF: it estimates demographic parameters from autosomal data and tests sex-bias models on X chromosome data. It better detects sex-bias than Q for data simulated with human-relevant demography. For recent growth, our test has more power (15% for mild bias, 40% for extreme bias) and accurately estimates pF whereas Q does not.

Our methodology applied to human genome-wide TGP data implies male bias in Africans and Europeans. The Q-based estimator which assumes constant size applied to an African population (YRI) gives an estimate of female proportion (pF) 0.359; our method using an old growth model gives pF of 0.465. For a European population (CEU), the Q-based estimator gives an unrealistic pF estimate of 0.080; our method using a complex demographic model gives pF of 0.435. The excess male bias in CEU could be due to male-biased migration out of Africa. Modeling size changes is essential to estimating sex-bias, and our novel approach can clarify pervasive signatures of sex-bias in sexual species and provide null models for selection scans.

17.10

Sex-biased expression and evolutionary rates in brown algae (Phaeophyceae; Heterokontophyta): the importance of genes uniquely expressed in males

Gareth Pearson, Baoning Liu, José Coimbra, António dos Anjos, Ester Serrão
Centre de Ciências do Mar (CCMAR), Universidade do Algarve, Faro, Portugal

In complex multicellular organisms, males and females share a common genetic background, with the exception of relatively small genomic regions in sex chromosomes. This suggests that variations in gene expression largely underlie sexual dimorphism. We investigated sex-biased expression in brown algae, an independent multicellular lineage with low levels of sexual dimorphism, by comparative transcriptome analysis of reproductive tissues (illumina RNA-seq) and patterns of adaptive evolution within the family Fucaceae (fucoid algae). Comparing transcriptomes from male and female reproductive tissue against the vegetative background in *Fucus vesiculosus*, many more male-biased (MBG) than female-biased genes (FBG) were identified; 1,127 and 174 transcripts, respectively. Orthologues from 6 species within the family were identified by reciprocal BLASTn of reference transcriptomes, and expression analysis showed that MBG were more consistently sex-biased across species than FBG. Among MBG in *F. vesiculosus*, over 60% were uniquely-expressed in male reproductive tissue, and these male-unique genes (MUG) showed some distinctive evolutionary patterns. Codon-bias was reduced in MUG, and to a lesser extent the remaining MBG, compared with female-, or non-biased genes (reduced GC3 content, higher effective number of codons), suggesting reduced selective pressure for translational efficiency and/or accuracy. Analysis of 439 gene alignments with codon-based branch-site models of adaptive evolution showed that the number of sites under positive selection was greatest in the MUG group, followed by MBG, FBG and non-biased genes. Together, our results suggest that tissue-specific expression is linked to reduced selective constraint and/or pleiotropy, and may drive accelerated rates of adaptive divergence.

362A

The effect of demography and mating behavior in the X chromosome of bonobos

Aida Andrés, Cesare de Filippo, Genís Parra, Juan Ramón Meneu, Romain Laurent, Gottfried Hohmann, Martin Surbeck, Linda Vigilant, Svante Pääbo, Sergi Castellano
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Bonobos and chimpanzees are our closest living relatives. The two species are genetically and phenotypically similar, but their societies differ in important ways. Compared with chimpanzees, bonobos live in peaceful and egalitarian societies. In bonobo groups, for example, male competition is moderate and reproduction is loosely determined by dominance rank. The genetic consequences of such differences are unknown. We sequenced to high coverage the exome of 20 chimpanzees and 20 bonobos from sanctuaries in Africa, as well as 20 Yoruba humans. We then compared the patterns of genetic variation of these populations focusing on the X chromosome, which is particularly sensitive to mating and sex-biased behaviors. We show that the X chromosome accumulates, compared with the autosomes, a higher proportion of non-synonymous and putatively deleterious alleles in bonobos than in chimpanzees (and humans), likely due to weaker natural selection. This is best explained by the mating behavior of bonobos making them particularly sensitive to the recent reduction in population size that the species has experienced. The interplay between demography and social patterns have, thus, profound consequences in the accumulation of functional variation in the X chromosomes of natural populations of this endangered species.

363B

Estrogen regulation of microcephaly genes and evolution of brain sexual dimorphism in primates

Bing Su [1,2], Lei Shi[1,2], Qiang Lin[1]

[1] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, [2] Yunnan Key Laboratory of Primate Biomedical Research, Kunming, China

Sexual dimorphism in brain size is common among primates, including humans, apes and some Old World monkeys. In these species, the brain size of males is generally larger than that of females. Curiously, this dimorphism has persisted over the course of primate evolution and human origin, but there is no explanation for the underlying genetic controls that have maintained this disparity in brain size. We tested the effect of the female hormone (estradiol) on seven genes known to be related to brain size in both humans and nonhuman primates, and we identified half estrogen responsive elements (half EREs) in the promoter regions of four genes (MCPH1, ASPM, CDK5RAP2 and WDR62). Likewise, at sequence level, it appears that these half EREs are generally conserved across primates. Later testing via a reporter gene assay and cell-based endogenous expression measurement revealed that estradiol could significantly suppress the expression of the four affected genes involved in brain size. More intriguingly, when the half EREs were deleted from the promoters, the suppression effect disappeared, suggesting that the half EREs mediate the regulation of estradiol on the brain size genes. We next replicated these experiments using promoter sequences from chimpanzees and rhesus macaques, and observed a similar suppressive effect of estradiol on gene expression, suggesting that this mechanism is conserved among primate species that exhibit brain size dimorphism. These results suggest that brain size dimorphism among certain primates, including humans, is likely regulated by estrogen through its sex-dependent suppression of brain size genes during development.

364C

Comparative analysis of the gene sets of mammalian sex chromosomes.

Toby Hunt [1], Matt Hardy[1], Jane Loveland[1], Javier Herrero[2], Chris Tyler-Smith[1], Jen Harrow[1]
[1] Wellcome Trust Sanger Institute, Cambridge, UK, [2] UCL Cancer Institute, London, UK

The sex chromosomes hold a unique place within the mammalian genome, determining the sex of the organism and holding the key to all sex-linked diseases. Moreover for domesticated animals, such as pig, they offer the possibility of improving reproductive traits and meat yield/quality via selective breeding/genetic modification.

The recent sequencing and assembly of new versions of the porcine sex chromosomes at WTSI produced good quality reference genomic sequence highly suitable for the manual annotation approach employed by the HAVANA group. We have produced full annotation for the pig X and Y chromosomes while updating our existing annotation for the equivalent human and mouse chromosomes in order to give us complete, up-to-date and consistent genesets for the sex chromosomes of these key mammalian species. Our annotation has high coverage of alternatively spliced transcripts, lncRNAs and pseudogene loci as well as detailed representation of functional potential at both gene and transcript levels. This unique data set has allowed us to conduct a comparative genomic analysis from key points across the mammalian lineage.

Here we present this analysis, providing an investigation of the patterns of X-Y/X-X/Y-Y homology within and between the species. In addition we highlight the expansion of the C/T antigens on the human X chromosome and the massive duplication of HSFY-like loci on the pig Y chromosome as interesting case studies which allow us to further understand the evolution of the mammalian sex chromosomes.

All sequence/annotation data is available through the VEGA browser (http://vega.sanger.ac.uk/) and human/mouse GENCODE genesets via FTP (http://www.gencodegenes.org).

365D

The genomic consequences of the X-linked "Sex Ratio" meiotic drive element in *D. pesudoobscura* and *D. persimilis*

Zachary Fuller [1], Randee Young[2], Nitin Phadnis[2], Steve Schaeffer[1]
[1] *Penn State, University Park, PA, USA,* [2] *Utah, Salt Lake City, UT, USA*

Meiotic drive factors are selfish genetic elements that violate Mendel's law of independent segregation by increasing their own transmission frequency at the expense of their homolog. *D. pseudoobscura* and *D. persimilis* males carrying the "Sex-Ratio" (SR) meiotic drive element fail to produce any Y-bearing sperm and exclusively produce female offspring. Three non-overlapping inversions on the right-arm of the X-chromosome are associated with the trait in *D. pseudoobscura*. The standard X-chromosome (ST) of *D. persimilis* has the same gene order as the SR X of *D. pseudoobscura*, and differs from the *D. persimilis* SR X by a single, large inversion. It is suggested an evolutionary history of meiotic drive and sex-ratio distortion may contribute to hybrid sterility, speciation and sex chromosome evolution. Because the genes underlying the SR phenotype reside within inverted regions of the chromosome where recombination is suppressed, traditional genetic approaches have failed to locate them. Here, we performed next-generation sequencing of wild-caught SR and
ST *D. pseudoobscura* and *D. persimilis* males to investigate the impacts of meiotic drive on the pattern and organization of genomic diversity and identify the molecular genetic basis of the SR phenotype. Extensive genetic differentiation and a high rate of nonsynonymous differences exist within the inverted regions of the X chromosome. We identified several candidate genes involved in processes such as spermatogenesis and chromatin condensation containing an excess of fixed nonsynonymous differences. Results from this work shed light on the population dynamics and genomic consequences of meiotic drive elements within and between species.

366A

Sex-biased demography of human populations using X chromosome-autosome comparisons

Joseph Lachance [1], Kristen Mullen[2,1], Matthew Hansen[2], Michael Chen[2], PingHsun Hsieh[3], Krishna Veeramah[4], Sarah Tishkoff[2]

[1] Georgia Institute of Technology, Atlanta, GA, USA, [2] University of Pennsylvania, Philadelphia, PA, USA, [3] University of Arizona, Tucson, AZ, USA, [4] Stony Brook University, Stony Brook, NY, USA

Because the number of X chromosomes differs for men and women, comparisons between sex-linked and autosomal genetic variation can inform our understanding of sex-biased human demography. Using 44 high-coverage whole genomes from a diverse global set of 11 human populations we quantified the strength of selective constraint on different chromosomes, found evidence of sex-biased colonization, and determined whether recent migrations involve matrilocality or patrilocality. Relative amounts of genic and intergenic diversity were similar across all studied populations (regardless of subsistence pattern or geography), and we observed that the strength of selective constraint on genes was greater for X-linked loci compared to autosomal loci - a pattern that is consistent with selection against deleterious recessive alleles. The ratio of X chromosome to autosome diversity (Q) was greater than the null expectation of 0.75 for African populations and less than 0.75 for non-African populations, with lower values of Q for populations located farther from Africa. This pattern is consistent with a male-biased serial founder effect model, and our computer simulations suggest a plausible out-of-Africa bottleneck size of 340 males and 70 females. Using the pairwise sequential Markovian coalescent, we found evidence of a large historic effective population size in West African Pygmies for autosomes. X-autosome comparisons of genetic distances between pairs of populations revealed male-biased gene flow between Pygmies and other African populations, and female-biased gene flow between Hadza and Sandawe hunter-gatherers, between Maasai pastoralists and African agriculturalists, and between Chinese and Japanese populations.

367B

The effects of male-ness on dosage compensation in the silk moth, *Bombyx mori*

Christopher Hamm, James Walters
University of Kansas, Lawrence, Kansas, USA

Heteromorphic sex chromosomes, such as those present in the Lepidoptera (where females are the heterogametic sex) are thought to evolve from autosomes that acquired a sex-determining locus. As time passes, these chromosomes acquire sexually antagonistic alleles that suppress recombination, eventually resulting in degeneration of the W chromosome. This erosion is thought to affect the stoichiometric balance of gene expression in the heterogametic sex, and evolutionary theory predicts that compensation should occur, raising X-linked expression in the heterogametic sex to counter act this. Contrary to theoretical predictions, emerging research suggests that dosage compensation in the Lepidoptera may be achieved in part by reducing Z-linked expression in the homogametic sex (males). Specifically both males and females show Z-linked expression levels averaging ~70% of autosomal expression. Recently, the sex determining locus of *Bombyx mori*, the silk moth, was discovered (Kiuchi et al. 2014, Nature, PMID:24828047). These researchers used siRNA embryos to knock down the male sex determining locus and observed increased expression in males, but not females, primarily on the X chromosome. However, expression relative to autosomes was not examined. We extend this analysis by comparing the absolute values of Z:autosome expression and find that suppression of male-ness appears to increase Z-linked in males to a level comparable to that of autosomes. This result provides further evidence for a dosage compensation mechanism in Lepidoptera that involves reduced male expression of the Z chromosome.

368C

Natural variation in a duplicated gene generates sexual antagonism in *Drosophila* pheromones

Bosco Rusuwa[1,2], Henry Chung[3], Scott L. Allen[1], Francesca Frentiu[1,4], Jocelyn Millar[5], Stephen F. Chenoweth[1]

[1] School of Biological Sciences, The University of Queensland, St Lucia, Australia, [2] Chancellor College, University of Malawi, Zomba, Malawi, [3] Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, USA, [4] Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Australia, [5] Department of Entomology, University of California, Riverside, USA

Sexually antagonistic selection is considered an important process for the evolution of sex chromosomes and sex-biased gene expression. Despite an intense empirical focus, we lack direct evidence of sexually antagonistic variants in nature. We have discovered a naturally occurring sexually antagonistic polymorphism affecting the cuticular hydrocarbons (CHCs) of *Drosophila serrata*. CHCs are dual-functioning traits in insects, with roles in both mate choice and environmental stress resistance. We used QTL mapping and NGS bulk segregant analysis to map a D. serrata CHC polymorphism to a group of fatty-acyl reductase (FAR) genes. *In situ* hybridisation of these genes in CHC-producing oenocyte cells, showed that only one of these, *DsFAR2*, is expressed in adult oenocytes. Resequencing multiple wild-derived genomes replicated the *DsFAR2* association and RNAi mediated knockdown of *DsFAR2*'s *D. melanogaster* ortholog revealed a similar phenotypic effect to that seen in D. serrata. Sequencing the *DsFAR2* locus with long-read technology revealed that the gene resides in a ~7Kbp duplication. Both CHC forms contain the duplication, but are fixed for different alleles at one copy. The common CHC form has a divergent allele, while the novel form has identical sequence at both copies. Competitive mate choice trials show that novel genotype males suffer a mating disadvantage relative to common genotype males, suggesting negative sexual selection. By contrast, the novel genotype is under positive natural selection in females, who benefit from higher desiccation and heat shock resistance. We suggest that variation in the strength of sexual antagonism across fitness components may explain why we see the coexistence of these two CHC forms in some natural populations but not others.

369D

Sex-biased gene expression analysis in the teleost fish common dentex (*Dentex dentex*) reveals the transcriptomic divergence between male and female gonads

Tereza Manousaki [1], Alexandros Tsakogiannis[1,2], Nikos Papandroulakis[1], Constantinos Mylonas[1], Costas Tsigenopoulos[1]

[1] Institute of Marine Biology, Biotechnology and Aquaculture (I.M.B.B.C.), Hellenic Centre for Marine Research (H.C.M.R.), Heraklion, Greece, [2] Department of Biology, University of Crete, Heraklion, Greece

Popularization of next generation sequencing technologies is gradually leading to an explosion of genetic knowledge in fields and species with limited prior such accomplishments. Aquaculture fishes are rapidly accumulating such efforts. We are now exploiting the advantages of the post-genomic era to unfold the biology of those valuable and biologically interesting species and ultimately improve the stocks. Sparidae fishes are not only being cultured for many decades across Mediterranean, but also represent a group of teleosts with tremendous variability in their reproductive strategies. Whilst most of them exhibit some form of hermaphroditism, a few - including our target species common dentex (*Dentex dentex*) - are actually gonochorists. We compared the expressed transcriptome of female and male gonads of adult individuals, aiming at understanding the genes that are involved in sex differentiation of that species. Apart from identifying and characterizing the global expression differences between the two sexes, we tracked down the expression profiles of target genes, known to be involved in the complex processes of sex determination and differentiation. In this study, we provide a basis for further understanding the molecular toolkits employed to make the two sexes by studying a gonochoristic species of a teleost family mostly comprised of hermaphrodites. This comparison sheds new light into the long-standing question of ‘how the two sexes differ’ and ‘which genes are responsible for that differentiation’.

370A

Mapping a novel sex determination gene in fire ants

Yu-Ching Huang[1], Nipitwattanaphon Mingkwan[2], Chih-Chi Lee[1], Laurent Keller[3], John Wang [1]
[1] Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, [2] Department of Genetics, Faculty of Science, Kasetsart University, Bangkok, Thailand, [3] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Sex determination is a fundamental biological process that is regulated by a bewildering large number of molecular mechanisms. In about 20% of insects, including all ants and bees, sex is determined by the ploidy of the individual. Males are haploid and develop from unfertilized eggs with one set of chromosomes. Females are diploid and obtain two sets of chromosomes from fertilization. In several species, the mechanism to ascertain ploidy has been proposed to be complementary sex determination (csd): heterozygous individuals at the sex locus develop into females, whereas hemizygous haploid or homozygous diploid individuals develop into males. Studies in the honeybee *Apis mellifera* have identified a single locus, *csd*, as the master sex determination gene. We are using the fire ant *Solenopsis invicta* to study the mechanisms of sex determination in ants. Our genetic analyses have revealed that fire ants have evolved a novel master sex determination gene. We have mapped the sex locus to a 131 kb region containing at least ten genes, all without similarity to known sex genes. By comparing genomes of different individuals we also identified a hypervariable region with at least 10 alleles, consistent with balancing selection acting on the sex determination locus to maintain its heterozygosity. We are currently conducting gene expression analyses and functional tests to identify and characterize this novel sex determination system.

371B

Toxic Love, Evolutionary genomics of the enigmatic Sex-Peptide

Jessy Rouhana [1,2], Tracey Chapman[1], Bregje Wertheim[2]
[1] *University of East Anglia, Norwich, UK,* [2] *university of Groningen, Groningen, The Netherlands*

Males transfer sperm and over 100 semen proteins to the female during reproduction. The effects of semen components often favour the interests of males whilst simultaneously generating costs in females, mediating a primordial sexual conflict. In the fruitfly Drosophila melanogaster, one enigmatic semen protein, Sex-Peptide (SP), causes strikingly diverse changes in female behaviour and physiology. The aim here is to investigate phenotypic and genomic variation in SP transfer among males and SP responses among females, using fully genome-sequenced lines. This will provide novel insights into the genomic signatures underlying sexual conflict.

372C

Analyzing the frequency of 300 Mendelian diseases amongst the Puerto Rican Population

Angelia Caro
University of Puerto Rico, Mayaguez, Puerto Rico

Early detection of inheritable diseases is fundamental to successful treatment and improvement of the quality of life for these different populations. The identification of sub-populations with high risk is necessary for maximizing the possibility of early intervention. The Puerto Rican population is an ethnically rich population in which prevalence is expected to vary greatly due to the vast genetic variation. Due to this, it is of extreme importance to map inheritable diseases in this population in order to improve prevention and treatment options. For this study samples will be obtained from geographically strategic municipalities so as to cover areas representative of the entire population. Replicate analysis will be conducted for each municipality using Ion Torrent sequencing technology. The characterization will be achieved via the Ion Ampliseq Inherited Disease kit which contains over 10,000 primer pairs to amplify the coding regions of over 200 inherited diseases such as early onset Alzheimer's, Familial Hypertrophic Cardiomyopathy, Parkinson's, and Wilson's disease. As a result, we expect to have a general census for each town relating the predominant genetic diseases as target sites for further study of factors that may affect prevalence. Furthermore, this study will also serve to correlate parallel studies which aim to define the ethnic diversity and the prevalence of cancer among this population.

373D

Different expression patterns of female- and male-biased genes in the avian brain

Jakob C. Mueller [1], Heiner Kuhl[2], Bernd Timmermann[2], Bart Kempenaers[1]
[1] Max Planck Institute for Ornithology, Starnberg-Seewiesen, Germany, [2] Max Planck Institute for Molecular Genetics, Berlin, Germany

The blue tit – an ecological model species since more than 60 years – has been used to investigate fitness consequences of different sexual behavioural and mating strategies. However, very little is known about general candidate genes for sex-specific behaviour. To this end we assembled the first draft genome of a single male blue tit and mapped the brain transcriptomes of five females and five males on this reference in comparison to other tissue's transcriptomes.

In the gonads we found a similar proportion of sex-limited/exclusive genes within the high number of male- and female-biased genes, respectively, whereas in the brain these proportions differed strongly. Most female-biased genes in the brain were also female-limited whereas most male-biased genes were not male-limited but expressed in both sexes. The male-biased expression can be explained by the predominantly Z-chromosomal location of its genes indicating incomplete dosage compensation for these genes, whereas the female-biased expression is independent of chromosomal location (the W chromosome was excluded from the reference). Further, the majority of female-biased genes are ncRNA genes in contrast to the mostly coding male-biased genes.

We thus hypothesize that the sexual dimorphism in the avian brain is mainly based on two distinct gene expression components: a qualitative female-exclusive expression of ncRNA genes and a quantitative male-biased expression of Z-chromosomal coding genes.

374A

The evolution of XX and XY males of *Rana temporaria* , a species with polymorphic sex determination The evolution of XX and XY males of Rana temporaria, a species with polymorphic sex determination

Melissa Toups, Paris Veltsos, Nicolas Rodrigues, Alan Brelsford, Nicolas Perrin
University of Lausanne, Lausanne, Switzerland

Depending on the population, sex-determination in *Rana temporaria* ranges from epigenetic or environmental to homomorphic sex chromosomes with specific Y haplotypes. In some populations, sex-determination is polymorphic; some males possess distinct Y chromosome haplotypes on chromosome 1 (referred to as XY males), while others do not (referred to as XX males). However, as adults, no phenotypic distinction between XX and XY males has been detected in morphology, hormone production, or reproductive success. In this study, we compare the gene expression between females and males during both metamorphosis and sexual maturity from both populations with a mix of XX/XY males, and XY males from populations with distinct Y haplotypes. We will determine the effect of the Y chromosome on sexual development. Specifically, does having Y haplotype result in a more masculine transcriptome – such as having more male-biased genes? Do Y-linked genes show any reduction in expression compared to the X-linked homologs, or do both XX and XY males have equal expression of genes on chromosome 1? Furthermore, we will determine which (if any) regions of the Y haplotype are the same for populations with a fixed Y haplotype and those that are polymorphic, which will allow us to determine if there is an ancient, nonrecombining region near the sex-determining locus, and give insight into the patterns of recombination suppression across populations.

L 4D

Assessing sex-biased demographic history in *Mus spretus* and *Mus castaneus* using population estimates of genomic diversity

Sara Keeble [1,2], Colin Callahan[2], Jessica Crisci[1], Peter Ralph[1], Jeffrey Good[2], Matthew Dean[1]
[1] Division of Biological Sciences, University of Montana, USA, [2] Molecular and Computational Biology, University of Southern California, USA

Mating system evolution can have profound impacts on genomic variation. With strict monogamy, the effective population sizes of males (N_{em}) and females (N_{ef}) are equal, and there are three copies of the X chromosome for every four copies of autosomes. However, in many species strong sexual selection results in higher variance in male reproductive success and a reduction in N_{em} relative to N_{ef}. This pattern is predicted to drive the effective number of X chromosomes towards the effective number of autosomes. Nei's π is a parameter used to estimate genetic diversity. Under equal breeding sex ratios, the ratio of π x/π A should equal 0.75. Departures from the expected ratio of 0.75 can indicate breeding sex ratio skews. Here, we compute Nei's π for putatively neutral loci in and subsequently calculate the ratio of X:autosome diversity in order to assess N_{ef} and N_{em} in wild caught *Mus spretus* and *Mus castaneus* individuals, two species of mice with different inferred intensities of male reproductive variance. 20 Mb of autosomal regions and 20 Mb of X-linked regions were used in analysis and were specifically chosen to be far from genes and in regions of high recombination to minimize the influence of selection. Consistent with theoretical predictions, X-linked variation is higher relative to autosomal variation in *M. spretus*, a species inferred to exhibit strong sexual selection based on relatively large testes. Our study highlights the effect of divergence in mating system on genomic variation.

18 Within- and between-host viral evolution

18.1

Using viral genetic data to study evolution and replication during chronic viral infection

Oliver Pybus
University of Oxford, Oxford, UK

Data generated by established and novel sequencing methods have the potential to transform our understanding of the evolutionary dynamics of chronic viral infections. Viruses such as HIV-1 and the Hepatitis C Virus (HCV) undergo continual mutation, replication, and adaptation over many years of chronic infection, and these processes leave a tangible footprint in sampled viral gene sequences. Here I introduce and discuss the analytical approaches that can be used to quantify and compare within-host virus evolution. Large scale comparative analyses of intra-host HIV-1 and HCV diversity reveals differences between the viruses in their evolutionary dynamics that are not apparent at the inter-host, or epidemiological, level.

18.2

Plant and algal genomes enclose footprints of past infections by giant virus relatives

Guillaume Blanc [2], Lucie Gallot[2], Aline Epert[1], Fabien Nogué[1], Florian Maumus[1]
[1] INRA de Versailles-Grignon, Versailles, France, [2] Laboratoire Information Génomique et Structurale, Marseille, France

Nucleocytoplasmic large DNA viruses (NCLDVs) are eukaryotic viruses with large genomes (100 kb-2.5Mb), which include giant Mimivirus, Megavirus and Pandoravirus. NCLDVs are known to infect animals, protists and phytoplankton but were never described as pathogens of land plants. Our results show that the bryophyte Physcomitrella patens and the lycophyte Selaginella moellendorffii have open reading frames (ORFs) with high phylogenetic affinities to NCLDV homologues. The P. patens genes are clustered in DNA stretches (up to 13 kb) containing up to 16 NCLDV-like ORFs. Molecular evolution analysis suggests that the NCLDV-like regions were acquired by horizontal gene transfer from distinct but closely related viruses that possibly define a new family of NCLDVs. Transcriptomics and DNA methylation data indicate that the NCLDV-like regions are transcriptionally inactive and are highly cytosine methylated through a mechanism not relying on small RNAs. Altogether, our data show that members of NCLDV have infected land plants.

18.3

Population genetics of Ebola virus in West Africa

Gytis Dudas [1], Andrew Rambaut[1,2]

[1] University of Edinburgh, Edinburgh, UK, [2] Centre for Immunology, Infection and Evolution, Edinburgh, UK, [3] Fogarty International Center, Bethesda, MD, USA

The outbreak of Ebola virus disease in West Africa has been the worst of its kind in recorded history. Due to advances in sequencing technology there are now more publicly available Ebola virus sequences from the West African outbreak than from all previous outbreaks combined. An early study using sequences sampled in Sierra Leone between May and June in 2014 observed that the evolutionary rate of the Ebola virus in humans in West Africa is much higher than the rate in the reservoir. Given the number of human-to-human transmissions in the current outbreak several hypotheses have been put forward to explain this rate contrast, such as intrinsic mutation rate differences between humans and the reservoir, relaxed purifying selection and rapid adaptation.

Using sequences sampled throughout the West African outbreak and from the main affected countries we explore and contrast the the within-host evolutionary forces with those at the level of the epidemic and on the long-term. We additionally assess the risk of altered virulence of Ebola virus, investigate the viral population dynamics during the extended circulation period in humans in West Africa and show what sequence data can teach us about the currently undetermined reservoir of the virus.

18.4

Population genomics of within host HIV evolution

Fabio Zanini[1], Jan Albert[2], Richard Neher [1]

[1] Max Planck Institute for Developmental Biology, Tuebingen, Germany, [2] Karolinska Institute, Stockholm, Germany

In untreated HIV infected individuals, the virus population is constantly adapting to selection pressures imposed by the host immune system. We have sequenced the HIV populations in 11 patients at about 10 time points each spanning 3-10 years during which HIV diverges by about 1%. The data set covers the entire HIV genome and is deep enough to accurately measure variant frequencies down to 1%. Using this time series data, we estimate the strength of purifying selection at loci conserved to different degrees on larger evolutionary time scales and quantify the corresponding within host evolutionary rates. While the exploration of sequence space by minor variants is very similar in HIV populations in different individuals, the loci of adaptation are typically specific to individual patients. The dynamics of neutral and deleterious alleles is dominated by linked selection and allele frequency spectra are inconsistent with standard neutral coalescent models. Instead, they are well described by predictions from recent theories of rapid adaptation that assume that allele frequency changes are governed by selection on changing genetic backgrounds of different fitness rather than drift. The data set provides a rich resource to study virus within host evolution and to test population genetic theories or inference methods.

18.5

A fighter after all: Accessory protein of minor HIV group under positive selection in arms race with human innate immunity factor

Christopher Monit, Richard Goldstein
Division of Infection and Immunity, University College London, London WC1E 6BT, London, UK

Human immunodeficiency viruses have been transmitted from nonhuman primates to humans on multiple occasions, but the spread of these virus groups has varied dramatically. HIV-1 group M is responsible for the AIDS pandemic, infecting 35 million people worldwide. In contrast, HIV-1 group O infects only tens of thousands of people and has largely failed to spread beyond West Africa. A suggested explanation for this difference has been that these viruses have to different degrees adapted to overcoming the innate immunity factors of their human hosts. Tetherin is a host innate immunity protein which acts to restrict virus growth by 'tethering' virus particles to infected cells. HIV-1 M uses its Vpu protein to mediate the destruction of tetherin, restoring efficient virus release. It was believed that HIV-1 O had no means of destroying human tetherin. However, recent data (Kluge et al. 2014) has shown that HIV-1 O acquired a means of antagonising human tetherin as efficiently as HIV-1 M, using its Nef protein.

We show that HIV-1 O Nef underwent positive selection in the period following transmission to humans and before its widespread dissemination. We find no evidence that HIV-1 M Nef underwent adaptive evolution in the equivalent period, supporting the hypothesis that HIV-1 O Nef acquired a new function while HIV-1 M Nef did not. Furthermore we identify specific sites in HIV-1 O Nef receiving high support for being under positive selection in this period, which will inform empirical studies into the mechanism of HIV-1 O Nef's anti-tetherin function.

18.6

The effects of a deleterious mutation load on patterns of influenza's antigenic evolution in humans

Katia Koelle, David Rasmussen
Duke University, Durham, NC, USA

Recent phylogenetic analyses indicate that RNA virus populations carry deleterious mutation loads. These mutation loads have the ability to shape patterns of adaptive evolution via genetic linkage to beneficial mutations that provide the genetic basis for adaptation. Here, we examine the effect of sublethal deleterious mutations on patterns of influenza A/H3N2's antigenic evolution. To gain intuition, we first develop and analyze simple population genetic and epidemiological models of influenza that incorporate a deleterious mutation load. These models show that the circulation of deleterious mutations should act to slow influenza's rate of antigenic evolution, while making it more punctuated in nature. These models further predict three distinct molecular pathways by which antigenic cluster transitions occur, and we find phylogenetic patterns consistent with each of these pathways in influenza viral sequences. By developing and simulating a more complex phylodynamic model that incorporates the occurrence of both antigenic and deleterious mutations in time, we show that the results obtained from the simpler models hold under more realistic scenarios. Furthermore, simulations of this more complex model indicate that transiently circulating deleterious mutations, in concert with antigenic mutations, can reproduce influenza's spindly hemagglutinin phylogeny, co-circulation of multiple minor antigenic variants, and high annual attack rates. This stands in contrast to simulations that exclusively consider the occurrence of antigenic mutations and ones that incorporate only deleterious mutations. Together, these results illustrate the importance of a deleterious mutation load on patterns on viral antigenic evolution and, more generally, on patterns of adaptive evolution in load-carrying populations.

18.7

Less efficient drugs lead to softer sweeps in intra-patient populations of HIV-1

Alison Feder [1], Soo-Yon Rhee[1], Robert Shafer[1], Dmitri Petrov[1], Pleuni Pennings[2]
[1] Stanford Universty, Stanford, CA, USA, [2] San Francisco State University, San Francisco, CA, USA

Viral populations tend to be characterized by exceptionally large population size and high mutation rates. These populations can be highly adaptable because they are not limited by the influx of mutations, and multiple positively selected variants can arise and sweep simultaneously. However, under combination antiviral therapy, a declining population size and the low rate at which fully resistant types emerge may bring populations into a mutation-limited regime, where a population must wait for the right variant which will then spread to the exclusion of others. These two regimes should leave markedly different patterns on the diversity profiles of the populations, which may allow us to test whether or not a population is evolving in a mutation-limited setting. We examine diversity among 7918 HIV-1 Direct-PCR sequences and use ambiguous reads to investigate the prevalence of hard and soft sweeps associated with resistance to antiviral drugs. Using these data, we confirm that drug resistance is associated with lower diversity levels, a common signature of selective sweeps. We further find that resistance to treatments with low success rates generate patterns consistent with soft sweeps, as marked by the relatively small decrease in diversity associated with resistance. Populations receiving high success rate treatments showed patterns more consistent with hard sweeps. This suggests that effective drugs may push HIV-1 populations into a hard sweep regime in which adaptation becomes mutation-limited. Finally, we suggest how this information can be leveraged to test new drug therapies in which information about efficacy is unavailable.

18.8

Extreme heterogeneity in genetic diversity and molecular evolution found in chronic HCV infection

Jayna Raghwani [1], Rebecca Rose[2], Isabelle Sheridan[6], Philippe Lemey[3], Marc A Suchard[4], Patrizia Farci[5], Paul Klenerman[6], Oliver G Pybus[1]

[1] Department of Zoology, University of Oxford, Oxford, UK, [2] BioInfoExperts, Thibodeaux, LA, USA, [3] Rega Institute, KU Leuven – University of Leuven, Leuven, Belgium, [4] Departments of Biomathematics, Biostatistics, Human Genetics, University of California, Los Angeles, CA, USA, [5] Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA, [6] Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK

In contrast to most fast-evolving RNA viruses, the hepatitis C virus (HCV) can cause both acute and chronic infection in humans, with viral clearance occurring in 15 to 20% of cases. However, in the majority of HCV patients that become chronically infected, the liver is expected to develop cirrhosis, cancer, and other related diseases. In recent years there has been great progress made in treatment of HCV, with the newly approved direct-acting antiviral drugs being reported to clear the virus successfully in 70% of patients. However, despite the clinical and scientific achievements, our understanding of HCV replication behaviour and thus of the within-host evolution is limited, especially compared to HIV. In this study, we undertake the first comprehensive analysis of HCV evolutionary dynamics during chronic infection by quantifying viral diversity and divergence through time. We investigate more than 4000 viral gene sequences obtained from 15 HCV patients, which have been sampled longitudinally over several years. These results are compared to those of 9 well-studied HIV subjects, which indicate key differences in these two chronically infectious RNA viruses. Notably, a significant degree of heterogeneity is observed in the molecular evolution and population genetic diversity in HCV, both among patients and over time, strongly suggesting the presence of complex replication dynamics. As a consequence, we suggest a novel mechanism by which HCV establishes chronic infection that can explain apparent paradoxes in the natural history of this virus.

18.9

Genome-wide recombination is a major driver of human cytomegalovirus evolution

Florent Lassalle [1], Daniel P Depledge[2], Matthew B Reeves[2], Amanda C Brown[3], Mette T Christiansen[2], Helena J Tutill[2], Rachel J Williams[2], Katja Einer-Jensen[4], Jolyon Holdstock[3], Claire Atkinson[5], Julianne Brown[6], Freek B van Loenen[7], Duncan A Clark[8], Paul D Griffiths[2], Georges MGM Verjans[7], Martin Schutten[7], Richard S Milne[2], Francois Balloux[1], Judith Breuer[2]
[1] UCL Genetics Institute, University College London, London, UK, [2] Division of Infection and Immunity, University College London, London, UK, [3] Oxford Gene Technology, Begbroke, UK, [4] QIAGEN-AAR, Aarhus, Denmark, [5] Department of Virology, Royal Free Hospital, London, UK, [6] Microbiology, Virology and Infection Prevention and Control, Camelia Botnar Laboratories, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK, [7] Department of Viroscience, Erasmus Medical Centre, Rotterdam, The Netherlands, [8] Department of Virology, Barts Health NHS Trust, London, UK

Human cytomegalovirus (HCMV) causes serious disease in congenitally infected infants and immunosuppressed transplant recipients. Using sequence-capture and next-generation sequencing methods, we significantly enrich for complete HCMV genomes available for uncultured and low passage clinical samples. We use population genetics and phylogenetic approaches on this dataset and show that the genomic landscape of HCMV is shaped by extensive homologous recombination. In the majority of the genome, intense recombination enforces the purge of deleterious variation and hence maintains a low polymorphism. In contrast, a small number of loci appear resistant to recombination and show unusual, sometimes extreme, levels of divergence among strains. These loci harbour genes known to be crucial for HCMV pathogenicity, such as glycoproteins B and H, but also genes whose functions remain unknown, e.g. the RL11 gene family, which forms a 10kb-long highly polymorphic and non-recombining locus. We further show that many of the hypervariable regions contain antigenic targets which are under positive selection, suggestive of strong pressure for immune escape. While any recombination between genotypes at these loci is likely to be restricted by their high sequence dissimilarity, we identify signals that support diversification at the RL11 and other loci to be due to past episodes of strong positive selection, possibly driven by immune pressure. The deep divergence at the RL11 complex between HCMV strains mirrors the pattern of accelerated between-species evolution of primate CMVs at this locus. and is suggestive of incipient speciation in HCMV.

18.10

Host-driven mutation drives genome evolution of sigma virus (DMelSV; Rhabdoviridae) in *Drosophila melanogaster*

Marta Wayne [1], Helen Piontkivska[2], Luis Matos[3], Michael Miyamoto[1]
[1] University of Florida, Gainesville, FL, USA, [2] Kent State University, Kent, Ohio, USA, [3] Eastern Washington University, Cheney, WA, USA

Sigma virus (DMelSV) is ubiquitous in natural populations of *Drosophila melanogaster.* Host-mediated, selective RNA editing of adenosines to inosines (ADAR) contributes to the control of viral infection by preventing transcripts from being transported into the cytoplasm, or by increasing the viral genomic mutation rate. However, the precise role of ADAR is unclear. Previous PCR-based studies showed ADAR occurs in DMelSV with low frequency. Here we use SOLID deep sequencing from multiple individuals within a single host population from Athens, GA, USA to comprehensively evaluate the patterns of sequence variation in sigma virus, and demonstrate that ADAR has a disproportionate contribution to between-strain differences across a worldwide sample of viruses. We also find ample mutations within hosts segregating at low frequency; and these mutations too show a large contribution by ADAR. Finally, we demonstrate that the dyads most susceptible to ADAR are underrepresented at the level of the viral genome, consistent with purifying selection on the viral genome. Thus, in DMelSV, host mutagenesis is constraining viral evolution both within and between hosts.

415A

Increased signals of positive selection in the HIV-2-specific accessory gene *vpx* in patients with symptoms of advanced disease compared to long-term viral controllers.

Katherine James [1,4], Joakim Esbjornsson[1,2], Thushan De Silva[1,3], Rosemary Roberts[1], Gilean McVean[4], Sarah Rowland-Jones[1]

[1] Nuffield Department of Medicine, University of Oxford, Oxford, UK, [2] University of Lund, Lund, Sweden, [3] University of Sheffield, Sheffield, UK, [4] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

HIV-2 is a non-pandemic retrovirus. In contrast to HIV-1, a large proportion of HIV-2 infected individuals maintain undetectable viral loads and CD4+ T cell counts in the absence of treatment, often for decades. The function of the HIV-2/SIVsmm specific accessory gene *vpx* is antagonism of the host restriction factor SAMHD1. SAMHD1 is thought to restrict retroviral replication through dNTP hydrolysis and RNase activity. The implications of SAMHD1 antagonism on HIV-2 infectivity remain unclear, however, *in vitro* assays have shown no evidence for differential SAMHD1 antagonism by alleles derived from clinically distinct patient groups.

The inter-patient diversity and evolution of *vpx* remains largely unstudied, due to the lack of sequences from primary patient samples. This project implemented a counting renaissance approach using longitudinal *vpx* sequences derived from proviral DNA to assess the selection acting on each codon of *vpx*. There was striking evidence for positive selection in viraemic patients, whereas signals of positive selection were almost entirely absent from viral controllers. Similarly, the signals of purifying selection were reduced in viral controllers when compared to symptomatic patients. Additionally, we used a novel and low bias method of HIV-2 whole genome sequencing (RNA-Seq) to show that *vpx* has lower diversity when compared to other HIV-2 genes. These results paint a previously un-described picture of dynamic selection pressures acting on *vpx*, particularly in patients with advanced disease, highlighting the importance of studying the evolution of accessory genes throughout the course of chronic HIV infection.

416B

Estimating the between-host heritability of viral traits:old-school parent-offspring versus next-gen phylogenetics

Gabriel Leventhal, Sebastian Bonhoeffer
ETH Zurich, Zurich, Switzerland

The heritability of a trait is one of the most widely used tools to quantifyhow fast a trait will evolve in a population. As a result, many differentmethods have been proposed to measure heritability in real populations.With the rise of the availability of genetic data, traditional methods suchas parent-offspring regression or sibling analysis have been supersededby phylogeny-based methods to estimate heritability. However, as with allphylodynamic models, tree-based methods for heritability estimation requirean underlying model that describes how the trait evolved along the tree.Using set-point viral load in HIV as an example, I will show that tree-basedestimates of heritability are very sensitive to model mis-specification suchas the absence of selection, both within- and between hosts. In contrast,estimates from parent-offspring (i.e. donor-recipient) regression aremore robust to such misspecification. The difficulty of obtaining goodparent-offspring pairs must thus be weighed against potentially stronglybiased estimates of heritability when the between- and within-host hostselective forces are unclear.

417C

Molecular evolution of virulence in honey bee RNA viruses

Emily Remnant, Niklas Mather, Thomas Gillard, Madeleine Beekman
Behaviour and genetics of social insects laboratory, School of Biological Sciences, The University of Sydney, Sydney, NSW 2006, Australia

Classical epidemiological theory predicts that virulence increases when pathogens are transmitted via a vector. A parasite now no longer requires its host to be alive to ensure transmission, thus allowing an increase in proliferation rate irrespective of the cost to the host.

Honeybees host a number of generally asymptomatic RNA viruses. The arrival of a mite vector, *Varroa destructor*, has been accompanied by an increase in virulence of RNA viruses. *Varroa* feeds on honeybee haemolymph, spreading viruses directly from bee to bee. This vector-induced change in transmission route appears to correlate well with virulence trade-off predictions. However in addition to changing transmission, some viruses replicate within *Varroa*. Observed changes in virulence could thus also be due to increased viral titres in the mite.

We use an experimental evolution approach to examine if, and if so how, transmission route affects evolution of RNA viruses. Using serial transmission experiments and *Varroa*-naïve honeybees we can exclude the effect of replication within the vector on the evolution of virulence. We find that while some RNA viruses increase in virulence, others do not. This indicates that some viruses are sensitive to changes in mode of transmission, leading to the selection of more virulent variants. To understand the molecular changes that underpin virulence due to changed transmission route, we are currently sequencing viral genomes to identify selection for specific polymorphisms and haplotypes. We can then compare our sequences to those obtained from honeybee populations that are sensitive or tolerant to *Varroa* and the viruses it transmits.

419A

Geometric constraints dominate the antigenic evolution of influenza H3N2 hemagglutinin

Austin Meyer, Claus Wilke
The University of Texas at Austin, Austin, TX, USA

We have carried out a comprehensive analysis of the determinants of human influenza A H3 hemagglutinin evolution, considering three distinct predictors of evolutionary variation at individual sites: solvent accessibility (as a proxy for protein fold stability and/or conservation), experimental epitope sites (as a proxy for host immune bias), and proximity to the receptor-binding region (as a proxy for protein function). We have found that these three predictors individually explain approximately 15% of the variation in site-wise dN/dS. However, the solvent accessibility and proximity predictors seem largely independent of each other, while the epitope sites are not. In combination, solvent accessibility and proximity explain 32% of the variation in dN/dS. Incorporating experimental epitope sites into the model adds only an additional 2 percentage points. We have also found that the historical H3 epitope sites, which date back to the 1980s and 1990s, show only weak overlap with the latest experimental epitope data, and we have defined a novel set of four epitope groups which are experimentally supported and cluster in 3D space. Finally, sites with dN/dS>1, i.e., the sites most likely driving seasonal immune escape, are not correctly predicted by either historical or experimental epitope sites, but only by proximity to the receptor-binding region. In summary, proximity to the receptor-binding region, rather than host immune bias, seems to be the primary determinant of H3 immune-escape evolution.

420B

Identification of endogenous retrovirus in the fixation process to cat genomes.

Sayumi Shimode [1], So Nakagawa[2], Takayumi Miyazawa[1]
[1] Kyoto University, Kyoto, Japan, [2] Tokai University, Kanagawa, Japan

Endogenous retroviruses (ERVs) are remnants of ancient retroviral infections of the host germ-line cells, and comprise approximately 10 % of the host genome in mammals. Most ERVs are inactivated through the accumulation of mutations and deletions; however, some ERVs are active and express viral proteins. The RD-114 virus is a replication-competent feline ERV, and several feline cell lines produce infectious RD-114 viral particles. All domestic cats are considered to have an RD-114 viral locus, although the locus had not been identified. In this study, we identified several RD-114 virus-related sequences (RDRSs) in domestic cats' genomes. We found that all domestic cats have an RDRS on chromosome C2 (RDRS C2a), but the other RDRSs have not been fixed in domestic cats' genomes, suggesting multiple invasions of cats' genomes by RD-114 virus-related virus. RDRS C2a has stop codons within all viral genes and phylogenetic analysis of RDRSs and RD-114 viral clones indicated that RDRS C2a belongs to the separated group from the other RDRSs and RD-114 viral clones. RDRS C2a, the oldest RDRS was not presented in the Leopard cat (*Prionailurus bengalensis*) genome. Our results indicated that RDRS C2a entered the cat's ancestor genome after the split from leopard lineage and the other new RDRS might have integrated into migrating cats in Europe. We also show that some RDRSs are useful for tracing cat's footstep.

421C

Patterns of sequence evolution in the viruses of *Drosophila* (and other insects)

Darren Obbard, Claire Webster, Fergal Waldron, Gytis Dudas
University of Edinburgh, Edinburgh, UK

Viral infections are universal, and viruses are arguably the most diverse form of life on earth. However, despite the recent rapid progress made in uncovering viral diversity, we actually still know surprisingly little about the evolution of viruses that do not infect vertebrates or plants. Motivated by the utility of *Drosophila melanogaster* as an experimental and comparative study system, we have used a metagenomic approach to identify approximately 20 novel viruses associated with *D. melanogaster* and *D. simulans*.

Many groups of insect viruses are present in *D. melanogaster* (Iflaviruses, Nodaviruses, Cypoviruses, Negeviruses, etc) but also some viruses which we believe may represent novel insect-infecting lineages (Partitiviruses, Luteoviruses). We have surveyed >2200 individual wild-collected flies to quantify the global prevalence and distribution of 16 different *D. melanogaster viruses* (including 12 newly discovered viruses), and using time-sampled data under molecular clock models we have inferred rates of intercontinental movement and host-switching between *D. melanogaster* and *D. simulans*.

We have also used these sequences to infer the patterns and strength of selection acting on *Drosophila* virus proteins. We find that viral protein sequences are generally highly conserved (dN/dS is low) and there is little evidence of strong positive selection. While this is in sharp contrast to what is seen in the Drosophila antiviral immunity genes that these viruses interact with, we argue that this is not as surprising as it may first appear.

422D

Effect of temperature on dengue virus fitness; does local adaptation matter?

Andrea Gloria-Soria [1], Philip Armstrong[2], Jeffrey Powell[1], Paul Turner[1]
[1] Yale University, New Haven, CT, USA, [2] The Connecticut Agricultural Experiment Station, New Haven, CT, USA

Dengue virus (DENV) is an important vector-borne RNA viral pathogen of humans, transmitted via bites of infected female *Aedes aegypti* mosquitoes. This vector is expanding its geographic range due to warming temperatures. However, it is uncertain whether differing DENV strains become locally adapted to extreme temperatures experienced by their *Ae. aegypti* vectors. Here, we examined the effects of temperature on DENV fitness (replication rate) in both live mosquitoes, and *in vitro* tissue culture. We obtained mosquitoes and DENV serotype-2 isolates from two thermal regimes in Vietnam, which differed by >3^{o}C in annual temperature minima and maxima, and by >7^{o}C during the colder months. To examine G x E effects, we tested whether DENV isolates from low versus high temperature regions showed superior fitness in mosquitoes from their same geographic location, but performed worse in mosquitoes from the other region. All mosquito x virus combinations were assayed at three temperatures (22, 27, and 32^{o}C) and DENV fitness was measured in *Ae. aegypti* cells as the DENV2 fitness baseline. Our results showed that temperature mainly determined viral fitness, with mosquito genetic background playing a minor role. These data suggest that DENV is poised to emerge via range expansion of its vector, regardless of the particular genetic background of the mosquito.

423A

Stochastic Population Dynamics in Bayesian Epidemic Parameter Inference with the Coalescent SIR Model

Alex Popinga [1,2], Tim Vaughan[1,3], Tanja Stadler[4], Alexei Drummond[1,2]
[1] Department of Computer Science, University of Auckland, Auckland, New Zealand, [2] Allan Wilson Centre for Molecular Ecology and Evolution, Palmerston North, New Zealand, [3] Massey University, Palmerston North, New Zealand, [4] Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

In recent years, the field of phylodynamics has been expanding rapidly with the development of new methods that use genetic sequence data to garner information about underlying epidemiological and population dynamics in infectious disease outbreaks. These methods allow for the simultaneous inference of the viral phylogeny and parameters that describe the dynamics. We extend one of these methods taken from Kingman's coalescent theory for stochastically changing populations that follow the Susceptible-Infectious-Removed (SIR) progression. We implement both the deterministic description and our stochastic extension as tree priors for Bayesian inference. We apply our models to simulated data with known parameters as well as influenza A (H1N1) sequence data sampled in the Canterbury region of New Zealand and compare our results with a recently published birth-death SIR method. We show that while all three methods are effective for data with large fundamental reproductive number R_0 and large population size S_0, the stochastic coalescent produces smaller error, bias, and has greater highest posterior density coverage. Additionally, we show that the smaller the true values are for R_0 and S_0, the stronger the advantage of using the stochastic model over its deterministic counterpart.

424B

Evolutionary consequences of delaying intervention for Ebola

Philip Johnson [1], Carl Bergstrom[2], Roland Regoes[3], Ira Longini[4], M. Elizabeth Halloran[2,5], Rustom Antia[6]

[1] University of Maryland, College Park, MD, USA, [2] University of Washington, Seattle, WA, USA, [3] Institute of Integrative Biology, ETH Zurich, Switzerland, [4] University of Florida, Gainesville, USA, [5] Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [6] Emory University, Atlanta, GA, USA

The basic reproductive number (R_0) is viewed as a good predictor both of the spread of existing pathogens and of the probability of emergence of novel zoonoses. We show this view needs to be re-evaluated when considering the global risk posed by supercritical zoonoses such as Ebola. We find that if intervention is delayed until the number of current cases reaches a threshold, the very pathogens that appear least problematic (low R_0, such as Ebola) pose the greatest threat of evolution. Further, if inter-vention only restrains an epidemic at a constant size rather then eliminates it, the probability of evolution continues to increase rapidly. These results highlight the importance of rapid control and elimination as soon as possible after emergence.

425C

The Limits and Patterns of Human Cytomegalovirus Genetic Diversity in Humans

Nicholas Renzette [1], Cornelia Pokalyuk[2,3], Jeffrey Jensen[3,4], Timothy Kowalik[1]

[1] Univerisity of Massachusetts Medical School, Worcester, MA, USA, [2] Goethe Universität Frankfurt, Frankfurt am Main, Germany, [3] École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [4] Swiss Institute of Bioinformatics, Lausanne, Switzerland

Human cytomegalovirus (HCMV) is a large dsDNA virus encoding the most complex genome of any human viral pathogen and is a leading cause of birth defects. Increasing evidence shows that HCMV exists as highly diverse populations in a broad range of human hosts, and the genetic diversity is a proposed novel therapeutic target. However, little is known about the limits or the patterns of HCMV genetic diversity. To address this deficiency, we have analyzed HCMV populations collected predominantly from symptomatic congenitally infected infants and sampled from a range of host compartments. We show that there is an upper limit to HCMV genetic diversity in this patient cohort, with approximately 35% of the genome being devoid of polymorphisms. These low diversity regions were distributed across 26 loci preferentially located in DNA processing genes. Furthermore, by developing the first genome-wide mutation and recombination maps for HCMV, we show that genetic diversity is positively correlated with these two rates. Lastly, we provide evidence that HCMV populations isolated from vascular compartments of different hosts are highly constrained and that polymorphisms in glycoproteins and regulatory proteins are enriched in these populations. This analysis provides the most highly detailed map of HCMV genetic diversity in human hosts to date, and informs our understanding of the link between HCMV diversity and disease.

426D

Evolution of HIV-1: geographic variation in substitution patterns at CTL epitopes

Helen Piontkivska [1], Reeba Paul[1], Madara Hetti Arachchilage[1], Austin Hughes[2]
[1] Kent State University, Kent, OH, USA, [2] University of South Carolina, Columbia, SC, USA

Human immunodeficiency virus (HIV) is a major public health challenge, with over 35 million people currently living with HIV (UNAIDS, 2013), of which a majority is residing in developing countries, primarily in Sub-Saharan Africa. While significant differences exist between levels of amino acid diversity among individual CTL epitopes, factors that influence variation in substitution patterns at individual epitopes sampled from different geographical regions with varying socioeconomic status are not well understood. Recently described set of so-called "associated epitopes" (e.g., Paul and Piontkivska, 2010), composed of highly conserved CTL epitopes that frequently co-occur together among different subtypes of HIV-1, allows us to evaluate whether differences in nucleotide diversity levels at these epitopes can be attributed to immune pressure alone, or whether other host population factors play a role in their evolution. In this study we examined the patterns of nucleotide diversity among different epitope and non-epitope regions in HIV-1 Pol gene sequences sampled from different countries to determine whether nucleotide diversity levels differ between geographical regions. Our results revealed significant differences in synonymous (πS) and nonsynonymous (πN) nucleotide diversity levels between groups of countries, suggesting that socioeconomic factors influence the evolution of viral epitopes (even if indirectly), albeit to different extent at associated and non-associated epitopes.

427A

Intrahost evolution of norovirus populations

Chris Ruis [1], Julianne Brown[2], Mette Christiansen[1], Daniel Depledge[1], Judith Breuer[1,2], Richard Goldstein[1]

[1] University College London, London, UK, [2] Great Ormond Street Hospital for Children, London, UK

Noroviruses are the leading cause of gastroenteritis worldwide and cause over 200,000 deaths annually. Norovirus infections of immunocompetent patients are typically acute and self limiting. Conversely, infections of immunocompromised patients are often chronic, lasting for weeks to years and can cause severe symptoms including malnutrition and intestinal dysfunction. This chronicity coupled with lack of immune control means the viral population within immunocompromised patients harbours many mutations and exists as a complex quasispecies. The evolution of the viral population within these patients is poorly characterised. Here, we use SureSelect technology to perform full genome deep sequencing of norovirus from multiple immunocompromised patients directly from clinical samples. We use a combination of variant analysis and haplotype reconstruction to characterise the evolution of the norovirus population at both local and global levels within these patients. We identify population-level shifts and correlate these with in depth clinical information to identify the clinical drivers of such events. Where possible, we employ techniques such as epitope prediction to identify putative sites driving this population level evolution. We characterise the selective pressure active on the viral population upon bone marrow transplantation, showing a large population shift consistent with reinstantiation of immune function. Importantly, we observe a decrease in the variability of the viral population and observe many rare variants rapidly increasing in frequency, suggesting a selective sweep. Our results show the norovirus population in immunocompromised patients is subject to changes in selective pressure and is able to rapidly adapt to such changes using its diverse quasispecies.

428B

Genotype-specific evolution of hepatitis E virus

Adam Brayne, Bethany Dearlove, Mukarram Hossain, Simon Frost
University of Cambridge, Cambridge, UK

Hepatitis E virus (HEV) causes large propagated epidemics of acute hepatitis in Asia and Africa, and low level, sporadic food-associated infections in the developed world. HEV comprises several genotypes with different geographic distributions, host ranges, pathology, and routes of transmission. Using a large annotated data set, we describe novel areas exhibiting positive selection and notable differences between genotypes. We find unusual areas of positive selection in overlapping reading frames and in the contentious HVR region. Furthermore, we demonstrate that open reading frame (ORF) 3 in genotype 1 evolves significantly more slowly than in genotypes 3 and 4, where ORF3 is under significant positive selection. We speculate that ORF3, whose function is related to immune evasion, is under relaxed selection pressure in human-specific genotype 1 due to high fitness, whereas adaptation to several host types provides ongoing selection pressure on ORF3 in genotypes 3 and 4. This work exemplifies important features of host-virus evolution, including different patterns between single host and zoonotic genotypes of HEV.

429C

Mammalian endogenous viral element database

So Nakagawa
Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

In mammals, it is widely accepted that approximately 10% of genome sequences correspond to endogenous viral elements (EVEs) including endogenous retroviruses (ERVs), which are thought to be derived from ancient viral infections of germ cells. Although most EVEs have been inactivated by insertions, deletions, substitutions, and/or epigenetic modifications, a few open reading frames (ORFs) of EVEs are still active and express viral proteins in the hosts. Indeed, several genes derived from EVEs have been found to be functional for host species, such as syncytins for placenta development in various mammals. However, no databases of EVE ORFs are available, and therefore evolutionary pathways of EVEs have not yet been understood comprehensively. Hereby, I developed EVE ORF databases for 19 mammalian species. I first identified EVEs using RetroTector and Repeat Masker. Obtained nucleotide sequences of EVE candidates were translated, lengths of which longer than 80aa were used in further analyses. For each sequence, motif sequences found in viruses were searched using hmmer3 with Pfam and Gypsy databases. Then all obtained EVE sequences as well as exogenous and endogenous viral sequences were searched against every genome sequence using BLAT. As a result, comprehensive EVE ORFs are identified for each species: for example, 144,013 ORFs (>80aa length) derived originally from EVEs are found in the cattle genome. This EVE database would be useful to study EVEs involved in various biological processes among mammals.

430D

Protein intrinsic disorder in human and simian immunodeficiency viruses (HIV/SIV) as a mediator for pathogen-host interactions

Felix Feyertag [1,2], Xiaowei Jiang[1,3], David Robertson[1]
[1] University of Manchester, Manchester, UK, [2] University of Nevada, Reno, USA, [3] University of Cambridge, Cambridge, UK

Protein intrinsically disordered regions (IDRs) play fundamental functional roles in biological systems; e.g., facilitating protein-protein interactions (PPIs). The HIV-1 genome encodes 9 genes with well-characterized interactions with human host-proteins, all of which have IDRs.

We analyzed disorder in 26 representative HIV-1 and HIV-2 genomes, and 2204 human proteins known to have human/HIV-1 PPIs. Additionally, we analyzed IDRs in 71 SIV genomes and identified proteins homologous to human proteins in 10 genera of non-human primate. We further classified host genes by gene ontology (GO) groupings, and investigated rates of intrinsic disorder between GO groups.

Our results indicate the highest rates of disorder were found in key regulatory genes *rev* (0.71±0.13) and *tat* (0.59±0.10), and significantly lower rates of disorder in *env* (0.10±0.03, $p<0.001$), *pol* (0.15±0.03, $p<0.001$), and *vif* (0.22±0.06, $p<0.001$). Interestingly, HIV-1 and HIV-2 had amongst the highest rates of disorder in *rev* (compared to *Chlorocebus*, $p<0.001$), *tat* (*Procolobus*, $p<0.001$), *gag* (*Procolobus*, $p<0.001$) and *nef* (*Colobus*, $p<0.001$). Mean disorder amongst human genes was 28%, comparable to homologous *Pan*, *Macaca* and *Gorilla* genes. When analyzing human host genes, 7/10 highest GO annotations were involved in regulation.

We investigated interaction between IDRs in HIV and human host genes, and performed comparative genomic analyses with SIV lineages and homologous proteins in simian genera, to investigate our hypothesis that pathogens mechanistically utilize IDR to manipulate virus-host PPI pathways, e.g. by targeting highly connected proteins. These results have direct implications on understanding evolutionary relationships between pathogenicity of HIV and why primates respond differently to species-specific SIV.

431A

Novel endogenous lentivirus in Galeopterus variegatus

Helena Fabryova, Tomas Hron, Jan Paces, Daniel Elleder
Institute of Molecular Genetics, ASCR, Prague, Czech Republic

Screening over 100 publicly available vertebrate genomes using an automatized BLAST approach we were able to confirm the presence of the already known endogenous lentiviruses in rabbit, domestic ferret, and gray mouse lemur. Not only did we confirm the previous findings, but we discovered a novel endogenous lentivirus in Malayan colugo, Galeopterus variegatus (denoted ELVgv). This was the first report of an endogenous lentivirus in an Asian mammal, indicating a long-term presence of this retrovirus family in the Asian continent. Sequence analysis of three ELVgv integrations points to evolutionary age between 5 and 27 million years ago. Galeopterus belongs to the mammalian order Dermoptera, which is close relative to the order primates. The only two extant genera in this order are Galeopterus and Cynocephalus. We were able to identify the same lentivirus in Cynocephalus volans, confirming the exceptionally old age of infiltration of dermopteran lineage by ELVgv. Further studies of the ELVgv sequences present in Cynocephalus and Galeopterus propose questions concerning the evolution of the relationships between the Dermoptera and lentiviruses. According to these analyses, ELVgv is probably the oldest lentivirus species described so far.

432B

Antiviral drug resistance as an adaptive process: a survey

Kristen Irwin [1,2], Jeffrey Jensen[1,2]

[1] *EPFL, Lausanne, Switzerland,* [2] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Antiviral drug resistance is a matter of grave clinical importance that, historically, has been investigated mostly from a virological perspective. While the proximate mechanisms of resistance can be readily uncovered using these methods, larger evolutionary trends often remain elusive. Recent interest by population geneticists in studies of antiviral resistance has spurred new metrics for evaluating the demographic history of infection, mutation rate, and selective pressures incurred during viral adaptation to antiviral drug treatment. We review recent advances in the field, with a focus on outstanding questions of clinical significance for a range of drug and viral types. We find that the demographic and selective histories revealed by population genomic inference are integral to assessing the evolution of resistance as it pertains to human health.

433C

Does transmission mode determine virulence in honeybee viruses?

Niklas Mather, Emily Remnant, Benjamin Oldroyd, Madeleine Beekman
The University of Sydney, Sydney, Australia

Sometime in the 20th century, the *Varroa* mite switched hosts from the Eastern to the Western honeybee, with disastrous consequences. *Varroa*'s spread into American and European bee populations was accompanied by outbreaks of viral disease, and a greatly increased rate of colony collapse. Intriguingly, there are no viruses unique to *Varroa*-infested colonies. Instead, the mite's presence appears to rapidly select for higher virulence in ordinarily benign pathogens, which *Varroa* carries between bees as it feeds on their haemolymph. These observations match the long standing, but untested, theoretical prediction that vector-borne pathogens should be highly virulent, because they do not rely on host activity to be successfully transmitted.

Varroa has not yet arrived in Australia, but most of its associated viruses are present in benign form, giving us a unique opportunity to experimentally determine the effect of transmission mode on virulence. We performed a long-term serial transfer experiment that mimicked the evolutionary conditions *Varroa* imposes on its associated viruses. We injected haemolymph drawn from asymptomatic pupae infected with Sacbrood Virus and Black Queen Cell Virus into pupae from the same colony. After a brief incubation, the haemolymph of the injected pupae was extracted and injected into the next generation. If vector-based transmission is associated with higher virulence, we expected that the virulence of the passaged viruses should increase as the experiment proceeded. We are currently using a combination of quantitative PCR, survival analysis and phylogenetics to determine whether such an increase in virulence has occurred.

434D

Recombination of globally circulating Varicella Zoster Virus

Daniel Depledge [1], Peter Norberg[2], Samit Kundu[1], Claire Atkinson[3], Julianne Brown[4], Tanzina Haque[3], Yusuf Hussaini[5], Eithne MacMahon[6], Pamela Molyneaux[7], Vassiliki Papaevangelou[8], Evelyn Koay[9], Julian Tang[9,10], Gillian Underhill[11], Tomas Bergström[2], Judith Breuer[1]
[1] University College London, London, UK, [2] University of Gothenburg, Gothenburg, Sweden, [3] Royal Free Hospital, London, UK, [4] Great Ormond Street Hospital, London, UK, [5] Ahmadu Bello University, Zaria, Nigeria, [6] St. Thomas' Hospital, London, UK, [7] Aberdeen Royal Infirmary, Aberdeen, UK, [8] University of Athens, Athens, Greece, [9] National University Hospital, Singapore, Singapore, [10] University Hospitals of Leicester, Leicester, UK, [11] Queen Alexandra Hospital, Portsmouth, UK

Varicella zoster virus (VZV), a human-infective alphaherpes with a worldwide prevalence exceeding 90%, is the causative agent of varicella (chickenpox) and herpes zoster (shingles), the latter caused by reactivation of the virus from latency. To prevent varicella in children, a live-attenuated heterogeneous vaccine preparation; vOka (Varivax®, Varilrix®) is used routinely in many countries worldwide while a similar preparation, but with a higher dose (Zostavax®), has recently been approved for use in the USA and Europe to prevent herpes zoster in the elderly.

A consequence of vaccination, as well as co- and/or super-infection with multiple wild type viruses, is the introduction of multiple VZV strains into the same individual. This creates conditions favoring genetic recombination, as has been observed in related human herpesviruses (e.g. HSV-1, EBV, HCMV) and this, importantly, could lead to the emergence of vaccine/wild-type recombinants.

Recent advances in sequencing methodologies have enabled sequencing of wild-type and attenuated VZV strains, yielding 115 full length genomes, the majority sequenced directly from clinical material. Following recombination analyses using GARD and RDP4, as well as analyses of linkage disequilibrium over varying genomic distances, significant insight is gained into the extent to which pervasive recombination has and continues to shape the evolution of VZV. Genes and domains under selection are also identified and discussed while the population structures within distinct sample types (vesicle fluid, blood, saliva) are examined in detail.

435A

Computational Analyses of the Influenza and Ebola Viral Genomes

Heidi Tessmer, Kimihito Ito
Hokkaido University, Sapporo, Japan

As our ability to obtain genomic data increases, so does our ability to conduct in-depth analysis of diseases, uncovering previously unknown patterns of evolution and spread. In the future, our ability to effectively combat zoonotic infectious diseases, such as influenza and Ebola virus disease, will require a thorough understanding of the evolutionary mechanisms of the viruses. These mechanisms can most efficiently be discovered through computational analysis of genomes from the nucleotide and protein levels up. This research focuses on rudimentary analysis of single and dinucleotide content in influenza and Ebola viruses. Previously identified trends in C+G content, as well as dinucleotide composition in these viruses, were corroborated and new viruses were analyzed, specifically the 2009 pandemic influenza virus and the Ebola virus from 1976 to the current epidemic. Our findings show similar patterns between human and avian H3N2 viruses, as well as between human influenza and the Ebola virus. Further research is needed to clarify whether these similarities are characteristic of all viruses or specific to the ones which we have chosen.

436B

Evolutionary dynamics of Bovine Viral Diarrhoea Virus 1 (BVDV-1) in Italy: phylogeography and contact networks

Francesco Cerutti [1], Erika Ebranati[2], Camilla Luzzago[3], Stefania Lauzi[3], Claudio Caruso[1], Loretta Masoero[1], Monica Giammarioli[4], Ana Moreno[5], Vincenza Cannella[6], Elena Canelli[5], Annalisa Guercio[6], Massimo Ciccozzi[7], Gian Mario De Mia[4], Pier Luigi Acutis[1], Gianguglielmo Zehender[2], Simone Peletto[1]

[1] Istituto Zooprofilattico Sperimentale Piemonte, Liguria e Valle D'Aosta, Turin, Italy, [2] Department of Clinical Sciences "Luigi Sacco", Section of Infectious Diseases, University of Milan, Milan, Italy, [3] Department of Veterinary Science and Public Health, University of Milan, Milan, Italy, [4] Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche, Perugia, Italy, [5] Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna, Brescia, Italy, [6] Istituto Zooprofilattico Sperimentale della Sicilia "A.Mirri", Palermo, Italy, [7] Department of Infectious Parasitic and Immunomediated Diseases, National Institute of Health, Rome, Italy

Bovine Viral Diarrhoea Virus genotype 1 (BVDV-1) is a RNA virus (genus Pestivirus) with large impact on cattle worldwide, causing economic loss in dairy and meat production.
5'-UTR sequences of the three main BVDV-1 subtypes circulating in Italy were considered: 1b (n=139), 1e (n=134), 1f (n=50). Phylogeography was performed with BEAST v1.8.0 grouping samples by areas of origin: Piedmont, Aosta Valley, Lombardy, Veneto, Emilia-Romagna regions and Central-Southern Italy. High similarity clusters were identified, and the relative sequences were input in Outbreaker (R package) to simulate transmission chains.
No significant geographic structure was observed in BVDV-1b tree, most of the sequences were from Lombardy and widely interspersed within geographically diverse clusters; subtype origin was traced to early XX century in Lombardy. BVDV-1e tree branched into three main clades and ancestry was estimated in Lombardy in 1990 (1988-1993). Piedmontese sequences clustered together, either due to later sample collection or to a recent introduction of BVDV-1e in Piedmont. BVDV-1f sequences clustered into three main clades, with a significant geographic structure: all Piedmont sequences clustered in two subclades, one including Aosta Valley sequences.
Within high similarity clusters, viruses carrying identical 5'-UTRs were identified in different farms thus suggesting a common infection source or a between-farm transmission. Npro gene sequencing and epidemiological investigations are ongoing for confirmation.
In conclusion, phylogeography and inferred contact networks suggest a regional transmission for 1e and 1f, while less evident for BVDV-1b. Understanding the origin and spread of BVDV is important for control strategies and prevention from introduction of new viral subtypes.

437C

Evolutionary origin and dispersals of oncogenic Human Papillomavirus 16 inferred from complete genomes and worldwide isolates

Ville Pimenoff [1], Cristina Mendes de Oliveira[1,2], Ignacio G. Bravo[1]
[1] Catalan Institute of Oncology/ IDIBELL, Barcelona, Spain, [2] Institute of Tropical Medicine, São Paulo, Brazil

Virtually all humans get infected in their lifetime by a variable number of papillomaviruses (PVs) at different anatomical sites. From more than 200 PVs retrieved from humans, a subset of closely-related PVs are identified as necessary agents for a large fraction of anogenital cancers. However, in most cases, PVs cause chronic, asymptomatic infections, eventually cleared by the host immune system.
Human Papillomavirus 16 (HPV16) is the most prevalent PV in anogenital cancers and in the genital tract of healthy individuals. Variation within the HPV16 lineage suggest certain degree of geographic structure or at least different HPV16 variants show differential prevalence in different regions. However, an in depth analysis of co-divergence between humans and HPV16 is still wanting. Moreover, there are no reliable estimates of viral substitution rate for HPVs, thus far, with all estimations made to date based on the assumption that the co-divergence of PVs with the host is correct.
Here, we present the most comprehensive phylogeographic analysis of HPV16 variant complete genomes and global variant isolates so far reported. We also compare the divergence time estimates obtained for HPV16 with human genome data. Our results support the hypothesis of limited co-divergent evolution of HPV16 with the human host, shaped by the heterogeneity of the human host immune response. In addition, our results suggest that other selective evolutionary phenomena might have played a more important role in the HPV16 lineage evolution than previously considered.

438D

Genomic diversity and evolution of Epstein-Barr virus

Fanny Wegner, Daniel Depledge, Florent Lassalle, Julianne Brown, Judith Breuer
University College London, London, UK

Epstein-Barr virus (EBV) belongs to the family of gamma-herpesviruses and is very successful in infecting humans. More than 90% of the adult population is latently infected for life. Primary infections are either symptomless or cause infectious mononucleosis. Due to the ability of the virus to immortalise B cells, EBV is associated with approximately 1.5% of human cancers worldwide (including Non-/Hodgkin's lymphoma, Burkitt's lymphoma, Nasopharyngeal Carcinoma (NPC) and Post-transplant lymphoproliferative disorders). Some EBV-associated diseases have higher incidence rates in specific parts of the world e.g. NPC, which is particularly prevalent in Southern China. Understanding the differences in pathogenesis of disease in the context of EBV variation is therefore of great interest.

Targeted enrichment of EBV sequences enables the direct sequencing and assembly of EBV genomes from clinical material such as blood. We recovered seven full EBV genomes from blood samples of immunocompromised paediatric patients and compared these to previously published sequences derived from tumours or lymphoblastoid cell lines. We found evidence for extensive recombination occurring between strains. Hotspots of linkage disequilibrium correlating with high nucleotide diversity were detected. However, while linkage disequilibrium is expected to decrease with increasing distance between sites, patterns of linkage between multiple distal sites in the genome were also found. Taken together, this points to the presence of a population structure within which recombination occurs. Finally, we report on several modes of selection acting on different parts of the genome and identify several potential new antigens and targets for antiviral therapy/vaccine development.

439A

Intra-patient co-evolution of adaptive immune system and pathogens

Jakub Otwinowski [2], Joshua Plotkin[2], Armita Nourmohammad[1]
[1] *Princeton University, Princeton, NJ, USA,* [2] *University of Pennsylvania, Philadelphia, PA, USA*

The adaptive immune system targets pathogens by mutating and selecting antibodies that bind to and neutralize antigens. Some pathogens, such as HIV, are able to persist in a host for extended periods of time, during which they also evolve to evade the immune response. We present a population genetics model of these coevolutionary dynamics between antibodies and antigens based on their binding phenotype, and we quantify how the antibodies and antigens adapt to each other. Some HIV patients are able to produce broadly neutralizing antibodies (BnAbs), which target conserved regions of the antigens where mutations may be lethal to the virus. Charactering these antibodies is extremely important for vaccine design against viruses. However, it is not clear when BnAbs have a selective advantage compared to other antibodies. Our model accounts for BnAbs by introducing high intrinsic fitness costs for mutations occurring at certain residues of the virus, which correspond to conserved regions, and we identify the regimes under which BnAbs may emerge.

440B

Inferring direct ancestry among densely sampled taxa using regularized phylogenetic trees

Prabhav Kalaghatgi, Thomas Lengauer
Max Planck Institute for Informatics, Saarbruecken, Germany

Phylogenetic analysis of pathogens can help characterize epidemic spread through a phylogeny-based reconstruction of transmission chains. The general approach in phylogenetic inference is to model evolutionary relationship using bifurcating trees with sampled taxa placed at the leaves of the tree. This approach may not be appropriate for pathogens collected during outbreaks where the likelihood of sampling an ancestor-descendant pair is not negligible. We model direct ancestry by placing sampled taxa at internal vertices of the tree and in addition allowing a multifurcating topology. We note that both of these modifications result in fewer unobserved ancestors (latent vertices) with respect to the bifurcating tree. We optimize trees under a likelihood framework that includes a penalty term for the number of latent vertices. In a preliminary analysis we tested this penalized likelihood approach on simulated trees with the number of taxa (n) ranging from four to fifteen, the number of latent vertices ranging from 0 to n-2 and branch lengths ranging from 0.005 to 0.05 subs/site. Each experiment was repeated 100 times. The penalized tree had a better test likelihood than the bifurcating tree in 77% of all experiments. In 83% of all experiments the penalized tree had the same topoogy as the simulated tree. Estimated branch lengths had a strong correlation ($\rho > 0.95$) with corresponding branches lengths in simulated troees. The penalized likelihood framework presented here exhibited good test performance and has the potential of reconstructing evolutionary relationships that are appropriate for pathogens sampled during outbreaks.

19 Novel functional approaches to evolutionary genomics

19.1

Developmental mechanisms underlying differences in cerebral cortex size in humans and other primates

Rick Livesey
University of Cambridge, Cambridge, UK

Relative cerebral cortex size and the total number of neurons in the cerebral cortex are thought to underlie the differences in cognitive ability between humans and other animals. Here we show that a primary determinant of cerebral cortex neuronal number in humans and non-human primates is a species-specific, cell-autonomous programme that controls the output of cortical progenitor cells, the neural stem cells of the developing cerebral cortex. Replaying cerebral cortex neurogenesis by directed differentiation of pluripotent stem cells from human, chimpanzee and macaque demonstrated that species-appropriate developmental timing was preserved *in vitro*. Human cortical progenitor cells exhibited distinct patterns of proliferation and differentiation, compared with those of the smaller-brained macaque, generating larger, more complex clones of neurons over longer periods of time. Each species' programme of cortical progenitor cell proliferation and clonal output was found to be species-specific. Control of developmental timing and clonal output from cortical progenitor cells was cell autonomous, unaffected by chimeric, inter-species cortical cultures. We conclude that cortical neuronal number, and thus cortical size, is regulated by a genetic programme that is implemented at the level of cortical progenitor cells and manifest in the clonal output of cortical progenitor cells.

19.2

Primate iPSCs provide a window into the evolution of key metabolic traits

Gregory Wray, Sasha Makohon-Moore
Duke University, Durham, NC, USA

Stem cells provide a valuable tool for identifying the genetic and molecular basis for adaptation and for studying gene-by-environment interactions. They allow one to control for genetic effects, remove environmental variability, and carry out replicate experiments. For species where ethical or conservation concerns preclude direct manipulation, stem cells also provide one of the few powerful experimental platforms available. We are using induced pluripotent stem cells (iPSCs) to identify genetic and gene-by-environment effects on the evolution of physiology during human origins. We differentiated iPSCs from chimpanzees and humans into adipocytes, the primary cell type of white adipose tissue. We then compared them to adipocytes derived from adult stem cells and *in vivo* adipose tissue using several ‘omic approaches in order to understand how faithfully they recapitulate the natural differentiation process. Our results demonstrate the utility of stem cells for providing insights into key metabolic traits that distinguish the two species that would be impossible to obtain using any other existing approach.

19.3

Positive Selection in Regulatory Elements Drives Changes in Gene Expression in the Adaptive Immune System of Primate Species

Charles Danko [1], Andre Martins[1], Hsiao-Pei Yang[1], John Lis[1], Lee Kraus[2], Adam Siepel[3]
[1] Cornell University, Ithaca, NY, USA, [2] University of Texas Southwestern Medical Center, Dallas, TX, USA, [3] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

The dramatic diversity of phenotypes observed in nature derives in large part from changes in gene expression across species. While RNA-seq studies have revealed many differences between primate species in stable mRNA concentrations, these observations reflect a mixture of changes at the pre- and post-transcriptional stages of mRNA biogenesis. Here, we attempt to disentangle these influences by systematically identifying changes in primary transcription that contribute to evolutionary differences in the primate immune system. We used PRO-seq to measure transcription in CD4+ T-cells isolated from humans, chimpanzees, and rhesus macaques. We combined our experimental approach with sensitive new machine learning tools to map the location of 47,886 regulatory elements (RE), including promoters and enhancers, and to measure the expression of 43,093 genes and non-coding RNAs. Approximately 3.3% of RE have undergone human lineage-specific changes, and 20% change activity in any one of the three primate species. Changes in distal enhancer activity are highly correlated with changes in the expression of nearby protein coding genes (R2 = 0.57). We applied INSIGHT to 166 human-specific RE, revealing signatures of both purifying and adaptive evolution. Divergences between species are enriched within putative transcription factor binding sites (TFBS) and DNase-I footprints, suggesting that adaptive changes modulate the binding affinity of transcription factors. We estimate that 35% of fixed nucleotide substitutions in these TFBS can be attributed to positive selection in the ancestors of modern humans. We show that adaptive changes in transcriptional regulation have contributed to human-specific changes in adaptive immunity.

19.4

Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies

Xueyan Li [1], Dingding Fan[2], Wei Zhang[3], Guichun Liu[1], Lu Zhang[2], Li Zhao[1], Xiaodong Fang[2], Lei Chen[1], Jun Wang[2], Guojie Zhang[2], Marcus Kronforst[3], Wen Wang[1]
[1] Kunming Institute of Zoology,Chinese Academy of Sciences, Kunming, Yunnan, China, [2] BGI-Shenzhen, Shenzhen,Guangdong, China, [3] University of Chicago, Chicago, Illinois, USA

Butterflies are exceptionally diverse but their potential as an experimental system has been limited by a lack of reference genome sequences and genetic manipulation technology. Here we use a hybrid assembly approach to construct high quality reference genomes for Papilio xuthus (contig and scaffold N50: 492 Kb, 3.4 Mb) and P. machaon (contig and scaffold N50: 81 Kb, 1.15 Mb), highly heterozygous species that differ in host plant affiliations and adult and larval color patterns. Integrating comparative genomics, genome scans of sequence divergence, and analyses of gene expression yielded multiple insights into butterfly evolution, including potential roles for specific genes in recent diversification. To functionally test candidate genes, we developed an efficient (up to 92.5%) CRISPR/Cas9 gene editing method that yielded striking phenotypes with three genes, Abdominal-B, ebony, and frizzled. Our results provide valuable genomic and technological resources for butterflies, and unlock their potential as a genetic model system.

19.5

Novel functional approaches to study regulatory evolution in cichlid fishes.

Claudius F Kratochwil [1,2], Axel Meyer[1]

[1] Department of Biology, University of Konstanz, Konstanz, Germany, [2] Zukunftskolleg, University of Konstanz, Konstanz, Germany

Cichlid fishes are one of the most diverse and species-rich vertebrate families and are famous for their astonishing rate of phenotypic diversification. It has been suggested that changes in non-coding DNA constitute a major component of their evolution. However, only a few non-coding regions with ecologically relevant effects have been identified so far. Research on non-coding DNA has been hampered due to the limited annotation of non-coding elements such as regulatory elements. The availability of six sequenced cichlid genomes and the increasing ability to perform functional experiments further elevate cichlids as a model that can help to understand the molecular underpinnings of phenotypically diverse traits. By performing ChIP-seq (chromatin immunoprecipitation with high-throughput sequencing) profiling of epigenetic modifications we were able to screen in a genome-wide, unbiased manner for active promoter and cis-regulatory elements in adult fin tissue of haplochromine cichlid fish. This novel dataset provides insights into the evolutionary dynamics and conservation of functional non-coding elements. To functionally test the regulatory elements discovered by this approach, we established the Midas cichlid *Amphilophus citrinellus* as a closely related model species to functionally assay and screen the activity of selected elements by transgenesis. Hereby, we combine selective putative regulatory elements with Green Fluorescent Protein (GFP) to assay their expression profiles *in vivo*. In summary, we introduce a powerful new approach that combines ChIP-seq and transgenesis and can largely help to identify unknown “loci of evolution”.

227A

Intraspecific Variation in the Rate and Phenotypic Effects of Spontaneous Mutation

Sarah Schaack
Reed College, Portland, OR, USA

Although there are now roughly a dozen reports providing direct estimates of mutation rates for multicellular eukaryotes, most studies quantify the rate for a single genotype and stop short of linking the mutations observed to their phenotypic effects. We will present recent data based on mutation accumulation experiments to examine a) levels of intraspecific variation in the mutation rate and b) examine the link between the rate of mutation and the phenotypic effects at the molecular, cellular, and whole organism level. This follows up on recent work showing the differential effects of spontaneous mutations in different environments, a key explanation for large levels of standing cryptic genetic variation in populations. Understanding the variation in mutation rate and the degree to which such genotypic variation results in phenotypic differences across genotypes, populations, species, and environments represents the current frontier in understanding how mutation rates evolve and therefore impact other phenomena at the interface of molecular and evolutionary biology.

228B

Evolution of a gene expression network underlying a disease state in humans and non-human primates

Courtney Babbitt [1], Gregory Crawford[2], William Nielsen[2], Gregory Wray[2]
[1] UMass Amherst, Amherst, MA, USA, [2] Duke University, Durham, NC, USA

We used a comparative approach to understanding differential disease susceptibilities between closely related species. Humans and chimpanzees have very similar genomes, but have different susceptibilities to a number of diseases, including epithelial cancer progression. We performed a serum challenge in human and chimpanzee cell culture over multiple time points to experimentally test this difference. We performed RNA-Seq and DNase-Seq (a measure of open chromatin) to understand how these species react differently to this important physiological response. Our results suggest that there are a number of important gene expression pathways in cell adhesion, oxidative stress, and cytoskeletal elements that have changed over evolutionary time to respond to this stressor, and there have also been significant changes in enhancer usage over evolutionary time. This experiment provides insights into the genetic pathways underlying the known differences in carcinoma rates between humans and chimpanzees.

229C

Structural, functional and evolutionary properties of genes that are neither duplicated nor deleted since Boreoeutherians.

Amelie Peres, Hugues Roest Crollius
Institut de Biologie de l'Ecole Normale Superieure, IBENS, CNRS UMR 8197, INSERM U1024, Paris, France

Finding genes that are involved in key steps for the life of an individual is an important goal in biology and medicine. A number of studies have identified so-called “essential genes”, defined from a functional point of view (e.g. that contain lethal mutations in model species). In this study we use a phylogenetic approach to identify genes that are never duplicated nor deleted in a given taxonomic group. These highly constrained genes are called “robust” and can be viewed as essential in a phylogenetic sense.

We focused on Boreoeutherian species and analysed 16 genomes available in Ensembl, where we identified 5754 robust genes since the Boreoeutheria ancestor (95 Mya), a number that vastly exceeds expectations under a model of uniform distribution of deletion/duplications in gene trees. In Human, these robust genes have a higher probability of being functionally essential (Fisher test, $P= 1.99x10^{-15}$) and of being involved in diseases (e.g. cancer: fisher, $P= 1.64x10^{-10}$; Alzheimer, $P= 3.08x10^{-22}$). They are enriched in pathogenic copy number variation (CNVs) and in GWAS reports. We have performed a logit regression on 14 structural, functional and evolutionary variables to identify putative discriminant properties of robust genes. Robust genes have a lower dN/dS ratio, are AT rich, larger, more expressed and preferentially positioned in larger chromosomes.

This study identifies a subset of approximately 25% of human genes that harbour unexpected phylogenetic, structural and functional properties, which together may help prioritise candidate disease genes and uncover new important biological processes.

230D

Evolutionary Changes in Promoter and Enhancer Activity During Human Corticogenesis

Steven K. Reilly [1], Jun Yin[1], Albert Ayoub[2], Deena Emera[1], Pasko Rakic[2], James Noonan[1,2]
[1] Yale University, Dept of Genetics, New Haven, CT, USA, [2] Yale University, Kavli Institute for Neuroscience, New Haven, CT, USA

Human higher cognition is attributed to the evolutionary expansion and elaboration of the human cerebral cortex. However, the genetic mechanisms contributing to these developmental changes are poorly understood. We used comparative epigenetic profiling of human, rhesus macaque, and mouse corticogenesis to identify promoters and enhancers that have gained activity in humans. These gains are significantly enriched in modules of coexpressed genes in the cortex that function in neuronal proliferation, migration, and cortical-map organization. Gain-enriched modules also showed correlated gene expression patterns and similar transcription factor binding site enrichments in promoters and enhancers, suggesting that they are connected by common regulatory mechanisms. Our results reveal coordinated patterns of potential regulatory changes associated with conserved developmental processes during corticogenesis, providing insight into human cortical evolution. Currently, we are testing the function of these human lineage gains of activity in humanized transgenic mice.

231A

Transcriptional Modularity and Functional Constraints in the Sexual Development of Nasonia vitripennis

Alfredo Rago [1], John H Werren[2], John K Colbourne[1]
[1] *University of Birmingham, Birmingham, UK,* [2] *University of Rochester, Rochester, NY, USA*

Genome-wide investigations are transforming our understanding of how changes in gene expression modulate phenotypic outcomes. It is now accepted that genes function within highly interconnected and context-sensitive networks rather than as independent units. The pervasive interactions among gene products (RNA and proteins) complicate models for the molecular evolution of adaptive traits, as the functions of genes are constrained by their partners.
This dependency among interacting genes can be reduced in networks composed by semi-independent modular clusters of genes. Modular networks can evolve either by increasing the degree of coregulation among functionally linked genes or by decreasing their interactions outside of the functional module (pleiotropy). Measuring how such network parameters change in response to alterations in the environment or genetic background is fundamental to understand whether and how modularity can evolve.
We designed a computational pipeline dedicated to the analysis of gene expression that explicitly tests how network structure and gene-gene relationships are altered either by environmental changes (ecological framework) or heritable variation (evolutionary framework). In this work, we apply our pipeline to the sexual development of the wasp Nasonia *vitripennis*. We define highly-coregulated transcriptional modules and detect their dynamic changes in coregulation and pleiotropy across the sexes in different life stages, highlighting the possible implications for the evolution of developmental processes.

232B

Genetic diversity of fluorescent protein genes generated by gene duplication and alternative splicing in reef-building corals

Shiho Takahashi-Kariyazono, Yoko Satta, Yohey Terai
SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan

Reef-building corals (Scleractinia) show various colors, and fluorescent proteins (FPs) are a major determinant of the colors. Gene duplication is considered a major mechanism to have generated *FP* gene family and color diversity. Examining gene duplications and evolution after those events may improve our understanding of *FP* gene family diversity. We isolated a novel *FP* gene family from *Montipora* sp. individual #5. This gene family consists of at least four genes and the four genes produce at least six different cDNAs. The cDNAs were categorized into two types based on the length of cDNA and this difference is owing to alternative splicing. The emission spectra of the *M5GFP* variants were nearly identical (518–521 nm). In addition to this gene family, we isolated ten paralogous *AdiFP10* sequences from cDNAs of two *Acropora* species, *A. digitifera* and *A. tenuis*. Based on the phylogenetic analysis, five sequences from *A. digitifera* and four sequences from *A. tenuis* appeared to be in a different cluster from *AdiFP10*, suggesting a new *FP* gene cluster. They were likely to have been generated independently in each species or generated by gene duplications in the ancestral lineage of *Acropora* and followed by an extensive gene conversion within each species. Our results clarified the diversification process of *FP* genes during the evolutionary history of corals. We showed possible correlation between differences in fluorescent color and modes in evolution, i.e., conserved or divergent evolution, after the duplication of *FP* genes.

233C

The genome of Antarctic copepoda Tigriopus kingsejongenesis and adaptation to life in the Antarctic Environment

Seunghyun Kang [1], Do Hwan Ahn[2], Sanghee Kim[1], Hyun Park[1,2]
[1] Korea Polar Research Institute, Incheon, Republic of Korea, [2] University of Science & Technology, Daejeon, Republic of Korea

Antarctic marine invertebrates face extremely cold temperatures and many of decapod crustacean and fish groups became extinct because of extreme climate for over the last 30 million years. In that matter, species which have survived in Antarctic region may have evolutionary strategies and understanding their adaptation mechanisms in response to the extreme environment has received considerable attention. However, genome-wide studies about the molecular basis underlying these mechanisms is still limited to fishes and microbes. Here we present the first draft genome sequence and annotation for Antarctic copepoda Tigriopus kingsejongensis, the first Antarctic Crustacean to be sequenced. We sequenced genomic DNA and RNA of T. kingsejongensis using Illumina Miseq platform and the libraries were prepared with average coverage of 120.7x. The final assembly consists of 48,368 contigs with an N50 contig length of 17.5 kilobases (kb) and 27,823 scaffolds with N50 contig length of 138.2 kb and a total of 39,717 coding genes were inferred using the MAKER annotation pipeline approach. The comparative genome analysis among 3,254 orthologs in 4 arthropod species (T. kingsejongensis, Tigriopus japonicus, Daphnia pulex and Drosophila melanogaster) revealed the T. kingsejongensis specific signals of molecular adaptation in genes associated with mitochondrial electron transport, deacetylase activity, proteasomal ubiquitin-dependent protein catabolic process, endoplasmic reticulum, and tryptophan metabolism. This suggest that T. kingsejongensis have changed adaptation mechanisms such as energy production and metabolism, proteolytic complex, and sterol biosynthesis. The results have important implications for understanding of Crustacean evolution and their adaptations to the Antarctic environment.

234D

5’ end overlapping genes in human and mouse genomes

Wojciech Rosikiewicz [1], Yutaka Suzuki[2], Izabela Makalowska[1]

[1] Department of Bioinformatics, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznan, Poland, [2] Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

For a long time, genes overlapping in the Metazoan, unlike the viral genomes, were not expected to be found. Nowadays more and more of different types of overlapping genes, depending on their position and the transcription direction, is reported in diverse species. Here we investigate role of 5' ends gene overlap in various human and mouse tissues / cell lines, trying to determine role of this phenomenon in regulation of gene expression and its scope in studied genomes. We analyzed gene coordinates and positions of alternative transcription start sites corresponding to them, what allowed us to detect 592 pairs of genes overlapping with their promoter regions in human and 113 pairs in mouse. We show that the number of genes identified as overlapping in each studied cell line differ enormously and that different cell line types shows different patterns of overlapping genes expression. What is more, we have elaborated ratio describing extent of promoter region overlap between two genes, which informs about the “preference” of expression using overlapping transcription start sites. We demonstrate, that such ratio may be used to find genes highly expressed from overlapping promoters, despite the fact, that theoretically they could be transcribed from non-overlapping start sites.

235A

The evolution of transcription errors

Weiyi Li [1], Jean-François Gout[1], W. Kelley Thomas[2], Michael Lynch[1]
[1] Department of Biology, Indiana University, Bloomington, Indiana, USA, [2] Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire, USA

Errors can occur at any level during replication and expression of genetic information. Although genetic mutations, which are mainly derived from replication errors, have been extensively studied in evolutionary research, many details regarding transcription errors remain largely unknown. With the recent combination of mutation accumulation approaches and the next-generation sequencing technology, the full spectrum and rate of mutations were well investigated. Based on these results, the drift barrier hypothesis, which argues that the lower limit of mutation rate is set by the power of random genetic drift instead of the intrinsic physiological limitations, has been proposed as a potentially unifying explanation for mutation rate evolution. However, evolution is a multi-layer issue covering genome, transcriptome and proteome. And transient intracellular errors, such as transcription and translation errors, are also of great importance. As a first step to investigate evolution of transcription errors, we recently developed a novel replicated sequencing (Rep-seq) method to accurately identify transcriptome-wide transcription errors and applied it in *Saccharomyces cerevisiae, Escherichia coli, Mesoplasma florum* and *Caenorhabditis elegans*. By comparing variations of the spectrum and rate of transcription errors across species, we can start to address the issue of transcription error evolution.

236B

A Haplotype Method Detects Diverse Signatures of Local Adaptation from Genomic Sequence Variation

Jeremy Lange, John Pool
University of Wisconsin-Madison, Madison, Wisconsin, USA

Identifying genomic targets of population-specific positive selection is a major goal in several areas of basic and applied biology. However, it is unclear how often such selection should act on new mutations versus standing genetic variation or recurrent mutation, and furthermore, favored alleles may either be fixed or remain variable in the population. Very few population genetic statistics are sensitive to all of these modes of selection. We have introduced and evaluated the Comparative Haplotype Identity statistic (χ_{MD}), which assesses whether pairwise haplotype sharing at a locus in one population is unusually large compared with another population, relative to genome-wide trends. Using simulations that emulate human and *Drosophila* genetic variation, we find that χ_{MD} is sensitive to a wide range of selection scenarios, and for some very challenging cases (*e.g.* partial soft sweeps), it outperforms other frequency- and haplotype-based statistics. We also find that, as with F_{ST}, our haplotype approach has the ability to detect surprisingly ancient selective sweeps. The simplicity and utility of χ_{MD} will make it an especially valuable tool in the search for genes targeted by local adaptation.

20 Evolution of molecular pathways and networks: Molecular evolution meets systems biology

20.1

The evolution of heritable human disease: what's a disease gene?

David ROBERTSON
UNIVERSITY OF MANCHESTER, MANCHESTER, UK

Genome-wide and exome sequencing studies are yielding insights into the extent of genetic variation in the genomes of humans, of which a significant component can be associated with disease susceptibility. However, it is now apparent that the scale and complex nature of inter-individual variation has been significantly underestimated. As a consequence consistently linking genotypic changes to specific disorders remains a massive challenge. In our research we have been investigating the evolution and properties of genes associated with heritable disease mutations via a systems-level integration of available molecular data, i.e., by studying variation in the context of molecular interaction networks and the organization of molecular function. In this talk I will discuss the importance of local interaction networks/modularity, "rewiring", dosage and systems-level compensation, for example, mediated by paralogous gene copies. For example, we have found a surprisingly strong link between gene duplication and disease-associated genes. On the face of it this would seem contrary to the expectation that duplicates will compensate for dysfunction. However, what appears to be happening is the presence of a duplicate gene, particularly those arising from whole-genome duplication (ohnologs), has been permissive for the accumulation of mutations and, in the case of heritable disease, increased the chance of the persistence of slightly deleterious mutations in line with the neutral theory of molecular evolution.

20.2

An adaptive scenario for the origins of complex innovations

Balázs Papp [1], Balázs Szappanos[1], Csaba Pál[1], Bálint Csörgő[1], Viktória Lázár[1], Gergely Fekete[2], Richard Notebaart[2], Xiaowen Lu[2], Claus Jonathan Fritzemeier[3], Martin Lercher[3]
[1] Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary, [2] Radboud University Medical Centre, Nijmegen, The Netherlands, [3] Heinrich-Heine-University Düsseldorf, Düsseldord, Germany

How innovations originate remains a central challenge in evolutionary biology. Innovation in metabolism allows the utilization of new nutrients, and arises through the integration of new metabolic reactions into the network. In many cases, metabolic innovations depend on the simultaneous acquisition of multiple reactions that provide little or no benefit individually. It has been argued that such complex innovations may arise through the non-adaptive exploration of phenotype space, but it remains unclear if such processes are indeed widespread and fast enough to explain the metabolic diversity observed even among closely related species. Here, we investigate how complex metabolic evolution can instead arise through purely adaptive processes. We traced in silico how bacterial metabolic networks can evolve across hundreds of different nutrient conditions. The analysis revealed that the *Escherichia coli* network can generally utilize novel nutrients through the addition of just one to three metabolic reactions. We demonstrate that temporally varying nutrient conditions can accelerate the adaptive expansion of metabolic networks: environments accessible through the addition of a single reaction serve as stepping stones towards the establishment of more complex pathways. Contingent gain of metabolic genes on the bacterial tree of life, distribution of nutrient utilization across 168 *E. coli* strains and results of a short-term laboratory evolutionary study in the same species provided empirical support for the scenario. We conclude that complex innovations in metabolic networks can evolve through a series of adaptive steps without the need to invoke non-adaptive processes.

20.3

Phenotypic and molecular evolution of the bacterial chemotaxis network

Alberto Nakauma Gonzalez, Sander van Doorn
University of Groningen, Groningen, The Netherlands

Rather than acting directly on individual genes, selection often acts on phenotypic properties that emerge from interactions between genes and/or proteins. Evolutionary theory is ill equipped to deal with such complexity, which makes it hard to disentangle the relationship between molecular evolution and phenotypic adaptation. One way to confront this challenge is to investigate the evolution of small signal transduction networks, whose biological function relies on interacting components but which are still simple enough to allow for a computational reconstruction of the mapping from genotype to fitness. We illustrate this bottom-up approach for the bacterial chemotaxis network, a well-studied signal transduction cascade that provides a molecular basis for fitness-relevant behavior. Starting from existing systems-biology models of the chemotaxis network in *E. coli*, we capture the biochemical constraints that restrict the effects of mutations and that result in phenotypic trade-offs between efficiency and the ability to show rapid sensory adaptation. We then calculate fitness based on simulated chemotactic performance for a large number of different mutant genotypes, systematically sampling the high-dimensional genotype space in which evolution occurs. The reconstructed adaptive landscape features high levels of epistasis, but we do not observe the neutral paths found in other types of reconstructed fitness landscapes. In addition, the topology of the fitness landscape is highly inhomogeneous. We discuss the consequences of these properties for phenotypic adaptation, and derive predictions on which components of the chemotaxis network are most likely to evolve as a response to environmental change.

20.4

Modeling Duplicate Genes in Populations, in Genomes, and in Complexes

Ashley Teufel[2,1], Anke Konrad[2,5], Joanna Masel[3], Liang Liu[4], David Liberles[1,2]
[1] Temple University, Philadelphia, PA, USA, [2] University of Wyoming, Laramie, WY, USA, [3] University of Arizona, Tucson, AZ, USA, [4] University of Georgia, Athens, GA, USA, [5] University of New Mexico, Albuquerque, NM, USA

Multiple processes give rise to the age distributions of duplicated genes observed in metazoan genomes. Beyond birth-death models that model a single duplication rate and a single loss rate, we have previously proposed a mechanistically inspired model to account for different loss processes and have applied this as a mixture process. We now extend this modeling to account for population genetic expectations of copy number variation that will never fix, of variation in the duplication process, and of greater mechanistic realism in the characterization of loss. This includes a discussion of the relationship between treating each protein independently and the function of a protein through inter-molecular interaction in a complex or in a pathway.

20.5

Turnover of phosphorylation sites in disordered regions: exploring a stabilizing selection hypothesis

Taraneh Zarin, Alan Moses
University of Toronto, Toronto, Canada

Intrinsically disordered regions of proteins are known to contain a wealth of post-translational regulatory sites involved in signaling pathways. While many of these are conserved amongst closely related species, others appear to be stochastically fluctuating both in number of modification sites and their location in the primary amino acid sequence through evolutionary time. As of yet, the mechanisms that determine these evolutionary patterns are not known. We hypothesize that these regulatory sites contribute to aggregate, quantitative functions that are preserved through stabilizing selection to maintain signaling fidelity. To test whether stabilizing selection is operating in disordered regions of proteins, we have focused on characterized phosphorylation sites in conserved MAP kinase signaling pathways that show drastically different numbers and positions in related yeast species. Using quantitative time-lapse microscopy, we are able to directly compare MAP kinase dynamics in fluorescently labeled cells with engineered differences in phosphorylation site organization. Consistent with stabilizing selection, we find that orthologous disordered protein regions from related species show similar MAP kinase dynamics to the wildtype, despite little conservation at the amino acid sequence level.

20.6

Inferences of selection in protein networks from genomic data and detailed network- and population-scale models

Ryan Gutenkunst
University of Arizona, Tucson, AZ, USA

I will first discuss recent "bottom-up" work from my group based on biochemically-detailed computational models of particular networks. Through sensitivity analysis, we used these models to assess the functional effects of smaller perturbations to protein activity than are accessible by traditional experiments. This measure of sensitivity predicted protein evolutionary rates as well as expression level, which has long been thought to be the dominant factor. I will also discuss a recent "top-down" genomic scan my group has undertaken for statistical signatures of selection in human African Pygmies. To calibrate our scan, we developed a complex neutral null model based on genome-scale simulations incorporating demographic history and heterogeneity of recombination and mutation rates. Remarkably, the majority of pathways inferred to be under selection by an F_{ST}-based enrichment test were not statistically significant using our more complete null model. Together, these results illustrate the increasingly important role detailed models play in understanding selection on complex networks.

20.7

The protein-protein network hub, Hsp90, does not provide robustness to new mutations in yeast

Kerry Geiler-Samerotte [1,3], Yuan Zhu[3], David Hall[2], Mark Siegal[1]
[1] New York University, NY, NY, USA, [2] University of Georgia, Athens, GA, USA, [3] Stanford University, Palo Alto, CA, USA

The molecular chaperone Hsp90 has been proposed to buffer, or mask the phenotypic effects of, mutations by mediating folding of mutant proteins. Since Hsp90 is at the hub of large protein-protein networks, the consequences of Hsp90-mediated buffering are pervasive. Buffering may obfuscate the mapping from genotype to phenotype, increase evolutionary rates of Hsp90 clients, and make protein networks robust to mutation. However, Hsp90 has also been proposed to have an opposite effect - sensitizing organisms to the effects of mutation - via the same mechanism. In this case, Hsp90 mediated protein folding allows gain-of-function mutations to have immediate phenotypic consequences despite concomitant destabilization of protein structure. Hsp90-dependent evolutionary innovations are particularly common in cancers. Here, we quantify Hsp90's ability to buffer vs. sensitize organisms to the effects of mutations. We use high-dimensional morphometric analysis of over five million yeast cells representing 78 strains isolated from natural populations and 94 lines that acquired spontaneous mutations under reduced selective pressure. Corroborating prior studies, we find that Hsp90 tends to buffer standing genetic variation in natural populations. In striking contrast, we demonstrate that Hsp90 predominantly modifies the effects of new mutations in the opposite way, sometimes buffering but more often sensitizing cells to these effects. These results validate theoretical predictions that selection increases the frequency of buffered variation in natural populations. Our findings force reconsideration of Hsp90's contribution to shaping phenotypic variation, and raise the question of whether any non-redundant gene product increases capacity to buffer the effects of mutations.

20.8

Genomic landscape of compensatory evolution

Csaba Pal
Biological Research Center, Szeged, Hungary

While core cellular processes are generally conserved during evolution, the constituent genes differ somewhat between related species with similar lifestyles. Why should this be so? In this work, we propose that gene loss may initially be deleterious, but organisms can recover fitness by the accumulation of compensatory mutations elsewhere in the genome. To investigate this process in the laboratory, we investigated 180 haploid yeast strains, each of which initially displayed slow growth owing to the deletion of a single gene. Laboratory evolutionary experiments revealed that defects in a broad range of molecular processes can readily be compensated during evolution. Genomic analyses and functional assays demonstrated that compensatory evolution generates hidden genetic and physiological variation across parallel evolving lines, which can be revealed when the environment changes. Strikingly, despite nearly full recovery of fitness, the wild-type genomic expression pattern is generally not restored. Based on these results, we argue that genomes undergo major changes not simply to adapt to external conditions but also to compensate for previously accumulated deleterious mutations.

20.9

Molecular Evolution Informs Systems Biology:Inference of Global Gene Networks through Correlated Evolution

Raghav Partha[1], Zelia Ferreira[1], Nolan Priedigkeit[1], Allyson O'Donnell[2], Nathan Clark [1]
[1] Univeristy of Pittsburgh, Pittsburgh, PA, USA, [2] Duquesne University, Pittsburgh, PA, USA

Functionally related genes are influenced by shared evolutionary pressures and hence tend to experience rates that covary over time. This signature, Evolutionary Rate Covariation (ERC), can reveal associations between genes by quantitatively finding those with covarying rates across a species-rich phylogeny. Genome-wide ERC datasets in yeast, *Drosophila*, and mammals find this signature between genes in the majority of known protein complexes and metabolic pathways. To demonstrate the inferential power of ERC we have used it to discover novel genes acting in *Drosophila* mating response and in yeast protein trafficking. This talk further explores the predictive power of ERC to characterize the evolutionary forces shaping gene networks. We explain that ERC correlations largely originate from parallel changes in evolutionary pressure and expression levels, and not by compensatory co-evolution as was previously assumed. Furthermore, to gain a broad perspective of relationships within gene networks, we used ERC to construct a global cellular map of associations between distinct pathways and complexes in yeast and in humans. This broad evolutionary network accurately lays out the major axes of information flow through the cell and reveals intriguingly novel links between disparate functional entities, such as between the splicing machinery and the nuclear pore. Importantly, our ERC-based map used only evolutionary rates and yet was able to capture much of the functional information derived through years of high-throughput experimental work. Evolutionary information can thus be used in many contexts to study relationships within gene networks as well as to infer them.

20.10

Coevolution of Rubisco and its chaperones

Maxim Kapralov
Liverpool John Moores University, Liverpool, UK

The major enzyme responsible for assimilation of inorganic carbon into organic biomass in plants, algae and bacteria is ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco), which performance is often one of the limiting steps in photosynthesis and can greatly affect plant survival and crop yield alike. Composed of eight large and eight small subunits, plant Rubisco is one of the largest and most abundant enzymes that command a suite of assembly and catalytic chaperones. We present both *in silico* and *in planta* findings that show the importance of ancillary protein complementarity within the Rubisco complex for its adaptive evolution. Further we discuss evolutionary aspects of Rubisco biogenesis and performance in plastids as well as their biotechnological implications for recombinant Rubisco bioengineering in plants.

20.11

Adaptive origin of the genetic robustness of metabolic fluxes

Wei-Chin Ho, Jianzhi Zhang
Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Genetic robustness refers to the phenotypic invariance in the face of mutation. Although genetic robustness has been well documented, its evolutionary origin is controversial. Genetic robustness could be an intrinsic property of biological systems (intrinsic hypothesis), a result of direct natural selection for genetic robustness (adaptive hypothesis), or a byproduct of natural selection for environmental robustness (congruent hypothesis). To differentiate among these hypotheses, we conduct a flux balance analysis of the reconstructed metabolic network of *Escherichia coli*, treating the flux of each metabolic reaction as a trait. We first show that, for a reaction, the amount of flux reduction from its wild-type value positively correlates with the amount of fitness decrease, demonstrating the potential adaptive value of flux robustness. We then show that the flux robustness of a reaction to genetic and environmental perturbations increases with the importance of the reaction to fitness. Furthermore, this correlation is stronger for the actual metabolic network than random metabolic networks of the same size and function, refuting the intrinsic hypothesis. In addition, our results regarding genetic robustness remain qualitatively unchanged after the control of environmental robustness. Together, these findings unequivocally support the adaptive hypothesis of the genetic robustness of *E. coli* metabolic fluxes, consistent with the theoretical prediction that natural selection is sufficiently powerful to promote genetic robustness in large populations.

20.12

Evolution of Hierarchy in Bacterial Metabolic Networks

Aaron Goodman, Dylan Rhodes, Laura Groenendaal, Marcus Feldman
Stanford University, Stanford, CA, USA

Typically hierarchy is thought of antithesis of self-organization. Hierarchical systems, organized from above, are often contrasted with systems organized by the spontaneous coordination of their component parts. However, recent work has shown hierarchy itself can be an emergent property of self-organized systems, since they found a high degree of hierarchy in such natural systems as supply chains, software systems, food webs, and neural networks.

We evaluate the degree of flow hierarchies in bacterial metabolic networks, and find that hierarchy plays an important role in evolution and adaptation of bacteria to different environments. We found metabolic networks in bacteria living in host-associated environments are more hierarchical than in free living bacteria and that differences degree of hierarchy is closely correlated with phylogenetic distances.

20.13

Reconstruction of ancestral metabolic networks reveals evolutionary gene-reaction-phenotype clustering in the *E. coli* pangenome

TY Pang, Martin Lercher
University of Duesseldorf, Duesseldorf, Germany

The metabolic network of *Escherichia coli* is the best-understood genome-scale system, and hence an ideal system to systematically analyse the relationship between evolutionary change and phenotype. Starting with 55 *E. coli* strains for which published metabolic network models exist, we reconstructed the genomes and metabolic networks of 54 ancestral strains using maximum parsimony. We applied constraint-based-modeling to determine phenotypic properties, i.e., ability to utilize different nutrients, of the ancestral strains. We applied flux variation analysis (FVA) to detect the reaction-phenotype associations. A reaction is associated with a phenotype if it has non-zero minimal flux in FVA, and is essential if its knockout destroys the phenotypic property. This reactions-phenotypes association of a metabolic model is represented by a bipartite network with nodes representing reactions and phenotypes with edges connecting associated reaction-phenotype-pairs. We investigated how gains and losses of reactions affect this reaction-phenotype-network. Further, we considered genes, reactions, and phenotypes as network nodes, and connected statistically significant co-gains and co-losses with edges to form a co-transfer-network. Our analysis revealed the clustered organisation of this gene-reaction-phenotype network, which may serve as an intuitive way to define functional groups in the *E. coli* pangenome. Small isolated clusters containing nodes with similar functions dominate ~80% of the network, and the remaining ~20% form an extended cluster. While most clustered gene-pairs are physically close and hence likely candidates for co-transfer in the same event, ~20% are instead separated by large genomic distances and are likely to be derived from different transfer events.

20.14

Evolution of bow-tie architectures in biology

Tamar Friedlander [1], Avraham E. Mayo[2], Tsvi Tlusty[3], Uri Alon[2]
[1] IST Austria, Klosterneuburg, Austria, [2] Weizmann Institute of Science, Rehovot, Israel, [3] Institute for Advanced Study, Princeton, New Jersey, USA

Bow-tie or hourglass structure is a common architectural feature found in many biological systems. A bow-tie in a multi-layered structure occurs when intermediate layers have much fewer components than the input and output layers. Examples include metabolism where a handful of building blocks mediate between multiple input nutrients and multiple output biomass components, and signaling networks where information from numerous receptor types passes through a small set of signaling pathways to regulate multiple output genes. Little is known, however, about how bow-tie architectures evolve. Here, we address the evolution of bow-tie architectures using simulations of multi-layered systems evolving to fulfill a given input-output goal. We find that bow-ties spontaneously evolve when the information in the evolutionary goal can be compressed. Mathematically speaking, bow-ties evolve when the rank of the input-output matrix describing the evolutionary goal is deficient. The maximal compression possible (the rank of the goal) determines the size of the narrowest part of the network - that is the bow-tie. A further requirement is that a process is active to reduce the number of links in the network, such as product-rule mutations, otherwise a non-bow-tie solution is found in the evolutionary simulations. This offers a mechanism to understand a common architectural principle of biological systems, and a way to quantitate the effective rank of the goals under which they evolved.

20.15

High-throughput methods reveal Evolutionary Dynamics of the core regulatory networks in placental mammals

Mauris Nnamani, Koryu Kin, Cong Liang, Gunter Wagner
Yale University, New Haven, CT, USA

The establishment of pregnancy in many placental mammals requires the coordination of embryo development, placental formation and the differentiation (decidualization) of endometrial stromal cells (ESCs) in response to the hormone progesterone, the second messenger cAMP, and in some species, fetal signals. The decidual process is characterized by the expression of a variety of phenotypic biomarkers including prolactin (PRL), WNT4, and insulin-like growth factor binding protein-1 (IGFBP-1). Decidualization evolved in the stem linage of mammals, and human decidualization requires many indispensable transcription factors including the homeobox protein HoxA-11, the forkhead protein FOXO1a, and C/EBP.

Here we propose that a key innovation for decidualization was the recruitment and reutilization of key transcription factors into endometrial fibroblasts cells. The chromatin modification, Trimethylation of histone H3 at lysine 4, is known to mark the start of actively transcribed genes. We show that the breath and intensity of these markers, coupled with transcriptome data, can be used to identify core regulatory networks that distinguish cell type identity genes. We tested the network with RNAi KD and found that FOXO1 and MYC KD affected the decidual reaction. Moreover, we discovered that part of the network functioned as a repressor of decidualization in the endometrial fibroblast network. We conclude that H3k4me3 chromatin modification is a promising tool for identifying, and provide insights into cell type specific core regulatory networks.

20.16

The interplay of gene regulatory network structure and evolutionary rate across different time and spatial scales.

Angela Early, Andrew Clark
Cornell University, Ithaca, NY, USA

The interactions among genes within regulatory networks are known to affect molecular patterns of species divergence, yet their effects on local adaptation are less clear. Here, we explore whether network structure affects patterns of population divergence and local adaptation in *Drosophila melanogaster*. In our analysis, we combine data on protein-protein interactions with genomic sequence data from 84 inbred *D. melanogaster* lines from five populations (Beijing, New York, the Netherlands, Tasmania, and Zimbabwe). Overall, our observations support a model in which genetic interactions amplify the strength and efficacy of purifying selection in functionally important regions, thereby constraining the evolutionary divergence of genes embedded within interaction networks. Across 12,000 protein-coding genes, rates of species-level nucleotide divergence correlate with intra-species levels of population differentiation (F_{ST} and K_{ST}; Spearman's partial rho=0.092, $P<0.001$) and nucleotide frequency shifts (Tajima's *D* and Fay and Wu's *H*; Spearman's partial rho=-0.037, $P<0.01$). In agreement with previous observations, measures of network centrality (betweenness and degree) correlate with levels of species divergence. Additionally, these network measurements also correlate with measures of population-level divergence and selection (Spearman's partial rho = -0.049, $P<0.001$). These correlations demonstrate that genome-wide selection pressures display some level of stability over time and spatial scales and suggest that network structure is a determinant of evolutionary rate under both scenarios. Using our *D. melanogaster* data, we conclude by demonstrating how incorporating measures of both network structure and species divergence into genome-wide analyses of local adaptation can better inform our identification of evolutionarily important outlier genes.

1A

Laboratory evolution of a pathogenic fungus in host cells rewires a signal transduction network

Sascha Brunke [1,2]
[1] *Hans-Knöll-Institut, Jena, Germany,* [2] *Center for Sepsis Control and Care, Jena, Germany*

Candida albicans is one of the most important fungal pathogens of humans, and its armament of virulence mechanisms is well investigated. For example, the ability to form hyphae allows the fungus to invade tissues and to escape from phagocytes. A complex network of signal transduction pathways integrates many different hyphae-inducing environmental signals to realize the filamentation program. The transcription factors Efg1 and Cph1 are central in this network, and mutants lacking both factors (*efg1*Δ/*cph1*Δ) are non-filamentous.

We found that an *efg1*Δ/*cph1*Δ mutant cannot escape from macrophages. This allowed us to perform a laboratory evolution experiment where *C. albicans* was kept in the harsh phagosome environment by continuous co-cultivation with macrophages. A months-long exposure resulted in a striking phenotypic alteration, as the non-filamentous mutant regained its ability to form hyphae and to escape from phagocytes. Filamentation was not limited to macrophages, but occurred in response to nearly all hyphae-inducing conditions. Importantly, this was also true during infection in mice, as much of the virulence lost by *efg1*Δ/*cph1*Δ was regained. Evidently, the two signal transduction factors were bypassed.

Using DNA and RNA sequencing, we detected a single nucleotide exchange which conferred hyphae formation capabilities to an independent *cph1*Δ/*efg1* Δ mutant. The mutation was found with the gene encoding a Ssn3 kinase, which forms part of the Mediator transcription co-activator complex. This adds a new and unexpected layer to the well-investigated signal transduction network of *C. albicans* hyphae formation. Detailed analyses of the new interactions inside this evolved network are now underway.

2B

Metabolic networks differ strongly in their evolvability

Claus Jonathan Fritzemeier [1], Balázs Szappanos[2], Balázs Papp[2], Csaba Pál[2], Martin J Lercher[1]
[1] Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany, [2] Biological Research Center, Szeged, Hungary

In parallel work, we have used a genome-scale metabolic network model to show how *Escherichia coli* can adapt to new environments via horizontal gene transfer. Typically, the addition of 2-3 reactions is sufficient for growth in a new environment that didn't support life for wildtype *E. coli*; here, viability is assessed via flux-balance analysis (FBA) simulations.

Is a small number of acquired genes generally sufficient for prokaryotic metabolic adaptation? To expand our one-species view, we repeated the analysis for 30 bacteria and archaea. To represent the space of possible metabolisms, we assembled 50+ metabolic networks [Ganter2013], including eukaryotes, into one supermodel with nearly 30,000 reactions and more than 10,000 unique metabolites. Due to inconsistent assignments of reaction directionalities, this supernetwork contains several hundred thermodynamically impossible energy-producing cycles, which had to be removed before our simulations.

We found that different metabolic networks vary greatly in the number of additional reactions required to establish growth in novel environments. Determinants of evolvability are original network size (which varies from a few hundred to more than 2500 reactions), the number of environments in which the original network is viable, network robustness, as well as general measures of network topology.

3C

Detection of pathways affected by positive selection in primate lineages ancestral to humans

Josephine Daub [1,2], Sébastien Moretti[3,4], Laurent Excoffier[1,4], Marc Robinson-Rechavi[3,4]
[1] University of Berne, Berne, Switzerland, [2] Universitat Pompeu Fabra, Barcelona, Spain, [3] University of Lausanne, Lausanne, Switzerland, [4] Swiss Institute of Bioinformatics, Lausanne, Switzerland

Gene set approaches are increasingly used to find signals of selection in the human genome. In this study we aim at detecting biological pathways affected by positive selection in ancient human evolutionary history, specifically in the four inner branches of the Primates tree that lead to modern humans. We tested all available gene trees of the Primates clade for the presence of selection in these four branches using the branch site test for positive selection. The results of all these tests were then used as input for a gene set enrichment test. We identified several pathways enriched for selective signals, mainly involved in immune response, sensory perception, lipid metabolism, and energy production. In addition to some gene sets being under selection in single branches, we found other gene sets to be significant in multiple branches. In the latter case, we find that different genes are responsible for the selection signal in different branches, suggesting that the same function has been optimized in different ways at different times in primate evolution.

4D

Strong negative correlation between longevity and substitution rates in mitochondrial genome of bivalves

Aurore Levivier [1], Nicolas Lartillot[2], Pierre Blier[1], France Dufresne[1]
[1] Université du Québec à Rimouski, Rimouski, Quebec, Canada, [2] UMR CNRS 5558, Université Claude Bernard, Lyon I, Lyon, France

Bivalves are good model systems to study evolutionary processes related to aging as they exhibit great variation in longevity. The veneroid, *Arctica islandica* is known to live up to 507 years whereas other species such as *Corbicula fluminea* only live five years. In this study, we used a comparative approach to examine: 1) how two life-history traits (generation time and lethal temperature) affect the longevity of 76 bivalve species, and 2) if patterns of mitochondrial DNA evolution are correlated to temperature, generation time, and longevity. A phylogenetic covariance model was built to measure substitution rates (dS), the ratio of nonsynonymous over synonymous substitution rates (dN/dS) and the ratio of radical over conservative amino acid replacement rates (Kr/Kc) in 76 bivalve species. Longevity was positively correlated with generation time and negatively correlated with lethal temperature. Substitution rates were positively correlated with lethal temperature and negatively correlated with longevity. Neither dN/dS nor Kr/Kc have an effect on any of the traits; likely due to the long time of divergence of some bivalve species (up to 530 millions of years). Therefore, we first clearly established a negative relationship between substitution rates and longevity in invertebrates.

5A

Origin and Evolution of Cellular Signaling Pathways:AKT/mTOR as a case study

Heloise Philippon [1], Céline Brochier-Armanet[1], Christine Brun[2], Evelyne Goillot[3], Guy Perrière[1]
[1] Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France, [2] Technologie Avancée pour le Génome et la Clinique, Marseille, France, [3] Laboratoire de Biologie Moléculaire de la Cellule, Lyon, France

The human AKT/mTOR signaling pathway plays a key role in autophagy and is activated through second messengers of extracellular signals named Phosphatidylinositol-3-kinases (PI3K). In order to understand the evolutionary history of the whole pathway, we first performed an in-depth phylogenetic analysis of the PI3K family. Divided into three main classes, PI3K are composed by catalytic and regulatory subunits. The catalytic subunits phylogeny indicates that two major duplications events occurred during the evolutionary history of eukaryotes. The most ancient arose before the Last Eukaryotic Common Ancestor (LECA) and led to the emergence of class III and class I/II catalytic subunits. The second duplication event, that led to the separation between classes I and II, took place in the ancestor of Unikonta. Given that classes I and III proteins directly induce cell autophagy in Human, the most interesting result was to find PI3K homologs in unicellular species, where autophagy could be seen as a self termination process! To make assumptions about their function(s) in these organisms, we reconstructed the phylogeny of 68 human proteins involved in the human AKT/mTOR pathway itself. For that purpose, we used the methodology previously developed for PI3Ks. Our preliminary results show that most of the proteins emerged in the last common ancestor of metazoans while key proteins such as AKT, mTOR and Beclin were already present in LECA.The comparison of the human protein-protein interactions network to the yeast and nematode ones is currently under way, in order to investigate the interaction conservation between homologous proteins.

6B

Natural selection in functional pathways and the emergence of evolutionary systems biology.

Marc Pybus, Pierre Luisi, Giovanni Dall'Olio, Brandon Invergo, Ludovica Montanucci, Hafid Laayouni, Jaume Bertranpetit
Institut de Biologia Evolutiva, Barcelona, Spain

Evolutionary analysis at the molecular level provide new tools to biology when considering the action of natural selection in genes and sets of genes in their functional setting of physiological pathways. Comparative analysis of selective pressures on sets of genes involved in a complex pathway or functional network may help disentangle the fine tuned purifying selection pressures.

As genes function in the context of molecular networks, with some occupying more important positions than others and thus being likely to be under stronger selective pressures, it is possible to interrogate how selection is distributes across the different parts of molecular networks. These analyses are telling us how evolution is shaping complex molecular pathways and networks, as the emerging function is a function of complex interactions.

These analyses may be undertaken at the pathway level (with a low number of interacting units but a very detailed molecular knowledge) to relate selection and the specificity of the reactions and function, or at the general level of all interactions among proteins. At the level of the human interactome, selection does not act equally at short or long time depth: genes with higher centralities are more likely to have been targeted by recent positive selection during recent human evolution. Our results indicate that the relationship between centrality and the impact of adaptive evolution highly depends on the evolutionary time-scale. Most likely, network adaptation occurs through intra-specific adaptive leaps affecting key network genes, followed by fine-tuning adaptations in less important network regions.

7C

Correlated evolutionary scenarios of metabolic functions

Murray Patterson, Daniel Kahn
University of Lyon 1, Villeurbanne, Rhone-Alpes, France

We are interested in the structure and evolution of metabolism inorder to better understand its complexity. We study metabolicfunctions in 1459 species within which several hundreds of thousandsof families of homologous genes have been identified [1]. Given aprotein sequence, PRIAM search [2] delivers probabilities of thepresence of several thousand enzymes (ECs). This allows us to inferreaction sets and to construct a metabolic network for an organism,given its set of sequences.
We then propagate these ECs to the ancestral nodes of the species treeusing maximimum likelihood methods. These evolutionary scenarios aresystematically compared using pairwise mutual information. Weidentify co-evolving enzyme sets from the graph of these relationshipsusing community detection algorithms [3,4]. This sheds light on thestructure of the metabolic networks in terms of co-evolving metabolicmodules. These modules are also interpreted from a functionalperspective using stoichiometric models of metabolic networks.
[1] Penel et al., BMC Bioinformatics, 10(6):S3, 2009[2] Claudel-Renard et al., Nucleic Acids Research, 31(22):6633--6639, 2003[3] Ahn et al., Nature, 446:761--764, 2010[4] Blondel et al., Journal of Statistical Mechanics, 2008(10):P10008, 2008

8D

Systems-level analysis of the antibiotic cross-resistance network

Réka Spohn [1], Viktória Lázár[1], István Nagy[1], Bálint Csörgő[1], Ádám Györkei[1], Ákos Nyerges[1], Balázs Horváth[1], Andrea Vörös[1], Róbert Busa-Fekete[2], Mónika Hrtyan[1], Balázs Bogos[1], Orsolya Méhi[1], Gergely Fekete[1], Balázs Szappanos[1], Balázs Kégl[1], Balázs Papp[1], Csaba Pál[1]
[1] Biological Research Centre - Hungarian Academy of Sciences, Szeged, Hungary, [2] Linear Accelerator Laboratory, University of Paris-Sud, CNRS, Orsay, France

Understanding how evolution of antimicrobial resistance increases resistance to other drugs is a challenge of profound importance. By combining experimental evolution and genome sequencing of 63 laboratory evolved lines, we charted a map of cross-resistance interactions between antibiotics in *Escherichia coli,* and explored the driving evolutionary principles. We demonstrate that 1) convergent molecular evolution is prevalent across antibiotic treatments, 2) resistance conferring mutations simultaneously enhance sensitivity to many other drugs, and 3) 27% of the accumulated mutations generate proteins with compromised activities, suggesting that antibiotic adaptation can partly be achieved without gain of novel function. By using knowledge on antibiotic properties, we examined the determinants of cross-resistance and identified chemogenomic profile similarity between antibiotics as the strongest predictor. In contrast, cross-resistance between two antibiotics is independent of whether they show synergistic effects in combination. These results have important implications on the development of novel antimicrobial strategies.

9A

Evolution of regulatory sequences considering expression noise

Georg Rieckh, Tiago Paixao, Nick Barton, Gasper Tkacik
IST Austria, Klosterneuburg, Austria

Gene regulatory elements not only encode the mean level of expression for different cellular environments, but also the variance around this mean. Equipped with a biophysically inspired genotype-phenotype map, we ask how this stochasticity influences the evolution of regulatory elements. Most importantly we show how this source of randomness interacts with another well-known source: finite population size.
We show that only for certain cases, increased gene expression noise is equivalent to a smaller effective population size and that even for large populations, the effect of gene expression noise can still be important. Furthermore, we point out how the topology of the gene network, promoter architecture and noise structure influence the final outcome of evolution -- both on the phenotypic and the genotypic level.

10B

Why are more robust transcription factor binding sites more evolvable?

Xinzhu Wei, Jianzhi Zhang
University of Michigan, Ann arbor, MI, USA

The relationship between robustness and evolvability has been a long-lasting controversy in biology. Recent studies of the neutral networks of transcription factor (TF) binding sites, RNA structures, and protein structures found that genotypic robustness hinders genotypic evolvability while phenotypic robustness (pR) promotes phenotypic evolvability (pE). Here we use mouse and yeast TF binding sites as a case study to understand the underlying cause of the positive correlation between pR and pE. We first show mathematically that the pR-pE correlation does not rely on the existence of a neutral network; even in random networks, pR and pE are positively correlated. However, the correlation is stronger in neutral networks than in random networks. We suggest that the neutral network of binding sites results from biophysical properties of protein-DNA binding rather than natural selection for robustness. Consistently, we find that robust genotypes are not preferentially used in vivo, nor do they occur more often in regulatory regions. We conclude that the positive correlation between pR and pE of TF binding sites is not a result of natural selection but a mathematical characteristic that has been fortuitously enhanced by the biophysical properties of protein-DNA binding. The same conclusion likely applies to the pR-pE relation in other systems.

11C

Identifying new enzymes of plant secondary metabolism - insights into the evolution of pyrrolizidine alkaloid biosynthesis

Lars H. Kruse [1], Christian Sievert[1,2], Dietrich Ober[1]
[1] Botanical Institute and Botanical Garden, Kiel University, Kiel, Germany, [2] The Biodesign Institute, School of Life Sciences, Arizona State University, Tempe, AZ, USA

Pyrrolizidine alkaloids (PA) are toxic secondary metabolites for chemical defense against herbivores in several, not closely related plant families. Until today only the first pathway-specific enzyme of PA biosynthesis is known, the homospermidine synthase (HSS). HSS catalyzes the formation of homospermidine and was recruited by duplication of a primary metabolism gene, which encodes the deoxyhypusine synthase (DHS). In Boraginales one independent gene duplication event resulted in the evolution of HSS. Further duplication events of the *dhs* gene are described for other angiosperm lineages. So far, in all studied PA containing plants, expression of HSS is restricted to specific cells and developmental stages of the plant.
We hypothesized that the PA biosynthesis occurs in the same tissue where HSS could be found. We identified candidate genes from different Boraginales species, which are associated with *hss* gene expression. cDNAs of candidate genes were cloned for heterologous expression and biochemical characterisation. We found a putative homospermidine oxidase that might be involved in the formation of the necine base, the typical backbone structure of PAs. Phylogenetic analyses suggest that this enzyme belongs to a subfamily of peroxisomal localised diamine oxidases. This subfamily also contains sequences encoding N-methylputrescine oxidases (MPOs) involved in tropane alkaloid pathway. MPOs seem to be recruited by gene duplication and following subfunctionalisation of a diamine oxidase involved in polyamine catabolism.

12D

Molecular characterization of Notch signalling during *Amphiura filiformis* (brittle star) arm regeneration

Prudence Lui, Anna Czarkwiani, Paola Oliveri
University College London, London, UK

The molecular evolution of regeneration remains difficult to address, because of the distant phylogenetic relationships between popular invertebrate and vertebrate regenerative models. Therefore echinoderms, deuterostome invertebrates, are an attractive model. They are closely related to chordates, yet show equally extensive regeneration like other invertebrates. In this work, we aim to characterize the Notch signalling pathway in the brittle star *Amphiura filiformis*, and ultimately determine if its role in embryonic development and adult arm regeneration is conserved.

We confirmed that most pathway components (*Afi-Notch, Afi-Delta, Afi-Serrate, Afi-Deltex, Afi-Numb, Afi-Fringe* and *Afi-Su(H)*) are present in the *A. filiformis* transcriptome. After cloning these genes, we assessed their expression during adult arm regeneration and embryonic development, using quantitative PCR and whole mount *in situ* hybridization. In adult regenerating arms most genes are expressed in the outer epithelium of the blastema, with the exception of *Afi-Delta* and *Afi-Serrate,* which are additionally expressed in the inner cell layers. Furthermore, we show that Notch pathway genes are expressed in the embryonic endomesoderm, in close resemblance to that of the sea urchin *Strongylocentrotus purpuratu.*

This work provides the first comprehensive view of Notch signalling in a regenerating echinoderm. Our results show that Notch signalling is present in both embryogenesis and adult arm regeneration of *A. filiformis*. This also suggests that the Notch signalling pathway components could be evolutionarily conserved in echinoderm embryogenesis and potentially adult regeneration. Future work will focus on discerning the role of Notch signalling in *A. filiformis* using functional perturbation analysis.

13A

Integrated gene regulatory network dynamics and evolution in the worm Caenorhabditis

Vanessa Vermeirssen [1], Jonas Defoort[1], Yves Van de Peer[1,2]
[1] *VIB-Ghent University, Ghent, Belgium,* [2] *University of Pretoria, Pretoria, South Africa*

Differential gene expression is a tightly controlled process that governs development, function and pathology of eukaryotic organisms. Several molecular interactions between regulatory proteins, RNA and DNA closely work together in order to establish proper gene expression in space and time. In order to get a systems level understanding of how different molecular interactions interrelate to form a coordinated response in gene regulation, we studied clusters of composite network motifs in integrated gene regulatory networks of the worm *Caenorhabditis elegans*. We found that these clusters of undirect protein-protein, genetic and homologous interactions, and direct interactions of protein-DNA, regulatory and miRNA-mRNA interactions form functionally relevant modules in gene regulation, thereby establishing the relation between the topology of gene regulatory networks and their function.
We investigated the dynamics of these network patterns during development and their evolutionary rewiring between species. Using the integrated network as ancestral network and mapping life cycle expression profiles of *C. elegans* and *C. briggsae* onto this network, dynamic species-specific gene regulatory networks and composite network motif clusters were created. Not only do these provide insight in how gene regulation changes over developmental stages, they also explain expression divergence of orthologs and paralogs. In conclusion, integration of molecular interactions by network motif clustering provides a means to identify the underlying gene regulation that leads to expression dynamics during development and expression divergence between species.

14B

Characterization of the Transcriptome of *Hermetia illucens* (Diptera:Stratiomyidae)

Raquel Monfardini [1], Alisson Gontijo[2], Tatiana Torres[1]
[1] *USP - Sao Paulo University, Sao Paulo, Brazil,* [2] *CEDOC - Chronic Diseases FCM Nova, Oeiras, Portugal*

Tissue development is finely regulated to maintain the functionality of adult organisms. In Drosophila larvae, ILP8 was recently discovered as a key protein involved on a metamorphosis delay mechanism and indicates the presence of abnormal imaginal disc growth. Despite its importance, little is known about its function and evolution; it is assumed that the *ilp8* gene is present in the genomes of true flies but absent in basal dipterans. To understand gene *ilp8* evolution it is necessary to determine when ilp8 evolved and its signaling pathways. The species chosen to pursue those questions was *Hermetia illucens* (Stratiomyidae), as a representative of intermediate taxa between the basal Diptera (without *ilp8*) and the true flies (with *ilp8*). As a first step, we obtained *H. illucens* whole transcriptome from RNA-seq data. To find *ilp8*, we browsed the transcriptome, using a specific motif. By doing this, it was possible to identify four ilp8 isoforms in *H. illucens* as well as in others species with an available transcriptome. With these sequences we will try and understand the evolutionary history of *ilp8*. Among the *H. illucens* annotated transcripts, we also found transcripts associated with metamorphosis, such as the Ecdysone receptor gene. We performed experimental assays using ethylmethanosulfonate (EMS) in *H. illucens* larva. EMS induces damage in tissues under development, triggering the metamorphosis delay response. This will allow the comparison among sets of genes expressed during normal and abnormal larval development and to identify differentially expressed transcripts (aside from *ilp8*) that may be involved with metamorphosis regulation.

15C

Exploring systems affected by the heat shock response in the malaria parasite via protein association networks

Hong Cai [1], Timothy Lilburn[2], Jianying Gu[3], Yufeng Wang[1]
[1] University of Texas at San Antonio, San Antonio, TX, USA, [2] Novozymes, Durham NC, USA, [3] City University of New York, Staten Island, NY, USA

The heat shock response is a general mechanism by which organisms deal with physical insults such as sudden changes in temperature, osmotic and oxidative stresses, and exposure to toxic substances. The malaria parasite *Plasmodium falciparum* is exposed to drastic temperature changes as a part of its life cycle and maintains an extensive repertoire of heat shock response-related proteins. As these proteins serve to maintain the parasite in the face of anti-malarial drugs as well, better understanding of the heat shock-related systems in the malaria parasite will lead to therapeutic approaches that frustrate these systems, leading to more effective use of anti-malarials. Here we use protein association networks to broaden our understanding of the systems impacted by and/or implicated in the heat shock response.

16D

Molecular evolution and structure of the GPCR signaling network explains their role in disease

Ana Barradas, David Robertson
University of Manchester, Manchester, UK

G protein-coupled receptors (GPCRs) are an important group of paralogous signaling molecules that play a central role in physiology. They are at the core of a signaling network that, when disrupted, frequently results in pathological outcomes. While GPCR dysfunction can be very broadly divided into gain-of-function or loss-of-function mechanisms, the molecular basis for malfunction is very diverse and may have their origins in defective network components other than the receptors. When compared to singletons, genes that have retained paralogues from their duplication history tend to be associated with human heritable disorders, especially if they have arisen by whole-genome duplication (WGD), and diseases caused by mutations in WGD genes have a tendency to be autosomal dominant. In order to assess the contribution of the duplication history of each GPCR network component to the emergence of disease, we determined the duplication status of the groups of genes in each component of the signaling network: GPCR, ligand, downstream signaling proteins, and GPCR-interacting proteins, and quantified their association to disease. We find that WGD genes have been retained in distinct proportions for the different components, and that although ligands exhibit a very small proportion of WGD, these have largely contributed to the disease development within the network. This finding is consistent with differences in genetic variation, where we observe a higher frequency of small insertions and deletions, and truncating mutations among the ligand genes. The majority of the GPCR-associated diseases are autosomal dominant, confirming the tolerance of WGD genes for variation and disease pervasiveness.

17A

Expression and Evolutionary Analysis of Starch Metabolism Genes in *Oryza sativa* L.

Ung-Han Yoon [1], Gang-Seob Lee[1], Tae-Ho Kim[1], Hyeon-So Ji[1], Yeon-Ki Kim[2], Jong-Yeol Lee[1], Tae-Ho Lee[1], Jang-Ho Hahn[1]

[1] National Academy of Agricultural Science, Jeonju, Republic of Korea, [2] GreenGene Biotech, Yongin, Republic of Korea

The expression analysis of rice starch metabolism genes and their comparative genomic analysis with other crops are important for utilization through crop improvement studies. We sequenced 51,383 ESTs from immature and germinating seeds of Ilpumbyeo for analysis of structure and expression of rice starch metabolism genes. We isolated 37 full length cDNA clones, and they are homologous to starch synthesis and degradation genes. We also searched for these genes and selected 156 genes in TIGR rice annotation DB. For the phylogenetic analysis, we aligned with both nucleotide and amino acid sequences of 37 kinds of rice starch metabolism related genes by MEGA 6 software. We inferred the evolutionary distances and built trees using other species of plants starch metabolism genes. To understand the functional role of starch synthesis and degradation genes, we were analyzed expression profiling of these genes in rice using 135K microarray. As a result, we observed high level of expression of the genes AGPase, starch synthase, starch branching enzyme, and pullulanase in immature seed stage. An amylase gene, Os08g047380 which is involved in starch degradation showed the strongest expression in germinating seed. It was revealed that these genes were expressed in a tissue specific manner. Further works focus on the comparative genomics and expression patterns on starch metabolism genes to other plant.

18B

Condition-dependent expression of genetic variation for flowering time in natural populations of *Arabidopsis thaliana*

Eriko Sasaki [1], Pei Zhang[2], Dazhe Meng[1,2], Magnus Nordborg[1]
[1] Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria, [2] Molecular and Computational Biology, University of Southern California, Los Angeles, California, USA

Understanding how genetic variation interacts with the environment is essential for understanding adaptation. In particular, the life cycle of plants is tightly coordinated with local environmental signals through complex interactions with genetic variation (*G* x *E*), the molecular basis of which poorly understood. We focused on temperature-dependent variation of flowering time among natural inbred lines of *Arabidopsis thaliana*. We collected flowering time data for 174 lines from Sweden under two growth temperatures (10 and 16ºC) and conducted genome-wide screening of *G* x *E* including interaction of *cis* regions of genes with environment (*cis* x *E*). Under warm conditions, the flowering time across lines was highly variable in comparison to cool conditions, indicating that genetic differences expressed under warm conditions were 'buffered' under cool conditions. We identified strong *cis* x *E* signals across genome, although the effects of single SNP were largely undetectable. Our results suggest that temperature-dependent variation of flowering time is due to many small effects, and will be extremely difficult to characterize molecularly.

19C

Mining, storage and evolutionary analysis of a high quality compendium of phosphorylation sites from high-throughput phosphoproteomic data.

Panayotis Vlastaridis[1], Pelagia Kyriakidou[1], Anargyros Chaliotis[1], Pavlina Chalyvopoulou[1], Stephen Oliver[2], Yves Van de Peer[3], Grigorios Amoutzias [1]
[1] University of Thessaly, Larisa, Thessaly, Greece, [2] University of Cambridge, Cambridge, UK, [3] VIB/University of Ghent, Ghent, Belgium

Protein phosphorylation is the most abundant post-translational modification and a very attractive area of research for understanding biological complexity, cell regulation, therapeutics, diagnostics and even for synthetic biology. By mutating phosphorylation sites, molecular pathways and phenotypes may be manipulated. We have mined the literature for HTP phosphoproteomic data with two expert annotators and generated a noise-free compendium of phosphorylation sites from several species using very stringent quality criteria. Next, we developed a java web application utilizing a MySQL database for manual literature annotation and a Solr server as a document search engine. In addition, we have developed one database system based on graphs (Neo4j) to organize and store the high-quality compendium of phosphorylation sites, together with evolutionary and functional information on the stored phosphorylation sites.

20D

Unraveling the effects of different types of essentiality in the evolution of the human testis proteome

Julia Schumacher, Hans Zischler, Holger Herlyn
Institute of Anthropology, Johannes Gutenberg-University, Mainz, Germany

The "knockout-rate prediction" implies that essential proteins, whose disruption leads to lethality or infertility, should evolve at lower rates than nonessential ones. However, male reproductive proteins, especially those with functions proximate to fertilization, have been reported to evolve rapidly. We explored proteins involved in human male sub- or infertility (‘infertility proteins') regarding their evolution, pleiotropy, and centrality in a human testis protein-protein interaction (PPI) network. As a proxy for human lethality, we used proteins with mouse orthologues whose deletion or loss of function causes prepubertal death (‘lethality proteins'). We show that both, infertility and lethality proteins differ in their patterns of evolution, expression, functionality, and network centrality when compared to the remainder of the dataset. Due to partial overlap between the infertility and lethality categories, we removed all proteins associated with prepubertal death in mice. This procedure enabled us to investigate the characteristics of infertility proteins relative to the remainder of the testis proteome without bias from the lethality category.

21A

Stimulus-dependent expression of the honeybee (*Apis mellifera*) homologs of the immediate early genes c-Jun and Egr-1

Frank Sommerlandt, Wolfgang Rössler, Johannes Spaethe
University of Würzburg, Biozentrum, Dept. of Behavioral Physiology & Sociobiology, Würzburg, Germany

Immediate early genes (IEG) are a group of genes that are rapidly and transiently expressed following neuronal activation and that do not rely on de novo protein synthesis. Based on these characteristics, IEGs are used as markers for neuronal activity in vertebrate brains. However, if their homologs in insects are functioning in a similar way, needs to be investigated. Here we report findings that two IEGs (the honeybee homologs of the transcription factors Egr-1 and c-Jun) are differentially expressed in honeybee workers after exposure to light and stress pheromone, compared with untreated controls. The expression differences are primarily restricted to higher order brain neuropils (the mushroom bodies, which serve as centers for sensory integration, learning and memory, and can be compared with the vertebrates' cortex area), whereas no expression differences were observed in the primary input regions (the antennal lobes and the optical lobes). These findings suggest that IEGs (and their products) are highly conserved key genes in the molecular processes of neuronal plasticity among the animal kingdom

22B

Timing of opsin's mRNA expression and the role of the transcription factor spineless during the pupal development of the honeybee (*Apis mellifera*) compound eye

Leonie Lichtenstein, Johannes Spaethe
University of Würzburg, Würzburg, Bavaria, Germany

Vision is an essential sensory modality in bees for flight orientation, recognition of nest sites and finding suitable food sources and mating partners. To achieve these tasks honeybees are equipped with a pair of compound eyes that consist of several thousand units (the ommatidia). Each ommatidium is composed of a set of nine photoreceptor cells, and each cell expresses a specific visual pigment (the opsin protein) and its conjugated chromophore (the retinal). Although the retinal represents the light sensitive part which is able to absorb light photons, the spectral sensitivity of a photoreceptor is determined by its opsin protein. The honeybee possesses three different types of photoreceptors which are most sensitive for ultraviolet (UV), blue and green wavelengths, respectively. Since Karl von Frisch first demonstrated color vision in honeybees more than 100 years ago, numerous studies have been performed to gain insight into the molecular, morphological and physiological basis of the compound eyes and their corresponding photoreceptors. However, little is known about the formation of the eyes during pupal development of the bees. In this study, we used real-time PCR to identify the temporal expression patterns of the UV, blue and green opsins during the pupal development of the honeybee worker compound eyes. Moreover, we explore the role of the transcription factor spineless during eye development and checked for a co-expression of specific opsins and spineless, which was shown in *Drosophila melanogaster* to be necessary and sufficient to determine the photoreceptor type "yellow" within the compound eye.

23C

Evolution and functional diversification of chaperonin paralogues genes in Cyanobacteria

Julia Weißenbach, Judith Ilhan, David Bogumil, Karina Stucken, Tal Dagan
Institute of Microbiology, Genomic Microbiology, Kiel,Schleswig-Holstein, Germany

The chaperonin GroEL and its co-factor GroES promote protein folding in an ATP-dependent manner. In most eubacteria the GroEL/S chaperonin is encoded by a single-copy bicistronic operon that includes the *groES* and *groEL* genes. Cyanobacteria encode in addition a monocistronic *groEL* gene, with the filamentous, heterocyst forming cyanobacteria encoding an additional *groEL/S* operon. A phylogenetic analysis shows that the chaperonin genes have been duplicated at least twice during cyanobacterial evolution. Here we study the functional diversification of GroEL/S paralogs in *Chlorogloeopsis fritschii* PCC6912. This cyanobacterium differentiates heterocysts and forms multi-seriate filaments. The *C. fritschii* genome encodes two *groEL/S* operons (*groE1*, *groE2*) and a monocistronic *groEL* gene (*groEL3*). A comparison of gene expression under stress conditions shows that *groEL1* is upregulated during temperature stress whereas the monocistronic *groEL3* is upregulated under light stress. Interestingly, the expression of *groEL2* is induced upon nitrogen deficiency during heterocyst differentiation. Transcriptional GFP fusions with each of the three *groEL* promotors further validated the increased expression of the paralogous *groEL* under the corresponding stress condition. In addition, *groE1* could complement a *groE* deficient *Escherichia coli* strain, but only when cultured at low temperature. Furthermore, to establish the GroEL-GroES specificity, we tested for protein interactions between the chaperonin subunits *in vivo*. GroEL3 subunits seem to neither form oligomers nor interact with any of the two co-chaperonins, whereas subunits encoded in the two operons may form hybrid complexes. Our results thus indicate that the GroE2 underwent a subfunctionalization while the GroEL3 underwent a neofunctionalization.

24D

Sequence co-evolution predicts residue-level protein interactions and oligomerization

Anna G. Green, Thomas Hopf, Charlotta Schaerfe, Debora Marks
Harvard Medical School, Boston, MA, USA

A detailed understanding of macro-molecular interactions is critical for interpreting the results of mutations on organismal phenotype. Such interactions remain difficult to measure in a systematic way due to biases in existing experimental methods. Our lab recently demonstrated that computational approaches using measures of residue co-evolution across proteins are able to detect functionally coupled residues, which are sufficiently close in space to yield three-dimensional structure of the protein complexes. We predicted the structures of 76 known and 32 previously unsolved protein complexes. Current work seeks to develop methods to determine the propensity of proteins to oligomerize using coevolution between interface residues. This research will provide tools to determine protein interactions and oligomerization from sequence alone, as well as insight into evolutionary constraint on macromolecular interactions.

25A

Pathway-dependent Robustness of Yeast Mitogen Activated Protein Kinase Kinase Kinase

Purnima S. Kompella, Sergio G. Peisajovich, Alan M. Moses
University of Toronto, Toronto, ON, Canada

Understanding how variations in a protein sequence affect function is crucial to our understanding of many aspects of biology, including pathology and evolution. While detrimental mutations lead to disease, neutral and beneficial mutations may lead to adaptation. Although the identification of function-altering mutations that lead to disease is crucial for targeting therapies, it is also important to understand how neutral variations in these genes can alter responses to therapies. Besides possible clinical applications, understanding how mutations affect phenotypes is important for understanding fundamental biological processes, such as evolution. Genes capable of tolerating mutations (i.e., robust genes) can accumulate large neutral variation in populations. Since diverse neutral pools of mutations are great reservoirs for potentially adaptive variations, robust genes are more likely to facilitate adaptation against future environmental or genetic changes. We are interested in understanding how the extent of tolerated variation differs for the same gene in different pathways. Since robustness depends on mutation rate,, we made random mutant libraries with increasing number of mutations per gene of the yeast MAPKKK, Ste11. Next, we measured the ratio of mutations that lead to wild-type like pathway function as detected by flow cytometry using a fluorescently tagged mating-responsive promoter or osmolarity-responsive promoter. Finally, we calculated the extent of tolerated variation as a ratio of neutral to non-neutral mutation. We find that the extent of tolerated variation for Ste11 differs based on pathway function.

26B

Identification of chimpanzee microRNA transcription start sites and divergence of the regulation of miRNA transcription between human and chimpanzee

Li-Ching Hsieh [1,2], Cheng-Kuo Lai[1]

[1] Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, Taiwan, [2] Biotechnology Center, National Chung Hsing University, Taichung, Taiwan, [3] Department of Physics, National Chung Hsing University, Taichung, Taiwan

Identifying the transcription start sites (TSSs) of microRNA (miRNA) genes is essential for understanding the regulation of miRNA transcription which is extremely important for determining the specific roles that miRNAs play in regulatory pathways. In addition, the divergence of the regulation of miRNA transcription between human and chimpanzee is an issue of interest and seems not to be addressed yet. In this study, 86 global-run-on-sequencing-validated TSSs of 125 human miRNAs were collected to search their homologous regions in the chimpanzee genome. A homologous TSS location was then confirmed as a putative chimpanzee miRNA TSS if it was able to be identified by a semi-supervised statistical model trained on Pol II ChIP-sequencing data of chimpanzee and several sequence-associated features. In addition, an un-annotated homologous miRNA region was considered to be a putative chimpanzee miRNA if it met some criteria. We then were able to obtain a set of high quality of chimpanzee microRNA TSS and miRNA pairs. From the results, most of the human TSS-miRNA pairs still kept in chimpanzee but a number of pairs were missing. We examined the missing pairs and found that one pair, mir-659 and its TSS, was human-specific, and the others like mir-3199-2, mir-4277, mir-8072, and etc. were lost in chimpanzee during evolution. We further studied the regulatory pathways of these miRNAs. These divergences of the regulation of miRNA transcription might be, at least in part, the molecular footprint of speciation events in the human and chimpanzee lineages.

27C

"Evolutionary Characterization of Molecular Pathogens in Antillean Manatee's (*Trichechus manatus)* "

Maria Rivera Diaz, Ashley Matchett
InterAmerican University of Puerto Rico, Bayamón Campus, (IAUPR-BC), Bayamón, Puerto Rico

Molecular evolutionary pathogen in the skin from Antillean manatee (*T. manatus*) is poorly characterized. *T. manatus* is an autochthonous species and classified endanger species. As an indicator species understanding the Manatee microbiome is key to understanding the developmental association of bacteria with skin and monitoring changes in the local ecosystem. Unusual subcutaneous infection was detected in a local manatee that had not been previously described. It is the purpose of this study to characterize this infection and to determine the causative factors, and whether this is evidence of a human transmitted disease or a new species-specific microbial flora. The aetiology of this infection was determined were isolated, with direct skin swab by techniques of PCR amplification by primer 16S rRNA gene to determine the microbial diversity species and improve phylogenetic resolution within the Enterobacteriaceae. To determine the exact nature and identity of the bacteria recovered, DNA was extracted from the pure cultivars of the viable bacteria and from the original skin swab and sequenced using Next Generation Sequencing (NGS). Overall bacterial population analysis of the recovered DNA indicates in protein homology bacterial that predominant presence of *Gammaproteobacteria*, *Proteobacteria* and *Renikea* and RNA homology *Proteobacteria* presence of *Pseudomonas aeruginosa* and other species and variety. Finally, from the methodological approach, further optimization of the DNA extraction and amplification procedures with analyst primer more specific with OriC-locus, DnaJ (Hsp40), GyrB, RpoB and HSP60, to determine the microbial diversity species and conclude which component of our dualist hypothesis is correct.

28D

A computational and mathematical model of GRN evolution with applications to speciation theory.

Joshua Schiffman, Sergey Nuzhdin, and Peter Ralph
University of Southern California, Los Angeles, CA, USA

An organism's phenotype is constructed by an elaborate process including complex environment-by-gene, and gene-by-gene regulatory interactions. In this poster I introduce a novel computational and mathematical model of evolutionary dynamics that includes mathematical representations of the genome, genetic regulatory network (GRN), and phenotype. This framework enables the study of GRN evolutionary dynamics; specifically the influence of variable molecular and population parameters on developmental systems drift (DSD), GRN rewiring, population reproductive isolation, and speciation. This includes simulation results demonstrating complete reproductive isolation via hybrid incompatibility between two allopatrically isolated populations despite identical initial population genetic makeup, the maintenance of identical selection pressures, and static environmental conditions. A preliminary mathematical analysis of the geometry of the fitness landscape is discussed and future analysis connecting the structure of the neutral fitness manifold to the rate of DSD and reproductive isolation is proposed.

21 Fungal evolutionary genomics: Unravelling mysteries from the Forgotten Kingdom

21.1

Evolutionary genomics of early branches of the fungal tree

Jason Stajich [1], Gerald Benny[2], Mary Berbee[3], Nicolas Corradi[4], Igor Grigoriev[5], Andrii Gryganskyi[6], Timothy James[7], Alan Kuo[5], Kerry O'Donnell[8], Robert Roberson[9], Matthew Smith[2], Joseph Spatafora[10], Thomas Taylor[11], Rytas Vilgalys[12], Merlin White[13]

[1] University of California-Riverside, Riverside, CA, USA, [2] University of Florida, Gainesville, FL, USA, [3] University of British Columbia, Vancouver, BC, Canada, [4] University of Ottawa, Ottawa, ON, Canada, [5] DOE Joint Genome Intsitute, Walnut Creek, CA, USA, [6] Lambert Spawn, Coatesville, PA, USA, [7] University of Michigan, Ann Arbor, MI, USA, [8] USDA ARS NCAUR, Peoria, IL, USA, [9] Arizona State University, Tempe, AZ, USA, [10] Oregon State University, Corvallis, OR, USA, [11] University of Kansas, Lawrence, KS, USA, [12] Duke University, Durham, NC, USA, [13] Boise State University, Boise, ID, USA

The 1000 fungal genomes (1KFG) project and other efforts worldwide have generated a wealth of sequence data supporting studies of evolutionary genomics. Major changes morphology and trophism have occurred in fungal history and some can be traced in the history and gene composition of extant fungi. This includes transitions from a flagellated zoospore life stage seen in chytrid fungi (Cryptomycota, Chytridiomycota and Blastocladiomycota) to the filamentous and yeast forms observed in the 'Zygomycete' and Dikarya lineages. The zygomycete lineages have been classified at times as a single monophyletic group and also split into an unresolved paraphyly. Our phylogenomic analyses provide substantial support for two monophyletic clades one containing Entomophtheromycota, Kickxellomycotina, and Zoopagomycotina (EKZ) and the other comprising Mortierellomycotina, Mucoromycotina, and Glomeromycota (MMG). Comparison of gene content among the species and their outgroups identified zygomycete specific genes and confirmations of gene losses and gains that correspond to the transition from a zoosporic ancestor to primarily filamentous or yeast growth forms. Genome sequences from more than 35 zygomycetes among nearly 300 fungal genomes were used to evaluate the phylogenetic position of these lineages, and compare gene content and phylogenetic relationships among groups of orthologous genes found in animals, zoosporic fungi and the Dikarya fungi. The Zygomycete Genealogy of Life project is building a phylogenomic framework to address genomic and morphological evolution, broad genome and transcriptome sampling of zygomycete lineages including host associated species, incorporating fossil-based dating, imaging of growth forms, and improved descriptions of these fungi in the Encyclopedia of Life.

21.2

History of Lachancea genomes inferred from extant species sequencing and extinct ancestor reconstructions

Véronique Sarilar[2], Guénola Drillon[1], Nikolaos Vakirlis[1], Ingrid Lafontaine[1], Jean-Philippe Meyniel[3], Cécile Neuvéglise[2], Gilles Fischer [1]

[1] Université Pierre et Marie Curie - CNRS, Paris, France, [2] INRA - AgroParisTech, Thiverval-Grignon, France, [3] ISoft, Gif sur Yvette, France

We have sequenced, assembled to one scaffold per chromosome the genomes of 7 *Lachancea* species. Together with the genomes of 3 previously sequenced species, this has provided us with a comprehensive genomic dataset within a monophyletic yeast genus. During this process, we developed a powerful annotation pipeline that defined a comprehensive gene repertoire of more than 50,000 CDS. We identified all gene gains and losses in all the branches of the *Lachancea* phylogenetic tree. Of major interest, we identified the loss of the ZMM genes responsible for the interfering meiotic CO pathway. We also characterized numerous cases of horizontal gene transfers and *de novo* gene creation. Secondly, we developed a new computational method, called AnChro, to reconstruct ancestral genomes. Reconstructed ancestral genomes comprise 8 chromosomes with one centromere per chromosome and on average account for more than 90% of the genes present in extant species. We showed that the rDNA array relocated during the evolution of the *Lachancea* clade probably as a result of two ancestral inversions. The entire dynamics of the sexual loci was recapitulated, revealing that the locus of the silent HMR cassette underwent many translocation, duplication and inversion events while HML remained stably located next to the MAT locus throughout the evolution of the clade. The regional correspondence between modern and ancestral chromosomes revealed that the last common ancestor to all *Lachancea* species had a mating type (MAT)-containing chromosome similar to the actual *S. cerevisiae* chromosome III. Finally, a chromosome fusion was identified in one extant species.

21.3

iGénolevures, a new consortium for yeast comparative genomics

Cécile Fairhead[3,4], Cécile Neuvéglise [1,2], iGénolevures consortium[4]
[1] INRA, Thiverval-Grignon, France, [2] AgroParisTech, Thiverval-Grignon, France, [3] Université Paris-Sud, Orsay, France, [4] CNRS, Orsay, France

Yeasts have played a pioneering role in the birth of comparative genomics. Soon after the genome sequencing of the model yeast *Saccharomyces cerevisiae* in 1996, genomic exploration of other related yeast species launched the development of comparative genomics. With more than a hundred complete genomes from different yeast species, the *Saccharomycotina* subphylum is still one of the eukaryotic groups that comprises the highest number of available genome sequences. However, despite their small and compact genomes, that are much less complex than other fungi, the difficult step of their structural and functional annotation has been underestimated and sometimes neglected. Here, we present the objectives of the new iGénolevures consortium, which bases its studies on expert genome annotation in order to provide reference genomes of high quality for further comparative genomics and population genomics investigations. Resulting studies will be facilitated and can be carried out in integrated topics such as phylogenomics, metabolic reconstruction, or structural evolution, a feature provided by the structural organization of the consortium itself. Indeed, the bioinformatics tools are spread across different websites such as PhylomeDB (http://phylomedb.org/), FUNGIpath (http://embg.igmors.u-psud.fr/fungipath/), and GRYC (http://gryc.inra.fr/). The iGénolevures consortium is also highly involved in non-coding RNA studies and expression of non protein-coding genes.

21.4

Long-read sequencing of Spore killer strains of *Neurospora* reveals highly divergent genomic architectures and a complex evolutionary history

Jesper Svedberg [1], Thomas M. Hammond[2], Hanna Johannesson[1]
[1] Department of Organismal Biology, Uppsala University, Uppsala, Sweden, [2] School of Biological Sciences, Illinois State University, Normal, Illinois, USA

Spore killer is a meiotic drive element found in natural populations of *Neurospora sitophila* and *Neurospora intermedia*. In crosses between strains carrying a *Spore killer* genetic element and strains that are sensitive to *Spore killer*, half of the spores will die and the surviving spores will all show the killing phenotype in further crosses.
Using PacBio SMRT RS II long-read sequencing (to 50x coverage) together with short-read Illumina data we have created high quality, full chromosome genome assemblies of seven strains of *N. sitophila* and *N. intermedia* which are representative of all known Spore killer types as well as of sensitive and resistant strains. This data has revealed a complex pattern of inversions, insertions and deletions, covering the approximately 2 Mbp region of suppressed recombination that surrounds the *Spore killer* genetic element. These structural differences may be instrumental in maintaining linkage between the genes causing the killer phenotype. The killer strains not only differ structurally from sensitive strains, but also from each other, suggesting they have evolved separately through different evolutionary routes even though they are found within the same, or closely related, species and exhibit very similar phenotypes.
Here we present a comparative analysis of the genomic architecture of the regions of suppressed recombination, providing novel insights into the evolution of the Spore killer phenomenon in Neurospora.

21.5

Adaptive genome remodelling by horizontal transfers in cheese-making fungi

Ricardo Rodríguez de la Vega [1,2], Jeanne Ropars[7], Manuela Lopez-Villavicencio[3], Jerome Gouzy[4,5], Erika Sallet[4,5], Emile Dumas[1,2], Sandrine Lacoste[3], Robert Debuchy[6], Antoine Branca[1,2], Tatiana Giraud[1,2]

[1] ESE, Université Paris-Sud, Orsay, France, [2] UMR8079, CNRS, Orsay, France, [3] UMR7205, Muséum National d'Histoire Naturelle, Paris, France, [4] UMR441, Institut National de la Recherche Aagronomique, Toulouse, France, [5] UMR2594, CNRS, Toulouse, France, [6] UMR8621, Université Paris-Sud, Orsay, France, [7] University of Ottawa, Ottawa, Canada

Horizontal gene transfers (HGT) swarm archeal and bacterial genomes, thus underlying many of their evolutionary transformations and adaptations. Although HGT is also recognized as a source of eukaryotic genome innovation, particularly in unicellular organisms such as many members of the fungi, evidence of HGT in eukaryotic genomes is relatively scarce. Here we will show that cheese-making fungi adaptation to the human-made cheese environment is linked to a rapid spread of HGT in two Penicillium species. More precisely, we will show that adaptation to the cheese medium has occurred through parallel gene family expansions associated with multiple recent horizontal transfers of crucial metabolic genes involved in the utilization of the cheese nutrients. The horizontally-transferred regions are flanked by specific transposable elements and display 100% identity in multiple strains and species of cheese-associated Penicillium fungi, indicating recent selective sweeps. The strains carrying these cheese fungus-specific horizontally-transferred regions were experimentally shown to have faster growth and greater competitiveness on cheese. While HGTs events have been reported in fungi, particularly in environments created by humans, the extent and timing of gene transfers observed here are unprecedented. Similarly, while gene duplication may have been involved in the domestication and diversification in crops and yeasts, the extent of gene family expansions and the xenolog origin of many of them appears exceptional. Our results have fundamental implications for understanding the mechanisms underlying rapid adaptation of eukaryotes to new environments and also have potential applications in food industry.

21.6

Population genomics of a highly divers sexual pathogen: the origin and fate of chromosomal structural variants in populations

Daniel Croll, Fanny Hartmann, Marcello Zala
ETH Zurich, Zurich, Switzerland

Fungal pathogen populations show extraordinary evolutionary potential to adapt to changes in the environment, host genotypes or chemical control agents. The main drivers of rapid evolution are frequent sexual reproduction, high dispersal capabilities and extensive standing genetic variation. Despite the ubiquity of evidence for rapid turnover occurring in fungal populations, little is known how the structure of the genome influences the evolution of genetic variation. In particular, chromosomal rearrangements and variations in recombination rates are expected to significantly affect the degree of linkage disequilibrium and, hence, the evolutionary potential of loci. We aimed to comprehensively study the fate of chromosomal structural variants in fungal populations of the ascomycete wheat pathogen *Zymoseptoria tritici*. First, we generated near-complete genome assemblies of four field isolates using a combination of short- (Illumina) and long-read (PacBio) technologies, and high-resolution genetic maps. Comparative analyses of the genome assemblies showed that a substantial proportion of the polymorphism in a field population is due to structural variation in the chromosomal sequences. Second, we used re-sequencing data of 130 additional isolates from a global collection to identify the population-level frequency of structural variants. Third, we calculated genome-wide linkage disequilibria. We found that the location of segregating structural variants correlated with variations in linkage disequilibria but also with independent estimates of recombination rates. The presence of segregating structural variants within populations provide an important genomic context to predict the evolutionary potential of loci in a pathogen genome.

21.7

A new intracellular parasite is a missing link between fungi and microsporidia

Karen Haag [1,2], Dieter Ebert[2], Timothy James[3], Jean-François Pombert[4], Ronny Larsson[5]
[1] Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil, [2] University of Basel, Basel, BS, Switzerland, [3] University of Michigan, Ann Arbor, MI, USA, [4] Illinois Institute of Technology, Chicago, IL, USA, [5] University of Lund, Lund, SE, Sweden

Intracellular obligate parasitism results often in extreme adaptations, whose evolutionary history is difficult to understand, because intermediate forms are hardly ever found. Microsporidia belong to an early-diverging clade of fungi, which evolved extreme physiologic and genomic simplification as well as exceptionally high rates of molecular evolution. They possess the smallest eukaryotic genomes with very few introns, short intergenic regions and bacterial-sized ribosomal genes. As observed in other eukaryotic intracellular parasites, mitochondria in microsporidia have degenerated into small double-layered organelles called mitosomes, which have lost the genome and cannot produce ATP anymore. Instead, they steal it from their hosts. We describe the evolutionary history of a gut parasite of the crustacean *Daphnia* with remarkable morphological similarity to the microsporidia, but genomic features of ancient fungi. This parasite, which we formally name *Mitosporidium daphniae* gen. et sp. nov., possesses mitochondria, genes for oxidative phosphorylation and an infection apparatus typical for microsporidia. Phylogenomics places *M. daphniae* together with the microsporidia in a clade that also includes the most ancient fungi, the Cryptomycota. Comparative genomics further supports the missing link status of *M. daphniae* highlighting both its microsporidian and fungal like characteristics, and reveals the intermediate evolutionary steps that led to extreme metabolic simplification. The new species demonstrates that the extreme reduction in energy metabolism genes as well as the loss of introns in microsporidia was preceded by a reduction in the machinery controlling cell cycle, DNA recombination, repair and gene expression that may have contributed to the characteristically accelerated rate of microsporidia evolution.

21.8

Transposable elements promote adaptive evolution in the genome of the fungal wheat pathogen *Zymoseptoria tritici*

Jonathan Grandaubert [1,2], Julien Y. Dutheil[1], Eva H. Stukenbrock[1,2]
[1] Max Planck Institute for Evolutionary Biology, Ploen, Germany, [2] Christian-Albrechts-Universität, Institute for Botany, Kiel, Germany

Fungal pathogens display a broad diversity of mechanisms enabling them to rapidly evolve under selection pressure and to overcome host defense systems. Unraveling the genetic and genomic basis of these mechanisms is a challenge and requires the integration of functional and evolutionary approaches. Our studies focus on the prominent wheat pathogen *Zymoseptoria tritici* that has co-evolved and spread with its host since wheat domestication more than 10,000 years ago. We applied population genomics approaches to unravel the genetic and genomic basis of adaptive gene evolution in *Z. tritici*. Twelve genomes were sequenced, de novo assembled and aligned to a reference genome. More than 1.5 million SNPs were identified from the 31-Mb genome alignment. The SNP map was used to assess genome wide patterns of linkage disequilibrium and generate a high-resolution recombination map. Gene alignments were analysed with models of sequence evolution in order to identify the patterns and strengths of selection along the genome. In agreement with predictions of background selection, a significant negative correlation was found between the strength of purifying selection and both recombination rate and gene density. Among the 11,839 predicted genes in *Z. tritici*, 875 show signatures of positive selection. Significantly higher *dN/dS* ratio values were found for genes encoding secreted proteins potentially involved in pathogenicity. These genes are located close to transposable element-rich heterochromatic regions suggesting a role of the latter in adaptation of putative pathogenicity-related genes. These results bring new insights in the genomics of rapid adaptive evolution of fungal pathogens.

21.9

Evolutionary analysis and manipulation of natural *Saccharomyces* genomes

Chris Todd Hittinger [1], Quinn Langdon[1], David Peris[1], William G. Alexander[1], Ryan V. Moriarty[1], Kayla Sylvester[1], Diego Libkind[2], Paula Goncalves[3], José Paulo Sampaio[3]
[1] University of Wisconsin-Madison, Madison, WI, USA, [2] CONICET-UNComahue, Bariloche, Argentina, [3] Universidade Nova de Lisboa, Caparica, Portugal

Saccharomyces cerevisiae is one of the premier genetic model systems, but other members of the genus *Saccharomyces* have been characterized minimally. Until recently, most other species have been known primarily through their genetic contributions to industrial interspecies hybrids. Here we focus on the genomic diversity and biogeography of the sister species *Saccharomyces eubayanus* and *Saccharomyces uvarum*. Population and phylogenomic analyses support the existence of at least four diverse sympatric or parapatric populations of these two species in Patagonia, South America. Gene flow is evident between populations of the same species, which are likely in secondary contact, while limited introgression has occurred between species. Rare northern hemisphere strains isolated from the wild are related to the southern hemisphere populations, but they have comparatively low sequence diversity or comprise nearly clonal mosaic lineages. In contrast, isolates from industrial settings are interspecies hybrids used in the production of beer, cider, or wine, suggesting that domestication may favor rare genetic events that are disfavored in nature. Expanding collections of *Saccharomyces* and other diverse yeasts highlight the need for generalized genetic manipulation strategies that work in prototrophic strains. We have developed, and continue to improve, novel cassettes for high-throughput genome engineering that deploy an inducible double-strand break generator and selectable/counterselectable marker. This approach enables pooled allele replacement and mutagenesis with higher efficiencies than those reported for fungal CRISPR/Cas9-based systems. We show its utility across the genus *Saccharomyces* and provide a roadmap for in vivo investigations of gene sequence evolution across fungi.

21.10

Experimental evolution with transgenic *Saccharomyces cerevisiae* strains illustrates massive amplifications of chromosomal segments forming macrotene chromosomes

Bernard Dujon, Varun, Khanna, Agnès Thierry
Unité de Génétique moléculaire des levures (UMR3525 CNRS, UMR927 Univ. P. M. Curie), Institut Pasteur, Paris, France

The importance of chromosomal dynamics in eukaryotic genome evolution is getting an increasing attention with recent results of genomics. Experiments of evolution using yeast, mostly focused on adaptation to limiting environmental conditions, revealed segmental duplications in addition to the more classical sequence alterations. But the spontaneous evolution of imperfect yeast genomes in absence of selection from the environment has received only limited attention, despite its relevance to cancer cell progression. We have examined this question using severely unfit *S. cerevisiae* cells in which essential tRNA synthetase genes (RS) were replaced by their orthologs from the distantly related *Saccharomycotina* species, *Yarrowia lipolytica*. These strains were under severe stress owing to the insufficient charge of their tRNA molecules. In all experiments performed, we found spontaneous mutants with significant growth rate recovery, among which many mutants exhibited massive amplifications of large chromosomal segments, far exceeding the number required for phenotypic restoration when encompassing the transgenic RS gene, or addressing another chromosomal segment not involved in any phenotypic selection. In the latter case, amplification of a 2 kb-long segment in up to 600 copies was found. In the different cases, chromosomes were consequently extended in size from 1.5 – 2 times (*macrotene*) without significantly affecting their stability. The formation of these *macrotene* chromosomes appears a one-step event, often associated with disomy. We will present a model in which accidental template switching in a replication fork, probably triggered by stress, may generate an incontrollable status after interference with a neighboring replicon.

21.11

Insights into the origin and evolution of yeast mating-type switching mechanisms from a chromosomal inversion system in methylotrophs

Sara J. Hanson, Kevin P. Byrne, Kenneth H. Wolfe
Conway Institute and School of Medicine & Medical Science, University College Dublin, Dublin, Ireland

Yeast mating-type switching, in which haploids are able to interconvert between cell types MATa and MATalpha to facilitate mating, is a complex molecular process comprised of several essential components and requiring multiple levels of regulation. The evolution of this process is still largely uncharacterized, with the well-studied systems in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* arising seemingly independently without evidence for a less complex ancestral switching mechanism. Recently, we described mating-type switching in the methylotrophic yeasts *Hansenula polymorpha* and *Pichia pastoris*, where a simpler chromosomal inversion mechanism is used to convert mating-type. In these species, the mating-type (MAT) region contains both MATa and MATalpha genes and is flanked by inverted repeat sequences. One set of MAT genes is silenced by their proximity to a transcriptionally repressed centromere or telomere. Switching occurs by inversion of the MAT region between the repeat sequences resulting in the opposite pattern of MAT gene expression. Mating-type switching, mating, and sporulation in these species are induced by nitrogen limitation, although the molecular components underlying this response are unknown. Using targeted gene deletions, RNA-seq, and bulk segregant analysis, we identify candidate genes involved in the mating-type switching response in methylotrophic yeasts.

505A

Chaos of rearrangements and degeneration of the non-recombining regions in the mating-type chromosomes of the anther-smut fungi

Tatiana Giraud [1,2], Helene Badouin[1,2], Eric Fontanillas[1], Michael Hood[3], Jerome Gouzy[4], Gabriela Aguileta[1,2], Elsa Petit[1,3], Michael H. Perlin[5], Christina Cuomo[6]
[1] Universite Paris Sud, Orsay, France, [2] CNRS, Orsay, France, [3] Amherst College, Amherst, France, [4] INRA, Toulouse, France, [5] Louiseville University, Louiseville, USA, [6] Broad Institute, Cambridge, USA

Dimorphic mating-type chromosomes in fungi are excellent models for understanding the genomic consequences of recombination suppression on adaptation. Their suppressed recombination and reduced effective population size are expected to limit the efficacy of natural selection, leading to genomic degeneration. We identified and assembled the whole sequences of the mating-type chromosomes (a1 and a2) of the anther smut fungi to investigate degeneration in their non-recombining regions. This revealed unprecedented details of structural rearrangements and convergence between a mating-type chromosome system and sex chromosomes, with extensive and ancient recombination suppression, extreme levels of transposable element accumulation, gene losses, and heterogeneity in allelic divergence levels between mating types. Genome sequence data was also obtained for twelve other Microbotryum species. We found strong evidence of degeneration across the genus in the non-recombining regions of the mating-type chromosomes, with significantly higher rates of non-synonymous substitution (dN/dS) than in non-mating-type chromosomes or in recombining regions of the mating-type chromosomes. The levels of degeneration did not differ between the a1 and a2 mating-type chromosomes, consistent with the lack of homogametic/heterogametic asymmetry between them, and contrasting with X/Y or Z/W sex chromosomes.

506B

The population genomics of adaptation in plant pathogens: the case of the anther smut fungi

Hélène Badouin [1,2], Gabriela Aguileta[1,2], Ricardo Rodriguez de la Vega[1,2], Jérôme Gouzy[3,1], Michael Hood[6], Sophie Siguenza[3,1], Alodie Snirc[1,2], Stéphanie Le Prieur[1,2], Zehua Chen[4], Su San Toh[5], Christina A. Cuomo[4], Michael H. Perlin[5], Antoine Branca[1,2], Tatiana Giraud[1,2]
[1] CNRS, Orsay, France, [2] Université Paris Sud, Orsay, France, [3] INRA, Toulouse, France, [4] Broad Institute of MIT and Harvard, Cambridge, USA, [5] University of Louisville, Department of Biology, Louisville, USA, [6] Amherst College, Amherst, USA

In the study of eukaryotic adaptive divergence, fungi present many experimental advantages, such as small genomes and the availability of many complexes of sibling species adapted to different hosts or habitats. The Microbotryum fungi, with dozens of host-specific sibling species causing anther smut disease in Caryophyllaceae family are particularly good models for addressing the question of the genomic processes of host adaptation. We have analysed dozens of genomes of each of two sister species, M. lychnidis-dioicae and M. silenes-dioicae, parasitizing S. latifolia and S. dioica, respectively, as well as a genome of each of ten other Microbotryum species specialized on other Silene, Dianthus or Saponaria plants. We have investigated signatures of selection both within and between species for getting insights into the genomic processes of host adaptation and coevolution. In particular, we have looked for footprints of recent selective sweeps within fungal species, and of gene gains and losses between species, as well as signatures of positive selection in the form of high rates of non-synonymous substitution. In planta expression have been used for validating the function of candidate genomic regions. Altogether, this study sheds light on the genomic processes of adaptive divergence in a model complex of fungal species, which has both fundamental and applied importance, being a model for understanding adaptation in crop fungal pathogens.

507C

Inferring evolutionary dynamics from genome variation patterns in *Saccharomyces cerevisiae* using the eukaryote pan-genome toolset

Giltae Song [1], Benjamin Dickins[2], Barbara Dunn[1], J. Michael Cherry[1]
[1] *Stanford University, Stanford, CA, USA,* [2] *Nottingham Trent University, Nottingham, UK*

Despite the release of a hundred *S. cerevisiae* strain sequences, comparative evolutionary analyses have been limited by a reliance on the single reference strain genome. Existing prokaryotic pan-genome analysis tools are difficult to apply to yeast owing to the complexity of gene structures in eukaryotes. We have developed a eukaryote pan-genome analysis pipeline that is freely accessible, and applied it to analyze 120 *S. cerevisiae* strains. Using our pipeline and these strain genomes, all genes can be classified into one of three categories of distribution within the species; core (always present), dispensable (in some organisms, but not all), and unique. We have investigated two aspects of variation within the pan-genome: (1) variation in the core genome, (2) patterns of absence or presence of genes in the dispensable and unique genomes. We have explored this genetic variation with a phylogenetic tree based on Single Nucleotide Polymorphisms (SNPs) in the core genome. An analysis of population structure based on SNPs in the core genome has also been carried out. We have investigated the dynamics of horizontal gene transfer, recombination, or gene gain and loss based on contrasting patterns in SNP-based phylogeny and gene-based clustering, combined with analysis of population structure. We also characterize, for a particular strain, whether genes not part of the core genome were part to an ancestor genome (or duplicated subject to divergent evolution), or whether they are horizontally acquired “novel” genes obtained, for example, by mating with another *S. cerevisiae* strain or through interspecies hybridization.

508D

Genomic footprints of adaptation in the blue cheese-making fungus *Penicillium roqueforti*

Emilie Dumas [1,2], Jeanne Ropars[1,2], Ricardo C Rodríguez de la Vega[1,2], Manuela López-Villavicencio[3], Jérôme Gouzy[4,5], Joëlle Dupont[3], Alodie Snirc[1,2], Stéphanie Le Prieur[1,2], Tatiana Giraud[1,2], Antoine Branca[1,2]

[1] Univ Paris -Sud, Ecologie, Systématique et Evolution, UMR8079, Orsay, France, [2] CNRS, Ecologie, Systématique et Evolution, UMR8079, Orsay, France, [3] Institut de Systématique, Evolution, Biodiversité, UMR 7205 CNRS-MNHN-UPMC-EPHE, Paris, France, [4] INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, Castanet-Tolosan, France, [5] CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, Castanet-Tolosan, France

Humans have been making cheeses for more than seven thousands years. Many bacteria and fungi coexist in the cheese ecosystem and they have evolved under strong human selection for producing the variety of flavors, textures and aspects of cheeses. The domestication process of these microorganisms make them ideal models for studying several aspects of adaptation. The comparison of several *Penicillium* genomes have revealed that cheese-making *Penicillium* have adapted to the cheese medium through multiple recent horizontal transfers (HGTs) of crucial metabolic genes. *Penicillium roqueforti*, used for the production of blue veined-cheeses, occurs in various habitats, i.e. cheese and non-cheese and displays substantial genetic diversity. In addition, some strains used for cheese production do not bear the HGTs. Here, we conducted a population genomics analysis focused on *P. roqueforti* to look at the genetic mechanisms other than HGTs underlying adaptation of *P. roqueforti* to its environment. We used genomic scans to look at pattern of diversity, recombinaison and selection across the genome. We identified at least five distinct genetic groups among the strains different from the ones previously described by microsatellites. Interestingly, preliminary results show a very dynamic pattern of gene gain and loss, each distinct genetic group presenting hundreds of highly specific gene sets. These results allow getting important insights into the process of genome adaptation in fungi.

509A

A dynamic transposon family in the yeast mitochondrial genome

Baojun Wu, Weilong Hao
Wayne State University, Detroit, MI, USA

Transposable elements (TEs) are an important factor shaping eukaryotic genomes. Although a significant body of research has been conducted on the abundance of TEs in nuclear genomes, TEs in mitochondrial genomes remain elusive. In this study, we took advantage of the abundant yeast population genomics data to determine TEs and their propensity in complete yeast mitochondrial genomes. We have observed compelling evidence of TEs propagating within the mitochondrial genome and being horizontally transferred between species. Mitochondrial TEs experience rapid diversification by nucleotide substitution, and more importantly, undergo dynamic merger and shuffling to form new TEs. Given the hyper mobile and transformable nature of mitochondrial TEs, our findings open the door to a deeper understanding of eukaryotic mitochondrial genome evolution on both genome dynamics and the origin of genome architecture.

510B

Estimating driving factors for codon usage bias by decomposing intra-genomic GC-content heterogeneity

Cedric Landerer, Michael Gilchrist
University of Tennessee, Knoxville, Knoxville, TN, USA

The usage of synonymous codons differs between organisms and even within a single genome. One explanation for the heterogeneity of codon usage bias (CUB) is varying GC-content but this assumes very weak selection on CUB. Another explanation is selection so that CUB and bias in the tRNA pool match, increasing translation efficiency for highly expressed genes.
The yeast *Lachancea kluyveri* is an example of an organism having a remarkable shift in GC-content: one contiguous ~1 Mb region of the genome has a GC-content of ~53% versus ~40% in the rest of the genome.
When ignoring this heterogeneity, existing models have to fail since the data is of contradictory nature. We hypothesized that mutational bias towards GC-rich codons is responsible for the shift in GC-bias, rather than differing selection pressures on translation efficiency.
We applied our mechanistic model for CUB that estimates selection against ribosome pausing in high expression genes, expression level, and mutational biases to deconstruct the contribution of selection and mutation to the evolution of CUB of individual genes.
Our results show that the whole genome experiences the same selection environment but mutation bias appears to be driving this significant intra-genomic shift. This work provides insights into the role of mutation bias on GC-content and provides a template for detecting differences in mutation and selection environments across tissues and organisms.

511C

Differences in gene expression between dried and wetted thalli of *Usnea* lichens.

Mieko Kono [1], Yoshihito Ohmura[2], Yoko Satta[1]

[1] The Graduated University for Advanced Studies (SOKENDAI), Kanagawa, Japan, [2] National Museum of Nature and Science, Tsukuba, Japan

Lichens are symbiotic associations of a fungus and photosynthetic partners (alga and/or cyanobacterium). They are often adapted to environments with cycles of extreme drying and transient wetting. Their ability to tolerate desiccation and revive metabolic activities immediately after rehydration is considered to have enabled their survival in extreme environment. Although constitutive protection mechanisms have been suggested, genetic backgrounds of their adaptation remain largely unknown. In the present study, transcriptomes of dried and wetted lichen thalli were examined for two *Usnea* species, *U. bismolliuscula* and *U. hakonensis* using the next-generation sequencing. In addition, the expression level of two enzymatic genes in the pentose phosphate pathway was quantified by quantitative PCR. These results revealed that a number of genes are constantly expressed in diverse biological processes irrespective of water conditions, supporting the hypothesis of constitutive protection mechanisms. However there exist genes that are specifically expressed in the dried lichens and involved in desiccation-tolerance. Taken together, it is concluded that not only constitutive but also inducible protection mechanisms against desiccation are developed at the transcription level for lichens to confront environmental challenges in their natural habitats.

512D

Hitchhiking DNA in Magnaporthe oryzae.

Ludovic Mallet [1,2], Cyprien Guérin[1], Joelle Amselem[3,4], Elisabeth Fournier[5], Hélène Chiapello[1,6]
[1] INRA, UR MIG, 78352 Jouy-en-Josas, France, [2] Evolutionary Bioinformatics, IEB, University of Muenster, Germany, [3] INRA, URGI, 78026 Versailles, France, [4] INRA, UMR BIOGER, 78850 Thiverval-Grignon, France, [5] INRA, UMR BGPI, 34398 Montpellier, France, [6] INRA, UR MIA-T, 31 326 Castanet-Tolosan, France

Magnaporthe oryzae is a successful pathogen of crop plants and a threat for food production world-wide. This species gathers pathogens of different Poaceae including rice and wheat and causes the main fungal disease of rice worldwide. The Evolutionary Genomics of Magnaporthe oryzae (GEMO) project is an attempt to identify the genomic determinants and evolutionary events involved in pathogenesis, host specificity and adaptation.

We have analyzed and compared a dataset of ten closely related genomes of the Magnaporthe oryzae/grisea species complex selected for their different main host and host range. We put emphasis on the horizontally acquired material that we predicted with a parametric detection method based on tetranucleotide signature.

We outline the general content of the predicted transferred regions and propose for some candidates the likely taxonomy of their potential donors. We depicted and compared the functional profiles of the host and acquired genes and investigated the intermingling of horizontal transfers with our de novo prediction of transposable elements as a potential adaptative and evolutionary feed. First results pointed out a few large transferred regions potentially acquired from distant species identified by alignment-free methods, including several plant-pathogen fungi.

These candidates called for further research around their potential contribution to phenotype, in a step to open the general study of the yet unresolved evolutionary tangram of the Magnaporthe oryzae pathogenesis.

513A

Genome variation across global isolates in the emerging fungal pathogen *Candida glabrata*

Laia Carreté [1,2], Cécile Fairhead[3], Toni Gabaldón[1,4]

[1] Centre for Genomic Regulation, Barcelona, Spain, [2] Universitat Pompeu Fabra, Barcelona, Spain, [3] Institut de Génétique et Microbiologie, Paris, France, [4] Institució Catalana d'Estudis Avançats, Barcelona, Spain

Candida glabrata is one of the most common pathogenic fungi in humans, ranking as the second causative agent of candidiasis worldwide. Despite its name is distantly related to the model pathogen *Candida albicans* and belongs to the Nakaseomyces clade,more closely related to Saccharomyces cerevisiae (Gabaldón et al., 2013). This indicates that virulence to humans has independently and recently emerged within this clade. Considering that virulence properties can vary significantly among strains of the same species, it is important to study the detailed genetic background of pathogenic and commensal isolates.

Here, we use a genome re-sequencing approach to analyse the variability among 32 different genomes from clinical and commensal *Candida glabrata* samples sampled from different countries. We did a computational analysis for detecting single-nucleotide polymorphism, ploidy, copy number variation and genomic re-arrangements. The sequenced strains are structured in six differentiated clusters, which do not cluster by geographical origin or site of infection. Despite an overall high similarity at the sequence level, most differences between strains consist of gene losses and gains, often involving cell-wall proteins. This is consistent with observed phenotypic variability among *C. glabrata* isolates in terms of their adherence and virulence properties. We find evidence for active recombination between distinct subpopulations, which is remarkable for species considered as asexual. In accordance we find evidence for active mating type switching in four strains.

Gabaldón et al.: Comparative genomics of emerging pathogens in the *Candida glabrata* clade (2013). BMC Genomics 14:623.

514B

Potential and pitfalls of eukaryotic metagenome skimming: A test case for lichens

Bastian Greshake [1], Simonida Zehr[1], Francesco Dal Grande[2], Imke Schmitt[2], Ingo Ebersberger[1]
[1] Department for Applied Bioinformatics, Institute for Cell Biology and Neuroscience, Goethe University, Frankfurt am Main, Germany, [2] Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany

Whole genome shotgun sequencing of multi species communities using only a single library layout is commonly used to assess taxonomic and functional complexities of large and diverse microbial communities. Here we investigate to what extent such metagenome skimming approaches are applicable for in-depth genomic characterizations of obligate symbiotic communities involving eukaryotes, e.g. lichens. Taking the sequencing study design into account, we address how to best assemble eukaryotic metagenome skimming data, what pitfalls can occur, and what genome quality can be expected from the data. To facilitate a project specific benchmarking, we introduce the concept of twin sets. These are simulated data resembling the outcome of a particular metagenome skimming study. We show that the quality of genome reconstructions from metagenome skimming data depends essentially on assembler choice. Individual tools, including the metagenome assemblers *Omega* and *MetaVelvet*, are surprisingly sensitive to low and uneven coverages. In combination with the common practice of empirical assembly parameter choice to maximize the assembly N50 value, their performance drop can culminate in precluding an entire genome from the assembly. In our benchmark setting, *MIRA*, an all-purpose overlap assembler consistently is least affected by varying coverage ratios for the sequenced organisms. The resulting genome reconstructions facilitate the identification of almost all genes annotated in the original data rendering them, in principle, suitable for high-resolution comparative genomics studies. Reconciling this expectation with the outcome of a real-world metagenome skimming of the lichen *Lasallia pustulata* indicates methodological problems causing the underrepresentation of one symbiont in the shotgun library.

515C

Exploring the unknown side of the fungal gene repertoire

Betina M. Porcel [1], Laura Do Souto[1], Benjamin Noel[2], Pascal Bento[2], Corinne Da Silva[2], Jean-Marc Aury[2], Arnaud Couloux[2], Simon Duprat[2], Eric Pelletier[1], Patrick Wincker[1]
[1] Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, Laboratoire d'Analyse Génomique des Eucaryotes (LAGE) ; Université d'Evry, UMR 8030; Centre National de Recherche Scient, Evry, France, [2] Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, Laboratoire d'Informatique Scientifique (LIS), Evry, France

The rapid pace of the genome sequencing field has provided an enormous amount of information on the gene content of many organisms. Genome sequencing allows the establishment of the complete inventory of genetic elements, particularly protein-coding genes, from which every organism performs its biological functions. Advances in sequencing, assembly and syntactic annotation methods have been fast, leading to the generation of a near-complete gene catalog for many species. However, a large fraction of the genes identified in any new eukaryotic genome sequence bear neither similarity to known sequences nor close relatives in databases. Interpretation of a genome sequence is seriously hampered by such lack of knowledge, which frequently affects 10-30% of the gene predictions in any newly sequenced eukaryote genome and even a larger set in poorly sampled taxa. Their absence of homology suggests that many of these genes are not part of the basic cell machinery, but may contribute to the specific functions performed by any eukaryotic single species.

The fungal kingdom encompasses an enormous diversity of taxa with varied ecologies, life cycle strategies, and morphologies from unicellular yeasts, to highly organized multicellular structures, such as mushrooms. The Dikaryome consortium explores uncharted areas of this group aiming to explain the relationship between these life style differences and their evolutionary content and history. The characterization of the "genetic signatures" of certain branches of the fungal tree of life through bioinformatics analyses will be presented.

516D

Next-generation RAD sequencing: a tool for evaluation the phylogeographic patterns of *Amanita ponderosa* (Malençon & R. Heim) in Iberian Peninsula

Ricardo Arraiano Castilho [1], Ismail K. Saglam[2], Michael R. Miller[3], Albano Beja-Pereira[1,4]
[1] CIBIO-InBIO, Porto, Portugal, [2] Faculty of Science, Department of Biology, Ancara, Turkey, [3] University of California - Department of Animal Science, Davis, USA, [4] Faculty of Sciences, University of Porto, Porto, Portugal

Ectomycorrhizal fungi are one of the most important constituents of forest ecosystems. Besides this little is known about the structure and organization of these organisms in this complex system. Here we investigate the phylogeographic patterns of one of the most charismatic wild and edible mushroom species. *Amanita ponderosa* occurs in southwest of Iberian Peninsula, some regions of North Africa and west coast Italy. Ectomicorhyzical of cork and holm oaks preferentially, it is very typical in Montado ecosystems characterized by agroforestry areas in open woodlands. Here we sequenced the partial genome of 206 individuals across all the distribution area through the Restriction-site Association DNA sequencing (RADseq) using a low frequent cutter restriction enzyme (*SbfI*) to generate a panel of genetic markers (SNP's). The markers were called and analysis conducted using statistical framework estimators from the site frequency spectrum. We clearly detected a longitudinal gradient of clustering populations in the Iberian Peninsula. These findings reveal a great applicability for RADseq in wild and non-model species, and promises to become an important technology for ecological population genomics studies.

517A

Genome sequencing and phylogenomics in ambrosia beetle fungal cultivars.

Dan Vanderpool, Ryan Bracewell, John McCutcheon
University of Montana, Missoula, MT, USA

Fungal farming by insects evolved independently in three insect orders, with the Attine ants and Macrotermite systems receiving the majority of the research attention. Genome sequencing of the ant and termite host and their respective fungal cultivars provided fascinating details on behavioral aspects and biochemical mechanisms of the symbiosis. The remaining insect-fungal agricultural symbiosis, that of the ambrosia beetles from the family Curculionidae, represents the most diverse assemblage of both insects and cultivars. Within this group agriculture evolved up to 11 times with two beetle subfamilies domesticating diverse fungal lineages. We sequenced the genomes of six different ambrosia beetle fungi in the order Ophiostomatales (Ascomycota, Pezizomycotina, Sordariomycetes). The resulting phylogenomic analysis of the cultivar group and previously sequenced genomes from other beetle associates confidently resolves historically unstable relationships within Ophiostomatales. In addition we confirm polyphyly of the ambrosial genus *Raffaelea,* indicating multiple transitions within Ophiostomatales from a beetle associated lifestyle to that of a more intimate beetle-fungus agricultural symbiosis.

22 Adaptive and non-adaptive evolution of gene expression and regulation

22.1

Differential adaptation to lakes and rivers: insights from transcriptome analyses in sticklebacks

Yun Huang [1], Philine Feulner[2], Frédéric Chain[3], Mahesh Panchal[1], Christophe Eizaguirre[4], Tobias Lenz[1], Irene Samonte[1], Martin Kalbe[1], Monika Stoll[5], Erich Bornberg-Bauer[6], Thorsten Reusch[7], Manfred Milinski[1]

[1] Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Ploen, Germany, [2] Department of Fish Ecology and Evolution, Eawag, Kastanienbaum, Switzerland, [3] Department of Biology, McGill University, Montreal, Canada, [4] School of Biological and Chemical Sciences, Queen Mary University of London, London, UK, [5] Genetic Epidemiology, Westfälische Wilhelms University, Münster, Germany, [6] Institute for Evolution and Biodiversity, Evolutionary Bioinformatics, Westfälische Wilhelms University, Münster, Germany, [7] Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

The three-spined stickleback (*Gasterosteus aculeatus*) has repeatedly colonized and adapted to diverse freshwater habitats since the last glaciation. Amongst the freshwater habitats, adjacent lake and river populations harbour distinct parasite communities, a recurring ecological difference proposed to drive adaptive differentiation between lake and river stickleback ecotypes. To study the rapid adaptation of sticklebacks to lake and river habitats, we analysed gene expression profiles of 77 whole-transcriptome libraries of two immune-relevant tissues, the head kidney and the spleen, obtained from wild-caught sticklebacks of the two habitat types across multiple geographic locations. Differential expression analyses identified 189 genes showing statistically significant habitat-specific expression patterns among the three European locations. Amongst these genes, 15 genes are annotated with putative immune functions, and 50 have been experimentally associated with immune-responses in sticklebacks, reinforcing the hypothesis that parasites contribute to adaptive evolution of lake and river sticklebacks. Furthermore, using genome sequences of the same individuals, we aim to associate transcriptional variation with genomic variation, i.e. single nucleotide polymorphism and copy number variation. Integrating genomic with transcriptomic data from replicated lake-river population pairs provides an excellent opportunity to investigate the adaptive evolution of gene expression and the underlying genetic basis.

22.2

The navigability of more than 1000 empirical adaptive landscapes of transcription regulation

José Aguilar-Rodríguez [1,2], Joshua L. Payne[1,2], Andreas Wagner[1,2]

[1] University of Zurich, Institute of Evolutionary Biology and Environmental Sciences, Zurich, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3] The Santa Fe Institute, Santa Fe, New Mexico, USA

The adaptive landscape is a central concept in evolutionary theory since the modern synthesis. Recent advances in high-throughput technologies have provided us with a first glimpse at the structure and navigability of a few incomplete empirical landscapes. Here, we have studied 1128 complete adaptive landscapes. Each landscape describes the binding specificity of a transcription factor to all possible short DNA sequences. The navigability of such landscapes via mutation and selection has important implications for evolution, as a transcription factor's specificity for its cognate DNA site determines its influence on gene expression. We find that the adaptive landscapes of transcription factor binding specificities are far less rugged than randomized landscapes and only slightly more than additive landscapes. They contain few peaks that are accessible from throughout the landscape. These peaks do not usually comprise single sites but broad plateaus that can contain dozens to hundreds of sites, indicating the robustness of high affinity binding to mutations. The binding sites found in more accessible and robust peaks are enriched in protein-bound regions of the mouse genome. Our findings suggest that the navigability of transcription factor binding affinity landscapes and the robustness of their peaks may have contributed to the enormous success of transcriptional regulation as a means to achieve novel phenotypes.

22.3

Dynamics of Transcription Factor Binding Site Evolution

Murat Tugrul, Tiago Paixao, Gasper Tkacik, Nick Barton
IST Austria, Klosterneuburg, Austria

Sequence specific binding of transcription factors (TFs) to DNA is an important mechanism determining gene expression levels. We use a biophysical model of directional selection on gene expression to estimate the rates of gain and loss of transcription factor binding sites (TFBS), under both point mutations and indels. Rates of gain and loss of a single site are typically slow. Gain of a single TFBS is extremely unlikely under neutral evolution or weak selection: strong selection in a large population is needed for rapid evolution of a site, and even then, the initial sequence must be close to a functional sequence. The stationary distribution of binding sequences is also typically approached very slowly. These results for a single TFBS suggest that evolution of longer sequences (i.e. enhancer and promoter) are important to make TFBS evolution possible over realistic time scales, making the length of the available functional regulatory DNA another important parameter in TF binding evolution. This is confirmed by simulations, which also suggest that both cooperativity and the ghosts of previously functional sites promote faster TFBS evolution.

22.4

Conserved microRNA editing in mammalian evolution, development and disease

Maria Warnefors [1,2], Angélica Liechti[1], Jean Halbert[1], Delphine Valloton[1], Henrik Kaessmann[1,2]
[1] Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland

Mammalian microRNAs (miRNAs) are sometimes subject to adenosine-to-inosine RNA editing, which can lead to dramatic changes in miRNA target specificity or expression levels. However, although a few miRNAs are known to be edited at identical positions in human and mouse, the evolution of miRNA editing has not been investigated in detail. In this study, we identify conserved miRNA editing events in a range of mammalian and non-mammalian species.

We demonstrate deep conservation of several site-specific miRNA editing events, including two that date back to the common ancestor of mammals and bony fishes some 450 million years ago. We also find evidence of a recent expansion of an edited miRNA family in placental mammals and show that editing of these miRNAs is associated with changes in target mRNA expression during primate development and aging. While global patterns of miRNA editing tend to be conserved across species, we observe substantial variation in editing frequencies depending on tissue, age and disease state: editing is more frequent in neural tissues compared to heart, kidney and testis; in older compared to younger individuals; and in samples from healthy tissues compared to tumors, which together suggests that miRNA editing might be associated with a reduced rate of cell proliferation.

Our results show that site-specific miRNA editing is an evolutionarily conserved mechanism, which increases the functional diversity of mammalian miRNA transcriptomes. Furthermore, we find that although miRNA editing is rare compared to editing of long RNAs, miRNAs are greatly overrepresented among conserved editing targets.

22.5

Evolution of microRNA across primates

Jennifer McCreight, Willie Swanson
University of Washington, Seattle, WA, USA

MicroRNAs are short, noncoding, single-stranded RNAs involved in post-transcriptional regulation in eukaryotes. The mature miRNA has a 7nt "seed region" that complementary base-pairs with the 3' UTR of mRNA, resulting in the degradation and therefore down-regulation of its targets. A single type of miRNA can have one to thousands of mRNA targets, establishing the potential for small changes in miRNA to have profound phenotypic effects.

Only four non-human primates have experimentally validated miRNA, and the number validated per species is a fraction of those in humans. In order to study how miRNA evolved across the primate lineage, I performed microRNAseq on fibroblast cell cultures from thirteen primate species with available genomes (human, chimpanzee, bonobo, gorilla, orangutan, gibbon, rhesus macaque, baboon, squirrel monkey, marmoset, galago, aye-aye, and mouse lemur). miRDeep2 identified hundreds of known and novel miRNAs. I then used baseml to investigate if any miRNA are conserved or evolving rapidly across the primate lineage, and also searched for mutations occurring in the functionally significant seed region. By predicting precursor miRNA secondary structure with the ViennaRNA package, I found divergent miRNA with striking structural differences, as well as some compensatory mutations that appear to conserve structure. In order to discover any miRNA that were significantly conserved or divergent in humans, I compared these findings to miRNA from the Broad Institute’s ExAc database of 61,486 humans. Because of our large sample population, we have identified more miRNA variation within humans than previous studies have suggested.

22.6

The target repertoire of each miRNA increases during evolution

Masafumi Nozawa [1,2]

[1] *National Institute of Genetics, Mishima, Japan,* [2] *SOKENDAI, Mishima, Japan*

MicroRNAs (miRNAs) play essential roles in gene regulatory networks. Recent advancements in sequencing technologies enabled the identification of miRNAs in many organisms and the elucidation of their origin and evolution. However, the mechanism by which miRNAs are integrated into existing gene regulatory networks during evolution remains elusive. Here, I investigated the evolution of three *Drosophila* miRNAs, namely, miR277, miR982, and miR954, which are expected to be at different integration stages because of their age differences. A combined approach of experiments and bioinformatics revealed that, in *D. melanogaster*, the oldest miRNA examined (miR277), which was generated before *Drosophila* radiation, has the largest number of target genes, whereas the youngest miRNA (miR954) possesses the smallest number. Analyzing other species, I found that only 57% of the miR277-target pairs existed in the common ancestor of *Drosophila*, and the remaining 43% were newly generated in the lineage to *D. melanogaster*. A number of miR277-target pairs were also lost during evolution, but old pairs were less likely to be lost than the expectation under random loss. Based on these results, I propose that miRNAs continuously acquire targets during evolution, but many of these pairs are selectively neutral, and their partnerships are dissolved by mutation and genetic drift. By contrast, a few miRNA-target pairs are biologically meaningful and maintained under purifying selection. Continuation of this process results in an increase in the number of target genes of a miRNA over time, which enables the miRNA to be integrated into the regulatory network during long-term evolution.

22.7

Genome-wide inference of natural selection on regulatory sequences in the human genome

Brad Gulko[1], Ilan Gronau[2], Leonardo Arbiza[4], Melissa J. Hubisz[4], Adam Siepel [3]
[1] Graduate Field of Computer Science, Cornell University, Ithaca, New York, USA, [2] Efi Arazi School for Computer Science, Interdisciplinary Center (IDC) Herzliya, Herzliya, Israel, [3] Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, USA, [4] Department of Biological Statistics & Computational Biology, Cornell University, Ithaca, New York, USA

For decades, it has been hypothesized that gene regulation has played a central role in human evolution, yet much remains unknown about the genome-wide impact of regulatory mutations. I will describe a recent project in which we used whole-genome sequences and genome-wide chromatin immunoprecipitation and sequencing data to demonstrate that natural selection has profoundly influenced human transcription factor binding sites since the divergence of humans from chimpanzees 4-6 million years ago. Our analysis uses a new probabilistic method, called INSIGHT, for measuring the influence of selection on collections of short, interspersed noncoding elements. We find that, on average, transcription factor binding sites have experienced somewhat weaker selection than protein-coding genes. However, the binding sites of several transcription factors show clear evidence of adaptation. Overall, regulatory elements seem to contribute substantially to both adaptive substitutions and deleterious polymorphisms with key implications for human evolution and disease.

I will also describe a novel computational method for estimating the probability that a point mutation at each position in a genome will influence fitness. These fitness consequence (fitCons) scores serve as evolution-based measures of potential genomic function. We have generated fitCons scores for three human cell types based on public data from ENCODE. Compared with conventional conservation scores, fitCons scores show considerably improved prediction power for cis-regulatory elements. In addition, they indicate that 4.2–7.5% of nucleotides in the human genome have influenced fitness since the human-chimpanzee divergence, and they suggest only modest impact from recent evolutionary turnover on the functional content of the genome.

22.8

Polygenic Adaptation in Recent Human History

Yair Field [1,2], Jonathan K. Pritchard[1,2]

[1] Department of Genetics, Stanford University, Stanford, CA 94305, USA, [2] Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

The extent and nature of adaptive evolution in recent human history is still poorly understood. Human population genetic research at the modern genomic era has focused on, and established much about, strong selective sweeps. Yet, our understanding of alternative, and more moderate, forms of selection, is still in its infancy. Recently, analyses of intra-population variation have been used to search for soft sweeps and polygenic adaptation, but only little has been thus found. A remarkable exception being, the compelling evidence for recent polygenic adaptation for increased height in northern Europe. We hypothesized that the comparative nature of these methods might have left much signal of population-shared polygenic adaptation to be found. We thus developed a novel non-comparative method to detect signals of recent adaptation, which is based on modeling the distribution of rare variants within a large whole genome sequencing dataset. Simulations show, that our method can be useful to detect a wide range of recent evolutionary events, including classic sweeps, soft sweeps and weak polygenic adaptation. Applying the method to ~1,600 individuals from the British ALSPAC cohort, we find reassuring evidence for the little known selection benchmarks, including a strong selection at the Lactase gene, and a polygenic adaptation for tall alleles. Notably, we find evidence for additional widespread polygenic adaptation, acting on both coding and regulatory regions, with a marked enrichment for immune functions. Our work thus shed new light on the evolution of current genetic variation in humans.

22.9

Functional characterization of an adaptive cis-regulatory polymorphism in Drosophila melanogaster

Amanda Glaser-Schmitt, John Parsch
LMU-Munich, Munich, Germany

Drosophila melanogaster, which presently has a worldwide distribution, expanded from its ancestral home range in sub-Saharan Africa within the past 15,000 years. This expansion into new habitats is thought to have been accompanied by extensive adaptation. However, the genes involved in this adaptation are almost completely unknown. Transcriptomic studies found that the gene CG9509, which encodes a choline dehydrogenase, consistently shows higher expression in cosmopolitan populations than in sub-Saharan populations. The expression difference is caused by sequence variation in an upstream enhancer that shows population genetic evidence for a selective sweep in cosmopolitan populations. Detailed molecular genetic analyses have revealed three SNPs that contribute to the expression difference. Interestingly, two SNPs with moderate effects on expression are fixed in cosmopolitan populations, while a third SNP, which has a large effect on expression, is at intermediate frequency. Using RNAi knockdown and a null mutant, we show that high expression of CG9509 is associated with diminished larval growth rate, smaller adult body size, and reduced wing load - phenotypic traits that may be advantageous in temperate environments. Our results demonstrate how large-scale transcriptomics can be combined with population genetic, molecular genetic, and phenotypic studies to characterize the functional effects of regulatory changes on gene expression and organismal traits.

22.10

Towards an understanding of the genetic basis of phenotypic change - changes in the expression of a microRNA underlie morphological change in Drosophila melanogaster

Sebastian Kittelmann [1], Saad Arif[2], Isabel Almudi[1], Maria Daniela Santos Nunes[1], Alistair Peter McGregor[1]

[1] Oxford Brookes University, Oxford, UK, [2] Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany

Trichomes are short non-sensory actin protrusions found on insect bodies. The transcription factor Shavenbaby is both necessary and sufficient for trichome development. It acts by integrating several upstream patterning pathways to activate a battery of downstream targets, many of which are directly involved in cell shape changes. One of these downstream genes is shavenoid the product of which is translationally repressed by miR-92a which thereby represses trichome formation. Different strains of Drosophila melanogaster exhibit differently sized trichome-free areas on the second femur, a quantitative trait referred to as the ‘naked valley'. Naked valley size is largely dependent on a 20 kb region on the third chromosome containing miR-92a. The sequence of miR-92a is similar between strains with different naked valley size, but miR-92a is differently expressed between these strains. Most likely at least one cis-regulatory element (CRE) at the locus controlling expression of miR-92a and possibly of its host gene jing interacting gene regulatory 1 has evolved between strains of D. melanogaster leading to differential expression of miR-92a. Several putative enhancer elements from the locus have the ability to drive GFP expression in the developing leg which makes them good candidates for the evolved CRE. Currently we are analysing the cis-regulatory logic of the locus in order to shed light on the evolution of morphological traits by changes in gene regulation in general and in the regulation of a the expression of a microRNA in particular.

22.11

Characterizing adaptively introgressed Neanderthal haplotypes in modern humans

Rachel Gittelman, Benjamin Vernot, Joshua Akey
University of Washington, Seattle, WA, USA

Approximately 2% of non-African genomes were inherited from Neanderthal ancestors, and recent studies have identified Neanderthal sequences that survive in contemporary individuals. Although the majority of surviving Neanderthal sequences are rare and largely influenced by the interaction of genetic drift and purifying selection, a number of high frequency Neanderthal haplotypes have been identified, consistent with the hypothesis of adaptive introgression. Here, we describe comprehensive population and functional genomics analyses on 18 putatively adaptive high frequency Neanderthal haplotypes in non-African populations. We performed extensive simulations to confirm signatures of positive selection and estimate the strength of selection that has acted on these sequences. Next, we carefully analyzed the adaptively introgressed haplotypes and find that many of the Neanderthal variants occur in predicted functional sites of non-coding DNA. Furthermore, we analyzed large-scale gene expression data from the GTEx Project and show that several adaptive Neanderthal haplotypes act as eQTLs to nearby genes, providing evidence that they play a role in gene regulation. Two notable examples include a ~100kb haplotype on chromosome 4p14 which is associated with expression of *TLR1*, *TLR6*, and *TLR10*, and a ~20kb haplotype on chromosome 11q23 that is associated with *POU2F3* expression. The TLR (toll-like receptor) genes and *POU2F3* are involved in pathogen recognition and skin biology, respectively, both of which were subject to strong selection during human evolution. These and other high frequency Neanderthal haplotypes provide compelling insight into the ways in which modern humans could have rapidly acquired adaptations to novel environments as they left Africa.

22.12

Adaptive and non-adaptive effects of a Drosophila melanogaster natural mutation through different molecular mechanims

Lain Guio, Josefa Gonzalez
Institute of Evolutionary Biology, Barcelona, Spain

Bari-Jheh is a Drosophila melanogaster natural transposable element insertion that is associated with adaptive and non-adaptive effects. We have shown that *Bari-Jheh*increases the resistance to oxidative stress by affecting the expression of its nearby genes: *Jheh1, Jheh2* and *Jheh3*. *Bari-Jheh* adds extra antioxidant response elements to the upstream region of J*heh1* and *Jheh2* leading to increase expression of these two genes. However, we also found that the expression of *Jheh3* is down-regulated under oxidative stress conditions. On the other hand, *Bari-Jheh* is associated with increased developmental time and reduced viability under non-stress conditions, which most probably represents the cost of selection of this insertion. Under non-stress conditions,*Jheh2* and *Jheh3* are both down-regulated. To further understand the molecular mechanisms underlying the adaptive and non-adaptive effect of this insertion, we have tested whether *Bari-Jheh* affects the epigenetic marks of the genomic region where it is inserted. We hypothesized that *Bari-Jheh* could be introducing histone marks that might be affecting its surrounding chromatin configuration leading to a down-regulation of *Jheh2*and *Jheh3* under non-stress conditions. Moreover chromatin configuration could be affected by stress conditions leading to further changes in the expression pattern of the nearby genes. To test these hypotheses, we performed ChIP-PCR assays in *Bari-Jheh* and in the promoter regions of *Jheh1*, *Jheh2* and *Jheh3*. We propose an explanation for the observed gene expression changes that lead from a non-adaptive to an adaptive effect of this insertion.

22.13

Expression Piggy-backing: Neighboring genes show correlated evolution in gene expression

Avazeh Ghanbarian, Laurence Hurst
University of Bath, Bath, UK

When considering the evolution of a gene's expression profile we commonly assume that this is unaffected by its genomic neighborhood. This is, however, in contrast to what we know about the lack of autonomy between neighboring genes in gene expression profiles in extant taxa. Indeed, in all eukaryotic genomes genes of similar expression-profile tend to cluster, reflecting chromatin level dynamics. Does it follow that if a gene increases expression in a particular lineage then the genomic neighbors will also increase in their expression or is gene expression evolution autonomous? To address this here we consider evolution of human gene expression since the human-chimp common ancestor, allowing for both variation in estimation of current expression level and error in Bayesian estimation of the ancestral state. We find that in all tissues and both sexes, the change in gene expression of a focal gene on average predicts the change in gene expression of neighbors. The effect is highly pronounced in the immediate vicinity (<100kb) but extends much further. Sex-specific expression change is also genomically clustered. As genes increasing their expression in humans tend to avoid nuclear lamina domains and be enriched for the gene activator 5-hydroxymethylcytosine, we conclude that, most probably owing to chromatin level control of gene expression, a change in gene expression of one gene likely affects the expression evolution of neighbors, what we term expression piggy-backing, an analog of hitchhiking.

22.14

Role of transcriptomic plasticity in adaptation to new environments

Saad Arif [1], Li Ying Tan[1], David Kingsley[2,3], Felicity Jones[1]

[1] Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany, [2] Department of Developmental Biology, Stanford University School of Medicine, Stanford, California, USA, [3] Howard Hughes Medical Institute, Stanford University, Stanford, California, USA

The significance of phenotypic plasticity in the preliminary stages of adaptive evolution has been much debated. Phenotypic plasticity in an ancestral group can both facilitate the adaptive response to a new selective regime and account for the repeated evolution of similar ecotypes. We are studying the relationship between gene expression plasticity and loci subject to natural selection during the adaptation of marine threespine stickleback fish to freshwater environments. Independent colonization of freshwater habitats by ancestral marine sticklebacks has resulted in the independent and parallel evolution of freshwater adapted ecotypes. Recent whole genome sequencing of multiple marine-freshwater population pairs identified >81 marine-freshwater divergent loci consistently underlying adaptation to the different environments. To identify genes involved in a plastic gene expression response to water salinity, we sequenced the transcriptomes of 8 siblings from a marine-x-marine cross, lab-reared in either freshwater or marine conditions. Analysis of differential gene expression between the salinity treatments yielded a set of 83 genes, many with known osmoregulatory function. Although studies in other organisms suggest gene expression is predominantly regulated in *cis*, none of these 83 differentially expressed genes fall proximal to the previously identified parallel adaptive loci. While plastic gene expression may play an important role in enabling marine fish to colonize freshwater environments, our data indicates that the genomic loci subsequently subject to natural selection are distinct from genes showing a plastic response to salinity. We are continuing to explore whether loci subject to selection regulate transcriptomic plasticity from a distance.

22.15

DNA methylation variation in Arabidopsis has a genetic basis and shows evidence of local adaptation

Manu Dubin [1], Pei Zang[1,2], Dazhe Meng[2], Marie-Stanislas Remigereau[2], Edward Osborne[3], Francesco Paolo Casale[4], Phillip Drew[5,6], Greraldine Jean[5,6], André Kahles[5,6], Bjarni Vilhjálmsson[1], Joanna Jagoda[1], Selen Irez[1], Victor Vornin[1], Qiang Song[2], Quan Long[1], Gunner Rätsch[5,6], Oliver Stegele[4], Richard Clark[3], Magnus Nordborg[1,2]

[1] Gregor Mendel Institute, Vienna, Austria, [2] University of Southern California, Los Angeles, USA, [3] University of Utah, Salt Lake City, USA, [4] European Bioinformatics Institute, Hinxton, UK, [5] Friedrich Miescher Lab, Tübingen, Germany, [6] Memorial Sloan-Kettering Cancer Center, New York, Uzbekistan

Epigenome modulation potentially provides a mechanism for organisms to adapt, within and between generations. However, neither the extent to which this occurs, nor the mechanisms involved are known. Here we investigate DNA methylation variation in Swedish Arabidopsis thaliana accessions grown at two different temperatures. Environmental effects were limited to transposons, where CHH methylation increased with temperature. This variation was detectable within 24 hrs of a change in temperature and appears to depend on the RNA directed-DNA methylation pathway.

In addition to temperature sensitivity, GWAS revealed extensive CHH methylation variation strongly associated with genetic variants in both cis and trans, including a major trans-association close to the DNA methyltransferase CMT2.

Unlike CHH methylation, CpG gene body methylation (GBM) was not affected by growth temperature, but was instead correlated with the latitude of origin. Accessions from colder regions had higher levels of GBM for a significant fraction of the genome, and this was associated with increased transcription for the genes affected. GWAS revealed that this effect was largely due to trans-acting loci, many of which encode chromatin associated or remodeling proteins. Finally we showed that many of these loci appear to be under selection in the swedish population and show signatures of adaptation to the local environment.

These findings provide arguably the strongest evidence to date of the role of DNA methylation in adaptation to local environment, and provide a basis for further dissecting how environmentally driven and genetically determined epigenetic variation interact and influence organismal fitness at the molecular level.

22.16

Alternative splicing: widespread fine-tuning regulatory process or costly errors?

Alexandra Popa[1], Baptiste Saudemont[2], Joanna Parmley[1], Olivier Arnaiz[3], Linda Sperling[3], Eric Meyer[2], Laurent Duret [1]

[1] CNRS, Université Lyon 1, Villeurbanne, France, [2] CNRS, ENS, Paris, France, [3] CNRS, Gif sur Yvettte, France

A large number of eukaryotic genes are subject to alternative splicing (AS). In several instances it has been clearly established that AS has been selected to increase the diversity of functional protein isoforms produced by a given gene. Furthermore, in some genes, AS is used as a way to regulate expression level, by creating unproductive transcripts (i.e. mRNAs containing premature termination codons), which get degraded via the action of the NMD machinery. However, like any other cellular machinery, the action of the spliceosome is subject to errors. Hence, in any transcriptome, a fraction of observed splice variants probably correspond to erroneous transcripts, which could incur a substantial cost for the organism. The extent to which splice variants represent functionally selected transcripts (either for protein isoform production or for regulatory purpose), as opposed to errors of the splicing process, has been a debated issue. The two hypotheses (functional variants vs. costly errors) make distinct predictions, that can be tested by analyzing transcriptomes of wild-type and NMD-deficient cells. We will present analyses of paramecium and mammalian transcriptomes, which suggest that in many eukaryotes, the vast majority of splice variants correspond to costly errors

22.17

Fixation of gene duplications due to beneficial increases in gene expression

Margarida Cardoso-Moreira [1,2], J. Roman Arguello[1,2], David Riccardi[1], Srikanth Gottipati[1,3], Lawrence G. Harshman[4], Jennifer K. Grenier[1], Andrew G. Clark[1]

[1] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA, [2] Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, [3] Translational Medicine and Think Team, Otsuka Pharmaceutical Development and Commercialization, Inc, Princeton, NJ, USA, [4] Department of Biology, University of Nebraska, Lincoln, NE, USA

Gene duplications are a crucial source of genetic innovation and serve as major contributors to phenotypic evolution. Despite their evolutionary importance there are critical gaps in our understanding of how gene duplications fix in natural populations. Two outstanding questions remain: 1) Are gene duplications more often fixed by genetic drift or by positive selection?, and 2) If gene duplications are fixed by positive selection, what are the targets of selection? Here we address these two questions using the Drosophila Global Diversity lines, a novel resource of 84 *D. melanogaster* genomes derived from 5 geographically diverse populations that is well-suited for studies of local adaptation. We detected ~ 2,200 polymorphic duplications, including ~500 whole-gene duplications. We investigated the potential impact of these duplications by generating gene expression profiles for each line. We found that ~50% of gene duplications lead to significant increases in gene expression, and that these changes in dosage are usually deleterious. However, the high-frequency gene duplications, for which we found evidence of positive selection, also had significant increases in gene expression. Our work suggests that gene duplications can be driven to fixation by positive selection in Drosophila, probably due to beneficial increases in gene dosage.

22.18

Long-term survival of duplicate genes despite absence of subfunctionalized expression.

Xun Lan [1], Jonathan K. Pritchard[1,3]

[1] Department of Genetics, Stanford University, Stanford, CA, USA, [2] Department of Biology, Stanford University, Stanford, CA, USA, [3] Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

Gene duplication is a fundamental process in genome evolution. However, new duplicates are susceptible to loss-of-function mutations unless they evolve distinct functions. One standard model is that duplicate genes can avoid mutational destruction by evolving divergent expression profiles. We examined this hypothesis using expression data from 46 human tissues. Surprisingly, we found that subfunctionalization of expression occurs very slowly, and is rare among duplications that occurred within the mammals. Most young duplicates occur in tandem and have highly correlated expression profiles, likely due to shared regulation. Moreover, a large fraction of duplicate gene pairs exhibit a striking asymmetric pattern in which the expression of one gene, the "major" gene, exceeds or equals expression of the other gene in all measured tissues. These asymmetrically expressed duplicates may be long-lived, albeit with the minor genes evolving under reduced selective constraint. Dosage sharing of expression may help explain long-term survival of minor genes.

22.19

Duplication and functional divergence of miRNAs in the spider *Parasteatoda tepidariorum*

Daniel J. Leite[1], Maria Ninova[2], Evelyn E. Schwager[1], Maarten Hilbrant[1], Alexandra Buffry[1], Sam Griffiths-Jones[2], Matthew Ronshaugen[2], Alistair P. McGregor [1]

[1] Oxford Brookes University, Oxford, UK, [2] University of Manchester, Manchester, UK

Gene duplication and divergence is an important mechanism for the evolution of gene regulatory networks. Sequencing of the *Parasteatoda tepidariorum* transcriptome has shown that this spider contains a large number of duplicated genes. However, the extent of duplication and diversification of non-coding genes like microRNAs (miRNAs) was not known. Indeed, while miRNA evolution has been investigated in insects, little is known about the function and evolution of thes genes in other arthropods, for example in chelicerates. To identify the genome wide repertoire of miRNAs in *Parasteatoda*, we sequenced and analysed small RNAseq data from 10 different embryonic stages of this spider. We identified a total of 247 miRNAs of which 111 had homologs in other species annotated in miRBase. Among these homologs, 24 miRNAs were found with at least two paralogs including duplicates of highly conserved animal miRNAs, for example iab-4 and miR-993, which are found in the Hox clusters of arthropods. Furthermore, analysis of duplicated miRNAs and their predicted targets, including Hox genes, suggests that the paralogs of these genes may have evolved functional differences in their regulatory interactions. Overall our results suggest that the lineage leading to *Parasteatoda* experienced a whole genome duplication leading to loss or retention and functional divergence of miRNA paralogs. Therefore this spider represents an excellent model to further explore the evolution of miRNAs and gene divergence and duplication more broadly.

22.20

Evolutionary changes in target genes of ZEB2, a transcription factor involved in brain development and mental retardation

Stefano Berto, Anna-Christina Erpf, Lisa Falkowski, Katja Jacob, Katja Nowick
University of Leipzig, Leipzig, Germany

Humans have larger brains and distinct cognitive abilities compared to other great apes. To gain insights into the evolution of these phenotypic differences, we examined human-specific differences in gene regulation and gene expression caused by the transcription factor ZEB2. ZEB2 plays an important role during brain development by being involved in the differentiation of neuronal progenitor cells. Furthermore, mutations in ZEB2 have been associated with Mowat-Wilson syndrome, a disorder characterized by microcephaly and mental retardation, making ZEB2 a prime candidate for driving brain-related differences between humans and other apes. By analyzing co-expression patterns and weighted topological overlap networks we found that ZEB2 has a significantly higher connectivity in a transcription factor network in humans compared to chimpanzees, indicating that the set of ZEB2 target genes differs between both species. We then determined target genes of ZEB2 experimentally in B-lymphoblastoid and fibroblast cell lines of humans, chimpanzees, and orang-utans using chromatin-immunoprecipitation and knock-down experiments followed by Next-Generation-Sequencing (ChIP-Seq and RNA-Seq). Having comparatively little variation in the predicted ZEB2 target genes between individuals of the same species, we identified multiple species-specific ZEB2 targets. The set of human-specific ZEB2 targets contains many genes with known functions in the nervous system, for instance genes coding for proteins of the synaptosome, as well as many non-coding RNA genes and genes coding for histones, suggesting human-specific changes in the regulation of epigenetic mechanisms. Taken together, gene regulatory pathways controlled by ZEB2 might have been involved in the evolution of the human brain and its cognitive abilities.

22.21

Evolution of toxin resistance via polygenic *cis*- regulatory adaptation

Santiago Naranjo, Justin Smith, Carlo Artieri, Mian Zhang, Yiqi Zhou, Michael Palmer, Hunter Fraser
Stanford University, Stanford, CA, USA

Despite a growing consensus that gene expression *cis*-regulation plays a major role in evolutionary adaptation, few experimentally validated examples of gene expression adaptations exist, and fewer still have been shown to impact fitness. We have found that the budding yeast *Saccharomyces paradoxus* has recently evolved resistance to citrinin, a naturally occurring mycotoxin. Applying a genome-wide test for selection on *cis*-regulation, we identified five genes involved in the citrinin response that are constitutively up-regulated in *S. paradoxus*. Four of these genes are necessary for resistance, and are also sufficient to increase the resistance of a sensitive strain when over-expressed. Moreover, *cis*-regulatory divergence in the promoters of these genes contributes to different components of the resistance phenotype, suggesting that resistance is actually a composite of multiple independent traits. Our results demonstrate how the subtle effects of individual regulatory elements can be combined, via natural selection, into a powerful polygenic *cis*-regulatory adaptation.

22.22

Disentangling the effects of mutation and selection on the evolution of gene expression

Brian Metzger, Dave Yuan, Fabien Duveau, Patrica Wittkopp
University of Michigan, Ann Arbor, MI, USA

A primary mechanism by which phenotypes evolve is through changes in gene expression. These changes in expression are known to be caused by both cis and trans-regulatory mutations. While selection is expected to act differently on cis and trans-regulatory changes due to differences in their degree of pleiotropy and dominance, whether there are fundamental differences in the types of regulatory changes created by cis and trans-regulatory mutations remains unknown. In order to test this hypothesis, we used the Saccharomyces cerevisiae TDH3 promoter to drive expression of a yellow fluorescent protein (YFP) and then used flow cytometry to quantitatively measure the effects of over 400 mutations that impact YFP expression. We found that most cis-regulatory mutations had little impact on the mean level of expression and instead increased the variability in expression between genetically identical individuals, i.e. expression noise. By contrast, trans-regulatory mutations had little impact on gene expression noise, but were capable of creating a wide range of effects on mean expression level. Comparing these effects to naturally occurring S. cerevisiae strains indicated the action of natural selection on both mean expression and expression noise. Interestingly, the effects of selection on mean expression was limited to trans-regulatory elements, while selection on gene expression noise was limited to cis-regulatory sequences. These results do not suggest an inability of selection to act, but instead suggest that inherent differences in the frequency and effects of mutations in cis- and trans-regulatory elements can play a dominant role in dictating how gene expression evolves.

22.23

Stabilizing selection and lineage specific evolution of stress response in *Saccharomyces sensu stricto*

Joshua Schraiber, Seungsoo Kim, Jennifer Madeoy, Joshua Akey
University of Washignton, Seattle, WA, USA

Studying the evolution of response to stress in laboratory conditions can shed light on how organisms cope with the variable environments that they face in nature. Using five *Saccharomyces sensu stricto* species as a model system, we investigated the impact of heat shock and dithiothreitol (a strong reducing agent) on gene expression. We identified genes that were differentially regulated in each species when subjected to stress conditions. In most species, we identified hundreds of genes with differential expression. In contrast, *S. baynus* has relatively few genes differentially regulated in either stress condition, suggesting that it may be more robust to environmental changes. We then focused on understanding the tempo and mode of stress response evolution by using an empirical Bayes phylogenetic method to pool information across genes. Overall, we find that stress response has evolved via strong stabilizing selection, suggesting that the species inherited a common set of response pathways from their common ancestor. Similarly, when looking at core stress response genes that are differentially expressed in every species, we again find support for stabilizing selection as the primary mode of evolution. Nonetheless, each species face specific environmental challenges, and when we examine species-specific differentially expressed genes, we find that they are characterized by rapid, lineage specific evolution. Thus, these genes may be important in shaping the adaptations that characterize each species.

22.24

Variation of gene expression response to developmental temperature of two natural populations affects their evolutionary response in novel thermal environment.

François Mallard, Raymond Tobler, Viola Nolte, Christian Schlötterer
Veterinärmedizinische Universität, Wien, Austria

Phenotypic plasticity, the ability of a genotype to express distinct phenotypes in different environments, is assumed to facilitate coping with new or rapidly changing environments. In the context of climate change, the phenotypic response to temperature has received particular attention since thermal plasticity could facilitate adaptation by allowing phenotypic accommodation followed by genetic assimilation. Although temperature is an important variable that affects many traits, including gene expression, the genetic basis of phenotypic plasticity evolution remains poorly explored. We performed an evolution experiment using two Drosophila simulans natural populations from distinct geographical origins: Portugal and Florida. Replicates of each population were evolved in one of two environments: cold, fluctuating between 10 and 20°C and hot fluctuating between 18 and 28°C. Using pool RNA sequencing, we then compared the gene expression profiles of the evolved and unevolved populations at two different temperatures, 15 and 23°C, in separate common gardens. The genetic expression of our two initial populations shows different patterns. The Portuguese founder population had a reduced overall phenotypic plasticity (<1000 genes differentially expressed between 15°C and 23°C) compared to the Florida one (> 4000 genes). After adaptation to the new temperature environments the Florida population became less plastic while the evolved Portuguese flies increased their plasticity. We show that different initial thermal plasticity could lead the evolution of gene expression into different directions, strengthening the role of phenotypic in the early steps of adaptation to novel changing environment.

22.25

Analysis of allele-specific expression in F1 hybrids reveals *cis*- regulatory changes associated with recent adaptive floral evolution in *Capsella rubella*

Kim Steige[1], Johan Reimegård[1], Daniel Koenig[2], Douglas Scofield[1], Tanja Slotte [1,3]
[1] Uppsala University, Science for Life Laboratory, Uppsala, Sweden, [2] Max Planck Institute of Developmental Biology, Tübingen, Germany, [3] Stockholm University, Science for Life Laboratory, Stockholm, Sweden

Cis-regulatory changes have long been suggested to contribute to organismal adaptation. While *cis*-regulatory changes can now be identified on a transcriptome-wide scale, in most cases the adaptive significance and mechanistic basis of rapid *cis*-regulatory divergence remains unclear. Here, we have characterized *cis*-regulatory changes in association with recent adaptive floral evolution in the selfing plant *Capsella rubella*, which diverged from the outcrosser *Capsella grandiflora* less than 200 kya. We assessed allele-specific expression (ASE) in leaves and flower buds at 18,452 genes in three interspecific F1 *C. grandiflora* x *C. rubella* hybrids. After accounting for technical variation and read-mapping biases using genomic reads, we estimate that an average of 44% of genes show some evidence of ASE, however only 6% show strong allelic expression biases. Flower buds, but not leaves, show an enrichment of genes with ASE in genomic regions responsible for phenotypic divergence between *C. rubella* and *C. grandiflora*. We further detected an excess of heterozygous TE insertions in proximity to genes with ASE, with a trend towards stronger ASE for genes near TEs targeted by 24-nt small RNAs. Finally, we leverage population genomic data to test for a shift in selection on regulatory regions showing ASE between *C. grandiflora* and *C. rubella*. Our results suggest that *cis*-regulatory changes have been important for recent adaptive floral evolution in *Capsella* and that differences in TE dynamics between selfing and outcrossing species could be an important mechanism creating rapid regulatory divergence.

22.26

Association mapping reveals the role of mutation-selection balance in maintaining genomic variation for gene expression.

Emily Josephs, Young Wha Lee, John Stinchcombe, Stephen Wright
University of Toronto, Toronto, ON, Canada

Genetic variation for quantitative traits persists within populations despite the expectation that prevalent stabilizing selection should reduce genetic variance. One hypothesis suggests that variation is maintained by a balance between new mutations and their removal by selection and drift, resulting in an excess of low-frequency variants and a negative correlation between minor allele frequency and selection coefficient1. Here, we test these predictions using the genetic loci associated with total expression variation (‘eQTLs') and allele-specific expression variation (‘aseQTLs') mapped within a single population of the plant Capsella grandiflora. In addition to finding eQTLs and aseQTLs for a large fraction of genes, we show that these loci are rarer than expected and exhibit a negative correlation between effect size and frequency. Overall, our results show that mutation-selection balance is the dominant contributor to genomic variation for expression within a single, inter-breeding population.

651A

Selection against microRNA target sites

Antonio Marco
University of Essex, Colchester, UK

MicroRNAs are powerful gene regulators that play an important role in the evolution of regulatory networks. At the population level, the study of polymorphisms allows the identification of microRNA target sites under positive or purifying selection. Mutations generate novel microRNA target sites, some of which may affect the expression of genes. Thus, some transcripts will avoid target sites for specific microRNAs. I have developed a theoretical model for microRNA target avoidance based on the comparison of allele frequency distributions in untranslated regions (UTRs). When applying this strategy to *Drosophila*, I detected that genes transmitted by the mother into the egg avoid target sites for microRNAs also deposited in the egg. That is, maternal genes avoid maternal microRNAs. Some of these results have been deposited in bioRxiv (http://biorxiv.org/content/early/2014/12/16/012757). MicroRNA target avoidance is also prevalent in humans, particularly in cancer-related gene transcripts.

652B

Evolutionary analysis linked to epigenetic modifications in the largest transcription factor family of humans

Adamandia Kapopoulou [1,2], Lisha Mathew[1,2], Didier Trono[1], Jeffrey Jensen[1,2]
[1] EPFL, Lausanne, Switzerland, [2] SIB, Lausanne, Switzerland

The KRAB-containing zinc finger (KRAB-ZNF) genes represent the biggest family of transcription factors in humans, yet for the great majority, their function and specific genomic target remains unknown. However, it has been shown that a large fraction of these genes arose from segmental duplications, and that they have expanded in gene and zinc finger number throughout vertebrate evolution. To determine whether this expansion is linked to selective pressures acting on different domains, we have manually annotated all KRAB-ZNF genes present in the human genome and assessed the evolutionary forces acting at the sequence level as well as on the epigenome. When compared across primates and across tissues, KRAB-ZNFs demonstrate species-specific expression rather than tissue-specific expression. Interestingly, those carrying a *nonsynonymous* SNP in their DNA-contacting amino acids exhibit significantly reduced expression in all tissues accompanied by repressive histone marks; these KRAB-ZNF genes also seem to be less strongly constrained than those without such polymorphism. This work represents the first large-scale effort to characterize selective effects on KRAB-ZNFs and to correlate those effects with epigenetic modifications and population genetic data.

653C

The novel gene *mai* is completely embedded in *citC* in enterohemorrhagic *Escherichia coli* O157:H7 and evolved recently by overprinting

Klaus Neuhaus [1], Lea Fellner[1], Christian Scherling[2], Svenja Simon[3], Michael Witting[4], Richard Landstorfer[1], Christine Polte[1], Sonja Vanderhaeghen[1], Daniel Keim[3], Philippe Schmidt-Kopplin[4], Siegfried Scherer[1]

[1] Lehrstuhl für Mikrobielle Ökologie, Wissenschaftszentrum Weihenstephan, Technische Universität München, D-85350 Freising, Germany, [2] Lehrstuhl für Ernährungsphysiologie, Wissenschaftszentrum Weihenstephan, Technische Universität München, D-85354 Freising, Germany, [3] Lehrstuhl für Datenanalyse und Visualisierung, Fachbereich Informatik und Informationswissenschaft, Universität Konstanz, D-78457 Konstanz, Germany, [4] Research Unit Analytical BioGeoChemistry, Deutsches Forschungszentrum für Gesundheit und Umwelt GmbH, Helmholtz Zentrum München, D-85754 Neuherberg, Germany

Horizontal gene transfer is important for evolution, but genes evolved *de novo* are increasingly acknowledged as a source of adaptation to novel niches. ‘Overprinting' expresses novel, but hidden alternative open reading frames (ORFs) overlapping annotated genes.

While the evolution of overlapping genes is thought to be unlikely due to information constraints, a surplus of long overlapping ORFs exists in enterohemorrhagic *E. coli* (EHEC) compared to statistical expectations [1]. The protein-coding capacity of these ORFs is an open question.

Strand-specific transcriptomics revealed that an overlapping ORF embedded in frame -2 relative to *citC* of EHEC EDL933 was transcribed in cow dung [2]. The overlapping ORF was strand-specifically arrested by introducing a nonsense mutation silent in *citC*. The mutant showed a phenotype in competitive growth against wild type in plain LB medium and medium supplemented with $MgCl_2$ or or malonic acid. Metabolome analyses revealed slight differences in metabolites between mutant and wild type, e.g., a metabotype of TCA intermediates and linked metabolites was found. The new embedded protein-coding ORF is designated *mai* for ‘malonic acid induced'.

The citrate lyase ligase gene *citC*, which is the mother gene of *mai*, is taxonomically broadly distributed (mostly in γ-proteobacteria and firmicutes). In contrast, *mai* is found only in the *Escherichia* / *Salmonella* clade.

New orphans are often short and weakly transcribed which applies to the novel gene *mai*. We suggest that overprinting is an important but unappreciated novelty source in prokaryotes.

[1] Mir etal (2012) PLoS One 7(9):e45103
[2] Landstorfer etal (2014) BMC Genomics 15:353

654D

Adaptive Introgression at the Amylase Gene Cluster

Miriam Linnenbrink, Ellen McConnell, Diethard Tautz
Max Planck Institute for Evolutionary Biology, Ploen, Germany

The amylase gene cluster contains *Amy1* (salivary gland amylase) and *Amy2* (pancreatic amylase) genes. Amylases are enzymes that breakdown starch into small oligosaccharides. In humans, Amy1 is highly polymorphic for copy number variation and correlates with starch consumption. In dogs, the pancreatic Amy2b appears to have increased copy number in response to domestication and a more starch-rich diet. A CNV study in several mouse populations shows variation only in additional copies of Amy2b. Based on a genome-wide SNP Chip array we found a strong selective sweep pattern around Amy2b in *M.m.domesticus* and *M.m.musculus* populations and an unexpectedly high rate of genomic introgression of haplotypes from *M.m.musculus* to *M.m.domesticus*. Interestingly, some domesticus populations harbour a stop codon in the first exon, resulting in a non-functional pseudogene. To determine the allele frequency distribution, we genotyped nine wild caught *M.m.domesticus* and two *M.m.musculus* populations at this locus. Both German *M.m.domesticus* populations are fixed for the non-functional allele, whereas five French *M.m.domesticus* and two *M.m.musculus* populations carry the functional variant. The non-functional variant may have been fixed by drift during colonization of Western Europe (possibly an allele surfing effect). The introgressing, functional variant now appears to have replaced the non-functional copy of Amy2b, which could be the reason for its adaptive spread. Functional differences of both Amy2b variants were shown by western blotting and activity assays on pancreas. An analysis of 80 mice overlapping with Wang et al. 2013 even raises evidence for a connection between amylase genotype, long-term-diet and gut enterotype.

655A

Comparative gene expression profile in *Spodoptera frugiperda* (Lepidoptera, Noctuidae) host races feeding on preferential and alternative host plants

Karina Brandao [1], Marcelo Brandao[2], Renato Horikoshi[1], Daniel Bernardi[1], Celso Omoto[1], Antonio Figueira[1]

[1] Universidade de Sao Paulo, Piracicaba, Sao Paulo, Brazil, [2] Universidade Estadual de Campinas, Campinas, Sao Paulo, Brazil

Ecological differentiation and speciation events in phytophagous insects can be induced by the use of alternative host plants found within their geographical distribution. In this scenario, the genes involved in physiological and behavioral responses to host plant recognition may be implicated in genetic differentiation, reflecting in lineages divergence. Here, we propose to identify the genetic and evolutionary mechanisms by which ecological speciation and reproductive isolation might occur in populations of an important pest insect in Brazil with host races differentiation, the moth *Spodoptera frugiperda*, by comparing their RNAseq expression profile. Larvae of "rice" and "corn" races were reared in both corn and rice leaves until 6th instar, and had their RNA extracted and sequenced in an Illumina HiSeq 2500. The transcriptome was assembled as a *de novo* hybrid assembling using ETSs data from SpodoBase project and the in house Illumina sequenced reads. Assembly quality control was assessed by assemblathon suite, and Chimeric assembled contigs were identified and removed using a suite of customized Perl script (blast CHECK), resulting in 71,425 contigs. From those we have found 9,849 candidate contigs that are statistically differentially expressed (>5 folds). The recent advances in sequencing technology has provided new strategies for studying lineage divergence, and we make use of one of these new technologies to postulate that the lineage divergence in *S. frugiperda* seems to be more related on lowering, or silencing, the transcription of some genes than on the activation of new genes when the different host races are reared in rice leaves.

656B

STUDY OF ADAPTATION TO HEAVY METAL STRESS IN THE HYDROTHERMAL MUSSEL *Bathymodiolus azoricus* (BIVALVIA, MYTILIDAE) USING TRANSCRIPTOMIC APPROACH.

Gonzalo Fuenzalida [1,2], François Lallier[1,2], Arnaud Tanguy[1,2]

[1] UPMC, Université Paris 06, Paris, France, [2] CNRS, UMR 7144 Station Biologique de Roscoff, Roscoff, France

The hydrothermal vents ecosystems are widely distributed around the world, mainly where new oceanic crust is formed. Despite the strong and stressful physicochemical conditions (toxic concentrations of heavy metals and gases, high temperature, anoxia, low pH) organisms ranging from bacteria to invertebrates have developed adaptations to this conditions. However the genetic mechanisms which conducted to the establishment of these adaptations (adaptive polymorphism in stress genes, gene duplication) are not well described but remained key factors to understand how the evolutionary forces are acting at population level. *Bathymodiolus azoricus* is a bivalve belonging to the Mytilidae family that is widely distributed in Mid-Atlantic Ridge, forming large communities around hydrothermal vents even in nearby areas where hydrothermal flow emerges at high temperatures and loaded with toxic components. Here we use mussel populations of Mid-Atlantic Ridge issued from three highly contrasted environments: Menez Gwen, Lucky Strike and Rainbow, which are characterized by an increasing gradient of heavy metals (cadmium, copper, iron, zinc) concentrations. This study aims to compare those populations using transcriptomic (microarrays and qPCR) approach to identify specific gene expression patterns that could be linked to heavy metal tolerance and initiate a study to identify adaptive polymorphisms through SNP analysis.

657C

The role of miRNA in the diversification of Midas cichlid fish

Paolo Franchini, Peiwen Xiong, Carmelo Fruciano, Axel Meyer
University, Konstanz, Germany

Cichlid fishes are an ideal model system for the study of biological diversification, and are a textbook example for adaptive radiation and speciation. Most research conducted to date on the molecular basis of speciation has neglected the role of gene regulation. However, in recent years, its role has been recognized as a powerful force in driving the diversification of gene function and speciation. We investigated the potential role of miRNA regulation in the diversification of five cichlid species of the Midas cichlid lineage (Amphilophus spp.) occurring in Nicaraguan crater lakes Apoyo and Xiloá (including a benthic and a limnetic species from each lake) and in the two large and old Lakes Nicaragua and Managua (both housing the ancestral species A. citrinellus). We predicted 236 miRNA genes using as references the miRBase teleost miRNAs and the Midas cichlid genome. These were used to target the miRNA binding sites on 8232 Midas 3' UTRs. Additionally, using Illumina sequencing, we obtained low coverage genomes for the five focal species that were used for SNP calculation in the annotated miRNA genes and 3' UTRs. As expected, we find the miRNA genes to be conserved, but we find signatures of purifying selection in the miRNA target sites when compared to their flanking regions and to simulated neutral expectation. However, the Lake Apoyo species pair showed a different pattern where relaxed purifying selection has been detected. Research on gene regulation by miRNA will provide fundamentally new insights into the processes of phenotypic diversification and speciation.

658D

Genomic location of conserved noncoding sequences sheds more light to gene expression dynamics

Isaac Adeyemi Babarinde [1,2], Naruya Saitou[1,2]
[1] *National Institute of Genetics, Mishima, Japan,* [2] *SOKENDAI (The Graduate University for Advanced Studies, Mishima, Japan,* [3] *The University of Tokyo, Tokyo, Japan*

Conserved noncoding sequences (CNSs) have been reported to be associated with gene expression dynamics of certain gene groups. However, the specific interaction between CNS and neighboring genes has not been fully understood. This understanding is crucial for identifying putative regulatory elements. We first confirmed that CNSs are functionally important and are less likely to be long noncoding RNAs. To elucidate the impact of CNSs on gene expressions, we investigated their genomic locations and the gene expression dynamics of the neighboring genes. CNSs tend to cluster around genes of certain ontology groups and genes expressed in certain tissues. Generally, CNSs are more associated with tissue- or stage-specific genes. Specifically, genes associated with development, nervous systems and transcriptions as well as genes expressed in brain and fetus have more CNSs. We investigated other factors that could lead to this observation, including gene evolutionary rate and gene density and found that gene expression dynamics is more associated with the genomic location of CNSs. Surprisingly, CNSs are found to be located far away from protein-coding genes, suggesting that CNSs are more likely to be distal regulatory elements. Contrary to several previous reports, CNSs were found to function more likely as repressing than enhancing elements. Taken together, this study suggests that CNSs are associated with genes which require highly controlled expression, offers new insights into the identification of conserved regulatory elements and highlights the relevance of CNS evolution to morphological diversity.

659A

Evolutionary aspects of gene expression during *Drosophila melanogaster* spermatogenesis

Julia Raices, Maria Vibranovski
University of Sao Paulo, Sao Paulo, Brazil

Genes that appeared recently in the evolutionary history of a taxonomic group are considered new genes. Although some of those new genes are functional, most of them become pseudogenes. However, when they turn to be functional, new genes can quickly convert to essential genes or bear an important function at different phases of the development of individuals of different species. In this way, new genes expressed during spermatogenesis - the system of male gamete development - are probably related to fertility, mobility, form and function of the sperm cells. And, therefore, must be more expressed during the late phases of the gamete development, bearing in mind that expression relates to functionality. In this work, we test the hypothesis that there is a relation between a gene's age and its expression during *Drosophila* spermatogenesis. According to this hypothesis, it's expected that new genes are more frequently expressed during post-meiosis, the latest phase of germline development. To answer those questions, bioinformatics and computational biology tools were used and proper statistical methods were applied to correlate already available data of gene age and expression during spermatogenesis phases in *Drosophila*. Our results have shown that the proportion of new genes expressed in late spermatogenesis (meiosis and post-meiosis) is significantly higher then in the beginning of the processes (mitosis). Also, we found that the expression level of new genes is higher than the expression of old genes during meiosis and post-meiosis, and the opposite pattern occurs during mitosis.

660B

Variable pathogen-recognising capacity across HLA-DRB1 allele lineages supports a limit to the divergent allele advantage model.

Quintin Lau [1], Yoshiki Yasukochi[2,1], Yoko Satta[1]

[1] The Graduate University for Advanced Studies, Hayama, Japan, [2] University of Tokyo, Tokyo, Japan

Genetic diversity in human leukocyte antigen (HLA) molecules is thought to have arisen from the co-evolution between host and pathogen and maintained by balancing selection. Heterozygote advantage is a common proposed scenario for maintaining high levels of diversity in HLA genes, and extending from this, the divergent allele advantage (DAA) model suggests that individuals with more divergent HLA alleles will bind and recognise a wider array of antigens. While the DAA model seems biologically suitable for driving HLA diversity, there is likely an upper threshold to the amount of sequence divergence. We used pathogen-binding capacity of DRB1 alleles as a model to further examine the DAA model, and this general approach showed evidence of optimal fitness at intermediate sequence divergence. Within the DRB1 locus, two distinct phylogenetic groups (denoted group A and B) were previously identified based on non-PBR nucleotide sequences; predictions in this study support that group A allele and group B allele lineages have contrasting pathogen recognition capacity, and this may maintain optimal rather than maximum sequence divergence. Furthermore, computer simulations revealed an inconsistency between the observed extent of polymorphisms and that predicted in the DAA model, supporting a threshold limit where the DAA model effectively works. By investigating the relationships among HLA alleles, and pathogens bound, we can provide further insights into the mechanisms on how humans have adapted to infectious diseases over time.

661C

A Whole Transcriptome Approach to Identification of Novel Reference Genes for Quantitative Gene Expression Studies in *Mimulus*

Kimmy Stanton [1], Patrick Edger[2], Josh Puzey[3], Philip Cheng[1], Dan Vernon[1], Nancy Forsthoefe[l1], and Arielle Cooley[1]

[1] Biology Department, Whitman College, Walla Walla, USA, [2] Department of Plant and Microbial Biology, University of California – Berkeley, USA, [3] Department of Biology, College of William and Mary, Williamsburg, USA

Species from the monkeyflower genus, *Mimulus*, have become widely used model organisms for evolutionary genetic studies of speciation and ecological adaptation. Many of these *Mimulus* studies utilize quantitative real-time PCR (qPCR), yet the accuracy of qPCR in quantifying gene expression is dependent upon the quality of the normalizer used to control for technical variation. Although *Mimulus* is an emerging model plant system, no reference genes have been formally validated for use in qPCR in this genus. Our study aimed to identify stable *Mimulus* qPCR reference genes by using whole transcriptome expression data collected through RNA-seq and to validate these RNA-seq results in a traditional qPCR experiment. Gene expression stability was compared across different tissue types of the widely used North American *Mimulus* species, *M. guttatus*, and its South American counterpart, *M. luteus*. We found poor expression stability of commonly used reference genes using both the RNA-seq and qPCR methods, showing that most of these genes are less than ideal normalizers of qPCR data in the genus *Mimulus*. These results are similar to results from reference gene validation in other model plant genera such as *Arabidopsis* and *Populus*. We also found that there was little correlation between gene stability as measured by RNA-seq and qPCR, sounding a cautionary note for utilization of RNA-seq expression data in choice of reference genes.

662D

tba

663A

Estimating the strength of selective constraint on intergenic sites in bacteria

Harry Thorpe, Sion Bayliss, Laurence Hurst, Edward Feil
University of Bath, Bath, UK

Whilst it is well established that specific intergenic mutations within bacterial genomes can have direct phenotypic consequences, there has been little effort to quantify the overall level of selective constraint on intergenic sites compared to protein coding regions, and intergenic mutations are still typically considered to be mostly neutral. The recent advances in bacterial genome sequencing, combined with an increasing emphasis on establishing genotype-phenotype relationships, means that a detailed examination of this null model is now both feasible and timely. We have examined the strength of selection on intergenic regions within two large whole-genome datasets of Staphylococcus aureus (1823 isolates), and Streptococcus pneumoniae (3042 isolates). Our approach was to extend the logic of dN/dS as a measure for selective constraint by proposing the use of dI/dS, where dI = intergenic SNPs / intergenic sites. Combining estimates of dI/dS with site-frequency spectra suggests that purifying selection on intergenic sites is on average three-fold stronger than on synonymous sites and three-fold weaker than on non-synonymous sites. Analysis of promoters and terminators shows that both are under moderate purifying selection, but promoters are more constrained than terminators. We also find evidence for epistatic interactions and a high rate of compensatory mutations within stem-loop structures of rho-independent terminators. Finally, we note a strong correlation in evolutionary rate between a specific gene and the regulatory elements controlling its expression; this suggests that they are subject to similar pressures and that mutations in one have consequences for the other.

664B

Evolution of gene expression and regulation in the *Plasmodium* apicoplast

Ellen Nisbet, Joanna McKenzie, Emma Carpenter, Christopher Howe
University of Cambridge, Cambridge, UK

The Apicomplexa, which include *Plasmodium* (malaria parasite) and *Toxoplasma*, contain a remnant chloroplast known as an apicoplast. In *Plasmodium* and other parasitic spieces, the apiocoplast is no longer able to carry out photosynthesis. However, the organelle is essential and is the target of important anti-malarial drugs such as doxycycline, an inhibitor of apicoplast protein synthesis. Here, we show that regulation of apicoplast gene expression is carried out by post-transcriptional RNA processing. We show that primatyapicoplast transcripts are polycistronic, and that there is extensive RNA processing. Such processing often involves the specific excision of tRNA molecules, and allows the release of mRNA molecules for overlapping genes. We have identified a conserved sequence motif which is associated with RNA cleavage, and show that an apicoplast-targeted protein binds to these sites. Surprisingly, RNA transcripts in chloroplasts of related alveolate species (photosynthetic Apicomplexa such as *Chromera* and dinoflagellate algae) undergo a number of unusual modifications, including RNA editing and the addition of 3' poly(U) tails. We show that these features are ancestral, and reflect the common evolutionary origins of the organelle.

665C

Cis-regulation at the *KLK* cluster another case of adaptive evolution in Asians?

Patrícia Isabel Marques [1,2], Filipa Fonseca[1], Tânia Sousa[1], Paulo Santos[1], Vânia Camilo[1], Zélia Ferreira[4], Victor Quesada[2], Susana Seixas[1]

[1] Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal, [2] Department of Biochemistry and Molecular Biology-IUOPA, University of Oviedo, Oviedo, Spain, [3] Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal, [4] Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, USA

The human *kallikrein* (*KLK*) cluster, located at chromosome 19q13.3-13.4, spans over 265 kb and encodes 15 serine proteases and a transcribed pseudogene. The neighboring genes, *KLK3*, *KLK2*, *KLK4* and *KLK5* play key roles in the cascades of semen liquefaction, skin desquamation and tooth enamel formation and among these, *KLK2* and *KLK3* were identified as targets of adaptive evolution in primates through mechanisms linked to reproductive biology. Several genome wide studies of positive selection also perceived some evidence that the *KLK3-KLK5* region could exhibit abnormal haplotype patterns in East-Asians only. To better understand the evolutionary forces shaping *KLK3-KLK5* region, we evaluated its genetic diversity in Asian (CHB+JPT) and non-Asian (CEU+YRI) populations. We detected significant deviations from neutral expectations of the site frequency spectrum in Asians using either "1000 Genomes" or our re-sequencing datasets and we identified multiple SNPs showing high levels of population differentiation, which were embedded in fairly homogenous haplotypes in spite of several recombination hotspots. Three variants stood out for their location on putative regulatory regions of *KLK4* and predicted functional effects (rs198968, rs1654556 and rs17800874), including the introduction of repressor motifs and microRNA binding sites. Luciferase reporter assays performed in different cellular models demonstrated that all three candidate variants may operate synergistically to downregulate *KLK4*. Taking into account the ubiquitous pattern of *KLK4* expression and its likely pleiotropic activity in different biological systems, we considered the hypothesis of the adaptive signature of KLK4 being correlated to teeth and skin phenotypic traits already proposed as selective advantageous in East-Asian.

666D

Cis -regulation evolution in *Arabidopsis*

Fei He, Agustin Arce, Juliette de Meaux
Botany Institue, University of Cologne, Cologne, Germany

To understand the genetic basis for adaptation is still a challenge. Genetic analysis of gene expression variation may contribute to a better understanding of its role adaptive traits for fitness. *Cis*-regulatory DNA has been suspected to play a pre-eminent role in adaptive evolution, but understanding the role of *cis*-regulatory mutations in gene expression divergence first requires an accurate analysis of the functional differences associated with these regions. The monitoring of allele-specific expression in F1 hybrid becomes an efficient method to detect cis regulation mutations. The ecological and phenotypic diversity has been shown in *Arabidopsis* genus. Here, we genome-widely draw the distribution of *cis*-regulatory differences in different hybrids of *A. thaliana, A. halleri* and *A. lyrata* in response to cold, drought and normal growth conditions by RNA-seq. Around two third of genes display ASE preferentially expressed the allele of *A. lyrata* or *A. halleri*. The derived cis change between species reflects their ecological diversity. *A. lyrata* shows more tolerant to drought and *A. halleri* shows tolerant to high heavy metal concentrations. In consistency, cis-regulatory changes derived in the *A. lyrata* lineage were enriched in proline transporter activities to maintain the osmotic pressure in stress, and cis-regulatory changes derived in the *A. halleri* lineage were significantly in excess among genes involved in metal ion transmembrane transporter activity. This finding provides the proof of selection on speciation during the adaptation.

667A

Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains

João Luis Reis-Cunha[1], Gabriela Rodrigues-Luiz[1], Hugo Valdivia[1], Rodrigo Baptista[1], Ana Tereza Vasconcelos[2], Daniella Bartholomeu [1]

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, [2] 2Laboratório Nacional de Computação Científica, Petrópolis, Rio de Janeiro, Brazil

American trypanosomiasis is a neglected tropical disease caused by the protozoan *Trypanosoma cruzi*, a highly polymorphic parasite currently divided into six discrete typing units (DTUs). CL Brener, the reference strain of the *T.cruzi* genome project, is a hybrid with a genome assembled into 41 putative chromosomes. Gene copy number variation(CNV), resulting from the gain or loss of genomic material, is well documented as an important mechanism to enhance gene expression and variability in *T.cruzi*. Chromosomal CNV(CCNV) is another level of gene CNV in which whole blocks of genes are expanded simultaneously. The extent of diversity in CCNV among *T.cruzi* strains based on a read depth coverage analysis has not been determined yet. Here, we identify the CCNV in different *T.cruzi* strains, by analyzing the depth coverage of short reads from these strains using the 41 CL Brener chromosomes as a reference. This study reveals explanations of some of the genome structural peculiarities of these DTUs. The TcI DTU strains are usually diploid with few aneuploidies, while the strains from TcII and TcIII DTUs present a high degree of chromosomal expansion. Chromosome 31, which is the only chromosome that was supernumerary in all six *T.cruzi* samples evaluated in this study, is enriched with genes related to glycosylation pathways, highlighting the importance of glycosylation to parasite survival. Increased gene copy number due to chromosome amplification may contribute to alterations in gene expression, which represents a strategy that may be crucial for parasites that mainly depend on post-transcriptional mechanisms to control gene expression.

668B

Developmental gene expression consequences of repeated subterranean colonization

Bethany Stahl, Joshua Gross
Dept. of Biological Sciences, University of Cincinnati, Cincinnati, OH, USA

Animals that colonize extreme habitats often evolve equally extreme phenotypes. However, our knowledge of the gene expression changes that accompany these colonization events remain largely unknown. To investigate this phenomenon, we study organisms that live in the unusual environment of a cave, which is marked by constant darkness and limited food availability. The blind Mexican cavefish, *Astyanax mexicanus*, has evolved an array of cave-associated phenotypes including regression of pigmentation and eyes, sleep loss, craniofacial aberrations, increased fat reserves and expanded sensory systems. Further, *Astyanax* has repeatedly colonized separate caves throughout Mexico, yet the "ancestral" surface form is extant, allowing for direct comparison to the "derived" cave-dwelling conspecific. In this study, we examine the numerous changes associated with subterranean colonization by measuring differential gene expression between two geographically isolated cavefish populations (Pachón and Tinaja), compared to surface morphs. We performed this analysis using RNA-seq profiling across four critical stages of early development wherein traits appear and others regress. We discovered certain genes implicated in cave evolution, such as *Mitfa* and *Slc5a8,* demonstrate similar patterns of expression across development in both cavefish populations. However, other genes including *Olfm4, Tas2r202, Sagb* and *Nrl* show unique expression profiles specific to each cave-dwelling lineage. This work reveals that cave-adapted traits arise through a combination of both shared and unique patterns of gene expression. Shared expression profiles may signal common environmental pressures driving the evolution of cavefish, yet patterns that are specific to each cave indicate that similar adaptive traits can arise through diverse genetic mechanisms.

669C

Lineage-specific loss of FGF17 in the avian orders Galliformes and Passeriformes: retention and loss of a gene arising from an ancient whole genome duplication

John Abramyan
University of British Columbia, Vancouver, BC, Canada

The developmental complexity of vertebrates is commonly attributed to two rounds of whole genome duplications which occurred at the base of the vertebrate radiation and gave rise to several multi-gene families of developmental proteins. These proteins function as molecular signals in critical cell-cell/tissue communication in order to direct the formation of the embryo. One of the more intriguing gene families to arise from these duplication events is the Fibroblast growth factors (FGF) 8 subfamily which is represented by *FGF8*, *FGF17*, and *FGF18* in tetrapods. While, *FGF8* and *FGF18* are found in all tetrapods and are critical for embryonic survival, *FGF17* is through to be lost from the genomes of various lineages ranging from frog to chicken. Due to the potential for genomic misassemble and sequencing error, the true nature of this missing FGF17 sequence has never before been confirmed. Utilizing recently available avian genomes, this study confirms the loss of *FGF17* to be a true, biological event which has occurred independently in at least two avian lineages. Interestingly, analysis of lineages which retain it reveals that FGF17 is still under strong purifying selection, despite being seemingly dispensable. Due to its close relationship to FGF8 and FGF18, and their shared receptors, mutated FGF17 protein has the potential for interfering with their function. Thus, FGF17 likely represents a molecular spandrel arising from a genome duplication and due to its high connectivity and potential for interference with other genes, is retained under strong purifying selection, despite itself not having a strong selective advantage.

670D

Adaptive evolution of the promoter region of the Sialyltransferase 8B (*STX*) gene

Naoko Fujito [1], Toshiyuki Hayakawa[2], Masaya Hane[3], Ken Kitajima[3], Chihiro Sato[3], Yoko Satta[1]
[1] School of Advanced Sciences, the Graduate University for Advanced Studies, Hayama, Kanagawa, Japan, [2] Faculty of Arts and Science, Kyushu University, Fukuoka, Fukuoka, Japan, [3] Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan

STX is an enzyme responsible for the synthesis of polysialic acids (PSA) in the human brain. PSA is typically attached to neural cell adhesion molecule (NCAM), and plays an important role in brain function. Variants of the *STX* gene have shown association with various mental disorders, such as schizophrenia, bipolar disorder and autism. Here we focus on the association between polymorphisms in the STX promoter region and schizophrenia. “Risk” and “protective” SNPs in the STX promoter region have been reported. The promoter activity is significantly different between risk and protective SNPs. This difference is considered to be involved in the risk of schizophrenia. Haplotype sequences of the 10 kb region surrounding the SNPs were determined using genomic DNA of 63 human individuals from a wide range of ethnic groups. Phylogenetic and population genetic analysis using the haplotype sequences shows that the ancestor of protective haplotypes emerged in Africa about 0.8 MYA. This is consistent with that only risk SNPs have been detected in apes. The protective haplotypes are prevalent especially in East Asian populations, and show lower nucleotide diversity than risk haplotypes. These findings raise the possibility of some local selection on protective haplotypes or on linked some other loci. To elucidate the cause of the difference in the nucleotide diversity between risk and protective haplotypes, extensive analysis of the data is ongoing.

671A

Genes involved in local adaptation and acclimation for thermal tolerance in *Tigriopus californicus*

Christopher Willett, Thiago Lima
University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

The study of genome-wide patterns of gene expression has become a useful way in which to gain an increased understanding of how organisms can respond to environmental factors and how these responses can evolve across populations. For thermal tolerance many studies have focused on adaptation to extreme, stressful thermal conditions. This study will focus instead on differential adaptation and acclimation to fluctuating but not extreme thermal regimes. Populations of the intertidal copepod *Tigriopus californicus* display both differential adaptation to thermal extremes and differential fitness under less stressful thermal regimes. This study will explore differences in gene expression using RNA-seq across different populations and different thermal conditions. Southern *T. californicus* populations have higher fitness than northern populations under moderately high, fluctuating conditions but lower fitness under slightly lower, constant thermal conditions. The results of these gene expression studies indicate that with acclimation, the higher fitness of these southern populations may be associated with the ability to handle these conditions with fewer changes in gene regulation across the transcriptome. This contrasts to results for tolerance of more extreme high temperatures where these southern populations appear to be able to more effectively upregulate a number of key genes that contribute to survival. These results suggest that patterns of evolution of gene expression changes underlying adaptation to chronic versus acute stress could differ dramatically.

672B

Intron/exon architecture is required to understand patterns of variation in nucleotide, codon or amino acid compositions of plant protein-coding genes

Adrienne Ressayre [1], Christine Dillmann[1], Sylvain Glémin[2]
[1] INRA/Univ Paris Sud, UMR Génétique Quantitative et Evolution, Gif Sur Yvette, France, [2] UMR CNRS 5554, Institut des Sciences de l'Evolution, Université Montpellier 2, Montpellier, France

Introns occupy a large proportion of most of the eukaryotic species gene space. During past decade, introns status has changed. On one hand, an increasing proportion of individual introns are now identified as hosting signals required for proper gene expression and on the other hand splicing processes have proven to be implicated in all the different steps of gene expression, from 5'-capping to first round of translation, export of mRNA into cytoplasm and mRNA stability. In parallel, the intron/exon architecture of genes is described as overlapping with patterns of gene chromatin architecture. Despite these observations, the potential impact intron on gene structure is still poorly documented, intron number and intron location being usually neglected in both gene and genome organization studies.

Plants have large genomes comprising both tens of thousands of genes and more than a hundred thousands of introns. Complete sequenced plant genomes therefore permit to document the consequences of intron presence into genes within a genome and to compare whether intron presence has similar impact on gene architecture between genomes of different species. We performed a comprehensive survey of the consequences of intron presence on nucleotide composition of genes in two widely divergent plant genomes (Arabidopsis thaliana and rice, Oryza sativa) and described the consequences of changes in the intron/exon structure of genes at a genome-wide scale. Our results converge to show that gene intron/exon architecture is required to properly described patterns of variation in nucleotide composition in plant genes.

673C

Local adaptation at the transcriptomic level in European populations of *Daphnia*

Mathilde Cordellier [1], Suda Ravindran[1], Ann-Kathrin Huylmans[2], Alberto Lopez[3]
[1] Universität Hamburg, Hamburg, Germany, [2] Ludwig-Maximilians-Universität, Munich, Germany, [3] Westfälische Wilhelms-Universität, Münster, Germany

Simultaneously phenotype and genotype, gene expression profiles are heritable, thus providing a substrate for evolution. In species with little or no gene flow between populations such as permanent *Daphnia* populations, local adaptation can lead to distinctive gene expression profiles within a single species. Selection drivers as diverse as parasitism, predation and heavy metal pollution were already shown to cause rapid genetic shifts in *Daphnia*.
In this study, we inferred the intra-specific variation at the transcriptome level in *Daphnia galeata*. A large scale RNAseq experiment was conducted on 24 clonally propagated genotypes from four European lakes. In total, 72 libraries were sequenced and long reads generated for assembly purposes. The de novo assembly was made using the EvidentialGene pipeline, allowing recovering a very high number of complete transcripts and of core eukaryotic genes. Short reads were mapped against this reference transcripts using NextGenMap. Candidate genes exhibiting differential expression patterns between populations were relatively few, partly due to the high variation within populations. A Q_{ST}/F_{ST} approach was used to identify genes departing from the neutral expectations. Expression levels of these candidate genes will be correlated with fitness measurements and sequence variation, in order to establish a link between phenotype and genotype. The outcome of this study will allow us to understand the genetic background of rapid adaptation to environmental changes in a key species of aquatic ecosystems.

674D

Sequence evolution and expression of the androgen receptor and other pathway-related genes in a unisexual fish, the Amazon molly, *Poecilia formosa* , and its bisexual ancestors.

Fangjun Zhu [1], Ingo Schlupp[2], Ralph Tiedemann[1]
[1] *Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam,Brandenburg, Germany,* [2] *Department of Biology, University of Oklahoma, Norman,Oklahoma, USA*

The all-female Amazon molly (*Poecilia formosa*) originated from a single hybridization of two bisexual ancestors, Atlantic molly (*Poecilia mexicana*) and Sailfin molly (*Poecilia latipinna*). As a gynogenetic species, the amazon molly needs to copulate with a heterospecific male, but the genetic information of the sperm-donor does not contribute to the next generation. Here, we study the sequence evolution and gene expression of the duplicated genes coding for androgen receptors (ARs) and other pathway-related genes in amazon molly, in comparison to its bisexual ancestors. Mollies possess – as most other teleost fish - two copies of the AR gene, i.e., ARα and ARβ. In comparing the analyzed species, ARα was evolutionary more conserved. We describe the gene expression pattern of AR and pathway-related genes in various tissues (i.e., brain, gill and ovary) among the 3 species. Expression levels in *P. formosa* are in most cases in the same range as in one or both of the bisexual ancestors. Over-expression in *P. formosa* relative to the bisexual species was detected in ERα (gill) and in CYP19A2 (ovary). The latter is interesting, as CYP19A2 is involved in the sex determination pathway. A tendency towards lower expression in *P. formosa* was only detected for ERβ1 (ovary). Note that this gene was expressed neither in brain nor in gill, in the analyzed molly species. All tissues in all species show a higher expression of ERα, relative to ERβ1, pointing towards a more important role of the former gene in estradiol synthesis pathway instead of ERβ1.

675A

Molecular Adaptations to Low Temperatures in Wild Tomato Species

Tetyana Nosenko, Wolfgang Stephan
Ludwig-Maximilians Universität München, Munich, Germany

Unlike its cultivated relative, the wild tomato species *Solanum chilense* adapted well to the extreme environmental conditions in its natural habitats including drought, high soil salinity, and extreme temperatures. In this study, we assessed low-temperature tolerance along an altitudinal gradient and investigated molecular mechanisms underlying this trait in low- and high-altitude populations of *S. chilense*. We observed no correlation between the chilling-temperature tolerance and environmental variables. In contrast, freezing-temperature tolerance had a strong association with the elevation and temperature variables and was significantly higher in the high-altitude populations.

Expression patterns of the key genes in the cold response pathway also differed between plants from the low- and high-altitude populations. *CBF* and dehydrin genes, which are known to be involved in the cross-talk between the cold and drought response pathways, had higher expression levels in the coastal population under both normal and chilling temperatures. High expression levels of these genes may represent an adaptation to drought and provide a molecular basis for the non-specific chilling-temperature tolerance in the low-altitude populations of *S. chilense* that occupy super-arid habitats. In turn, significantly higher constitutive expression levels of the *ICE1* and *FAD7* genes observed in the high-altitude population of *S. chilense* may represent an adaptation to steep diurnal temperature changes in the mountains and contribute to the plant freezing-temperature tolerance.

676B

The evolution of plant-pathogen defense-layers, focusing on miRNA negative regulation, in plant-parasite interactions

Sophie Rommel, Thorsten Kloesges, Laura Rose
Heinrich Heine University, Düsseldorf, Germany

Pathogens are major driving forces of host evolution. The shorter generation times and larger populations of most pathogens (relative to their hosts) are likely to give them an evolutionary advantage over their hosts in the coevolutionary arms race. Yet, despite these large differences in evolutionary potential, most plants are well-defended against a wide range of pathogens. We argue that negative regulation combined with expressional fine-tuning of plant-defense pathways is a key to achieve matching evolutionary speed. We investigated how the layers of negative regulation in plant-pathogen defense pathways are modulated during pathogen infection. In plants, the miR482/2118 gene family targets and negatively regulates resistance genes (R-genes). Studying the evolutionary dynamics and function of this R-gene-regulating miRNA family in the Solanaceae, we show, that miR482/2118 is subjected to high rates of evolution similar to their targets, but that different evolutionary constraints act on the seven family members. Comapring controlled inoculations of wild and cultivated species of *Solanum* by the pathogen *Phytophthora infestans,* we record how transcriptional responses of miRNAs and responding R gene tragets are tightly intertwined with natural variation in pathogen spread and strength of disease resistance. In conclusion, our data will show whether miRNA mediated R gene regulation contributes to host defense or if it is potentially an "Achilles heel" for plants - allowing an easy target for pathogens to further suppress host responses.

677C

The effect of human-specific promoter indels on gene expression

Frances M Marín-Maldonado, Juan C. Martínez-Cruzado, Taras K. Oleksyk
University of Puerto Rico, Mayaguez, Puerto Rico

Approximately six million years ago the human and chimpanzee lineages diverged from a common ancestor resulting in two distinct species with a great quantity of different morphological, behavioral, cognitive and other phenotypic traits. However, their genomes are more than 98.5% identical at protein-coding loci, and differences between these two species hover around 2.4%, excluding repeats and low complexity DNA and including insertions and deletions (indels). It is thus believed that most of the genetic foundation for the differences among these two lineages lies at the level of gene regulation. For instance, some of the major phenotypic traits that are distinct in these species, such as cognitive ability and fertility, may be related to observed significant differences in gene expression patterns in the nervous and reproductive system. Our work focuses on lineage-specific large (> 10 bp) indels in promoter regions that may affect gene expression and protein product levels. We started by identifying through the alignment of syntenic regions of human, chimpanzee, gorilla, orangutan and macaque, 64 indels, located 5' and at distance no more than 2 kb away from its nearest gene transcription start site. After performing PCR and electrophoresis analysis to validate real features from computational artifacts, we found two distinct genes in which their indels are found only in humans relative to all other primates. Our goal is to assess how these indels could be associated to different patterns of gene expression among these two species using in vitro gene expression assays.

678D

Unraveling the association between mRNA expressions and mutant phenotypes in a genome-wide assessment of mice

Ben-Yang Liao, Meng-Pin Weng
Division of Biostatistics & Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County 350, Taiwan

High throughput gene expression profiling has revealed substantial leaky and extraneous transcription of eukaryotic genes, challenging the perceptions that transcription is strictly regulated and that changes in transcription have phenotypic consequences. To directly assess the functional implications of mRNA transcription, we analyzed mRNA expression data derived from microarrays, RNA-seq, and *in situ* hybridization, together with phenotype data of mouse mutants as a proxy of gene function at the tissue level. The results indicated that, in spite of the presence of widespread ectopic transcription, mRNA expression and mutant phenotypes of mammalian genes or tissues remain associated. The expression-phenotype association at gene level was particularly strong for tissue-specific genes, and the association could be underestimated due to data insufficiency and incomprehensive phenotyping of mouse mutants; the strength of expression-phenotype association at tissue level depended on tissue functions. Mutations on genes expressed at higher levels or expressed at earlier embryonic stages more often result in abnormal phenotypes in the tissues where they are expressed. The mRNA expression profiles that have stronger associations with their phenotype profiles tend to be more evolutionarily conserved, indicating that the evolution of transcriptome and phenome are coupled. Therefore, mutations resulting in phenotypic aberrations in expressed tissues are more likely to occur in highly transcribed genes, tissue-specific genes, genes expressed during early embryonic stages, or genes with evolutionarily conserved mRNA expression profiles.

679A

No evidence that bacterial gene regulation has adapted to mitigate the accumulation of toxic metabolites

Adam Gyorkei [1], Balazs Szappanos[1], Laurence Hurst[2], Balazs Papp[1]
[1] Biological Research Centre, Szeged, Szeged, Csongrád, Hungary, [2] University of Bath, Bath, Somerset, UK

It has often been suggested that removal of toxic small molecules from the cell is necessary to the survival of the organism. In case of toxic intermediate metabolites, prevention of accumulation helps averting cellular damage. One strategy to achieve this is the tight regulation of producing and consuming enzymes of the damaging metabolite. In yeast, it has previously been suggested that toxic intermediate metabolites have their enzymatic genes clustered on the chromosome to promote coregulation. However, a direct link between coregulation and metabolite toxicity remains unestablished. Here we focused on Escherichia coli to systematically test this theory by employing a unique toxicity prediction algorithm that has been specifically designed for this organism. We tested two possible strategies to achieve tight coregulation for metabolic genes of toxic intermediates: coregulation via shared regulators and allocation of genes within the same operon, a bacteria-specific mechanism of strong co-regulation. We show that while toxic intermediates have their neighbouring genes more often in the same operon than expected by chance, this effect is only characteristic to a handful of metabolic pathways. Furthermore, coregulation or mRNA-level coexpression of neighbouring genes occur with similar frequency in case of both toxic and non-toxic intermediates. Taken together, even in an organism with huge population size, we failed to find any signature of selection to specifically increased co-regulation of metabolic genes of toxic intermediate metabolites. Our result also suggest that evolution of genome organization and regulation of genes on a large scale is shaped by other forces than metabolite toxicity.

680B

Domestications, Sex and Stress effects on Adrenal Gene Expression and steroid hormone levels in Chicken

Amir Fallahshahroudi, Martin Johnsson, Per Jensen
Linköping University, Linköping, Sweden

Domestication is the evolutionary process where organisms genetically adapt to live in human desired conditions . Even though it appears, animals have been domesticated for diverge incentives, all of them have the common trait of being more docile and tame toward humans. In every studied species, being less fearful is complemented by decreased hypothalamic-pituitary-adrenal (HPA) axis reactivity in domesticates. We have already shown that the ancestral red jungle fowl (RJF) display more extreme behavioural and physiological response to restraint stress compared to its domesticated descendant, *White Leghorn* (WL). Changes in gene expression lead to dramatic evolutionary phenotypic changes. Aim of the study was to decipher the molecular basis of hampered fearfulness and physiological stress reactivity of domesticated chicken. The goal was achieved by studying adrenal gene expression at baseline and after restraint stress in domesticated *White Leghorn* and ancestral red jungle fowl. The magnitude of adrenal gene expression difference between the breeds was substantially larger than reported brain gene expression difference between domesticates and their wild ancestors. We expect that many of our observed DE genes between the breeds are not evolutionary or functionally important and their expression difference may mainly be explained due to random fixation of gene expression following a neutral model. Functional analysis showed a significant overrepresentation of genes related to "channel activity" pathways. Several GABA receptors, glutamate receptors, opioid receptors, serotonin receptor, dopamine receptor as well as genes coding various potassium, calcium and sodium channels were among the DE genes in mentioned pathway.

681C

Evolutionary effects of DNA methylation at the nucleotide, exon, and gene levels

Trees-Juen Chuang [1], Feng-Chi Chen[2], Huai-Kuang Tsai[3], Yi-Ching Chen[3], Tai-Wei Chiang[1], Jen-Hao Cheng[3], Zing Tsung-Yeh Tsai[3]

[1] Genomics Research Center, Academia Sinica, Taipei, Taiwan, [2] Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan, [3] Institute of Information Science, Academia Sinica, Taipei, Taiwan

DNA methylation at CpG dinucleotides is known to significantly increase the rate of cytosine-thymine transitions and thereby the level of sequence divergence. However, highly methylated genes were observed to evolve slowly. Therefore, we first asked whether DNA methylation is correlated with increased or decreased protein evolutionary rates. We considered the genic positions of coding exons, and suggested that the first exons appear more prone to the mutagenic effects, whereas the other exons are more influenced by the regulatory effects of DNA methylation. Furthermore, we showed that in mammalian exons, the correlations between DNA methylation and the conservation of individual nucleotides are dependent on the type of exonic sequence (coding or untranslated), the degeneracy of coding nucleotides, background selection pressure, and the relative position where the nucleotides are located. We demonstrated that the biological roles of DNA methylation can vary significantly within a short genomic distance, supporting a notion that not all of the methylated CpG dinucleotides are equally functional. Our results thus suggested that the "functional resolution" of DNA methylation may be finer than previously recognized. We also investigated the impact of pre-transcriptional DNA methylation, transcriptional transcription factor (TF) and post-transcriptional microRNA (miRNA) on the evolutionary rates of mammalian proteins, suggesting that the relative importance of these regulatory factors in determining the rate of mammalian protein evolution is: promoter methylation » miRNA regulation > gene body methylation > TF regulation. We indicated that promoter methylation and miRNA regulation have a significant dependent effect on protein evolutionary rates.

682D

The protective role of melanocytes in the evolution of modern humans: Comparison of the transcriptional profiles of melanocytes from dark and light skinned individuals following UVB exposure.

Saioa López [1,2], Isabel Smith-Zubiaga[2], Alicia García de Galdeano[2], María Dolores Boyano[2], Oscar García[3], Jesús Gardeazábal[4], Conrado Martinez-Cadenas[5], Arrate Sevilla[2], Neskuts Izagirre[2], Concepción de la Rúa[2], Santos Alonso[2]

[1] University College London, London, UK, [2] University of the Basque Country, Leioa, Spain, [3] Ertzaintza Forensic Unit, Erandio, Spain, [4] Cruces University Hospital, Barakaldo, Spain, [5] Jaume I University of Castellón, Castellón, Spain

Skin color is one of the most conspicuous examples of the biological adaptations to the environment that anatomically modern humans went through. To investigate the mechanisms by which melanocytes protect human skin from the damaging effects of ultraviolet-B radiation (UVB) we have analysed the whole-genome transcriptional profile of 12 cell lines of melanocytes, 6 from light and 6 from dark skinned individuals. We have investigated both the transcriptional profile at basal conditions and after UVB at different time points (6, 12 and 24 hours). We used SurePrint G3 Human GE Microarrays (Agilent), which include probes for around 28,000 genes and 7,419 long intergenic non-coding RNAs. Our results suggest that an interaction between ribosomal proteins and the P53 signaling pathway may occur in response to UVB in both dark and light melanocytes. Furthermore, we also observed that dark and light melanocytes show differentially expressed genes after irradiation, in particular at the first 6h after UVB. These are mainly associated with inflammatory reactions, cell survival or melanoma. Finally, the comparison of the transcriptional profiles between LM and DM under basal conditions, and the application of natural selection tests in human populations allowed us to support the significant evolutionary role of *MIF* and *ATP6V0B* in the pigmentary phenotype.

683A

Evolutionary conserved expression of neuroglobin in the CNS of mammals

Andrej Fabrizius [1,2], Thomas Hankeln[2], Thorsten Burmester[1]

[1] University of Hamburg, Hamburg, Germany, [2] University of Mainz, Mainz, Germany

Neuroglobin (Ngb) is a recently discovered vertebrate globin that is almost exclusively expressed in the nervous system. Ngb is highly conserved and occurs in all vertebrates, with the exception of Agnatha (hagfish and lampreys) and Chondrichthyes (sharks and rays). Despite 15 years of research, the function and even the main sites of expression of Ngb are still a matter of debate. To relate the postulated physiological function(s) of Ngb to its expression pattern, we have re-evaluated Ngb expression by bioinformatic analysis using publicly available transcriptome data (RNA-Seq). In the adult murine brain, we found the highest Ngb-mRNA expression within the hypothalamic region of the brain stem, which was up to 20 fold higher than in the mouse cortex or the total brain. Other mammals, such as prairie vole, dog, pig, sheep and cattle, also revealed elevated levels of Ngb expression in the hypothalamus compared to other brain regions. The high regional expression differences in the mammalian brain raise the question of distinct Ngb functions in highly and lowly expressing cell types. The hypothalamus connects the nervous and the endocrine system. It contains a number of small nuclei which control our sleep-wake cycle, hunger, thermoregulation, water balance etc.. Our results suggest the involvement of Ngb in a specific hypothalamic regulatory or metabolic pathway conserved in mammalian evolution.

684B

Early adaptive mutations restore the ancestral gene expression state during thermal stress adaptation

Alejandra Rodriguez-Verdugo [1], Olivier Tenaillon[2], Brandon S. Gaut[3]
[1] Swiss Federal Institute of Technology Zürich, Zürich, Switzerland, [2] INSERM, Universities Paris Diderot and Paris Nord, Paris, France, [3] University of California Irvine, Irvine, USA

An aspect of adaptation that remains largely unexplored is the temporal change of phenotypes during the adaptive process and the molecular mechanisms underlying these changes. Here we have focused on the phenotypic effects of early adaptive mutations during thermal stress adaptation. We explored two questions: 1) What are the molecular mechanisms underlying the large fitness advantage conferred by early adaptive mutations? 2) What is the phenotypic contribution of an early adaptive mutation compare to phenotypic variation accumulated during adaptation? We addressed these questions based on 114 populations of *Escherichia coli* evolved for 2000 generations at 42ºC. We focused on three mutations affecting the *rpoB* gene – which encodes the beta subunit of the RNA polymerase – that were driven to high frequency in early stages of adaptation in 12 populations. We measured their growth curves and gene expression (mRNAseq) at 42ºC, and compared them to the growth and gene expression of the ancestor at 37ºC and 42ºC. The three mutations changed the expression of hundreds of genes and conferred large fitness advantages through the restoration of global gene expression back towards a pre-stressed state. When we compared the phenotypic characteristics of one single mutant, *I572L*, to those of two high-temperature adapted clones with this mutation we observed that mutation *I572L* contributed to most of the expression changes while later mutations increased fitness but did not substantially change gene expression. We conclude that early mutations in a global transcriptional regulator cause changes in gene expression that are under positive selection.

685C

Tissue specificity evolution of protein coding genes in vertebrates

Nadja Kryuchkova [1,2], Marc Robinson-Rechavi[1,2]
[1] *University of Lausanne, Lausanne, Switzerland,* [2] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*

One of the major properties of genes is their expression pattern. While there have been recent studies of the correlation of expression between species for a few tissues, little is known about the evolution of tissue-specificity itself. There are several methods to measure this tissue-specificity. In this study we compare available methods and use them to study evolution.

Eight methods for tissue-specificity were analyzed on their robustness to the choice and number of tissues and to data normalization. For RNA-seq, the results of most methods, established for ESTs and Microarrays, depend on the data pre-processing. Yet tissue-specificity is better detected with RNA-seq than with Microarrays.

We show that new genes and genes with more paralogs tend to have more specific expression. Previous findings, that more broadly expressed genes evolve under stronger purifying selection, could be confirmed.

We found a major difference between the evolution of tissue-specificity in paralogs and orthologs. Orthologs are strongly conserved between all tetrapod species, with the lowest Pearson correlation between human and frog r=0.73. Paralogs in human show much lower conservation, with correlations of 0.60 for paralogs predating Primates. Moreover, tissue specific paralogs are often specific for different tissues. In general both for orthologs and paralogs the correlations get weaker with divergence time.

686D

Auxin Regulatory Module Evolution during C4 Plant Evolution

YAO-MING CHANG [1], Chun-Chieh Shih[2], Wen-Hsiung Li[1,3]

[1] Biodiversity Research Centre, Academia Sinica, Taipei, Taiwan, [2] Institute of Information Science, Academia Sinica, Taipei, Taiwan, [3] Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

Auxin, a plant phytohormone, plays an important role in vasculature development of vascular plants through the feedback regulation between auxin responsive factors (ARFs) and AUX/IAA proteins. In flowering plants, more than twenty genes in both ARF and AUX/IAA families have been identified but their regulatory relationships are still largely unknown. In this study, we analyzed three time-course transcriptome datasets from maize foliar (C4), husk (C3-like) and rice (C3) embryonic leaves and found that most ARF transcription factors are coexpressed with AUX/IAA genes in all three types of leaves but their co-expression patterns differ from each other, indicating that the auxin regulatory modules of leaf development have evolved during the evolution from C3 to C4 plants. In addition, we compared the promoter regions of each group of ARF homologs across seven C3 and C4 flowering plants and found that many of them have not been conserved across the species. That is, the ARF genes with different expression profiles among C3, C3-like and C4 leaves have been affected by changes in *cis*-regulation. Therefore, we propose that changes in ARF-AUX/IAA regulation have contributed to the evolution of vasculature structure from C3 to C4 leaves.

687A

Apple miRNAs and their role in Fire Blight resistance

Elzbieta Kaja [1], Michal Szczesniak[1], Timothy McNellis[2], Michael Axtell[2], Izabela Makalowska[1]
[1] Adam Mickiewicz University, Poznan, Poland, [2] Penn State University, State College, USA

Micro RNAs (miRNAs) are small, single stranded RNA molecules, which are key players in multiple biological processes in plants and animals such as: plant development, hormone signaling or stress response. In apple (Malus domestica), 200 microRNAs are known, which probably represent only a fraction of miRNAome diversity.
In the first part of our research we characterized miRNAs, which are specific for Gala apple scions grafted on four different rootstocks: B.9, G.30, M.27 and M.111, presenting diverse Fire Blight resistance. SOLiD sequencing of small RNAs has been performed, all the reads have been mapped to the apple genome and searched for conserved and apple-specific miRNAs. Performed analyses allowed us to extend the apple miRNA repertoire by 38 conserved and 78 novel, apple specific, miRNA. Performed analyses let us define four apple miRNAs potentially involved in fire blight resistance in apple trees: mdm-miR169a, mdm-miR160e, mdm-miR167b-g, and mdm-miR168a,b. These miRNAs are known to be involved in response to stresses across other plant species, usually by targeting stress response proteins. Our data suggests that apple microRNAs might be considered as regulators and markers of fire blight resistance.
The second part of our research is focused on the apple response to E. amylovora infection. Using Illumina sequencing method to identify miRNAs and their expression levels in apple leaves inoculated with bacteria, we identified 234 novel miRNA candidates. A few of them seem to be differentially expressed in response to inoculation or just leaf cutting, which may suggest their role in bacterial stress response in apple trees.

688B

Evolution of gene expression patterns in primate stem cells

Beate Vieth [1], Stephanie Wunderlich[2], Martin Kircher[3], Volker Kittke[1], Ulrich Martin[2], Wolfgang Enard[1]

[1] LMU Munich, Munich, Germany, [2] Hannover Medical School, Hannover, Germany, [3] University of Washington, Seattle, WA, USA

{INTRO} Induced pluripotent stem cells (iPSC) are nowadays central tool to investigate the molecular and cellular basis of differentiation processes and human diseases. Comparing iPSCs from different species allows to study the divergence and conservation of these processes. {METHOD + RESULTS} We used iPSCs from humans, a bonobo, a gorilla and a cynomolgus monkey and embryonic stem cells (ESCs) from humans and rhesus macaque to generate expression profiles by RNA-Seq (Wunderlich et al., 2014). We find that expression distances within species are very similar between ESCs and iPSCs, among different individuals and among different clones of the same individuals suggesting that non-genetic factors dominate expression differences within species. In contrast, gene expression distance between species is three-fold higher. To assess the amount of constraint acting on primate gene expression evolution, we analysed expressed pseudogenes and found that they evolve 3.5-fold faster than protein-coding genes. When comparing these patterns to RNA-Seq data generated from the same species in brain, liver, heart, kidney and testis (Brawand et al., 2011), we find similar magnitudes of expression divergence in these differentiated tissues, with testis as an exception that evolves significantly faster. Expression of pseudogenes evolves also in these tissues significantly faster than protein-coding genes. {CONCLUSION} In summary, we show that primate gene expression evolution evolves under considerable constraint in stem cells and tissues and that primate iPSCs are a promising system to study constraints and divergence of differentiation processes and human diseases.

689C

Epistatic interactions do not constrain the evolution of gene expression.

Gábor Boross, Csaba Pál, Balázs Papp
Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary

Changes in gene expression can have a large phenotypic impact and therefore the expression level of genes is often subject to selection. Recently, it has been proposed that epistasis between genes can influence the evolution of gene expression. According to the theory, pairs of genes where simultaneous deletion has a more detrimental effect than expected (negative epistasis) were hypothesized to evolve reduced expression noise to avoid concurrent low expression of both gene products. However, the core assumptions of this verbal theory remain untested. Here, we combine quantitative biochemical modelling with meta-analysis of functional genomic data to demonstrate that epistasis between gene deletions cannot generally predict fitness changes caused by decreased expression and, non-intuitively, even when decreased expression of the two genes show a negative epistasis, it does not aggravate the fitness cost of stochastic expression. Taken together, stochastic variation in epistatic partners is unlikely to drive noise minimization or constrain expression divergence on a genomic scale. Our results also offer an explanation of why epistatic interactions are poor indicators of drug synergy.

690D

Construction of transcriptome databases of the Antarctic native angiosperms have adapted to extreme environments

Yoonjee Kang, Hyoungseok Lee, Hyun Park, Jungeun Lee
Korea Polar Research Institute, Incheon, Republic of Korea

Land plants are often exposed to unfavorable conditions for growth. Extreme temperatures, drought, high salinity and high-UV radiance, are typical environmental stress factors that inhibit the growth and development of plant and these environmental stress can alter cellular structures and cause damage to physiological functions. The Antarctic is one of the toughest environment for life to survive. Because of the harsh conditions, an extremely small number of species has been survived and only two native angiosperms are present, *Colobanthus quitensis* (Kunth) Bartl. (Caryophyllaceae) and *Deschampsia antarctica* Desv. (Poaceae). Despite they have been studied as an extremophile that has succesfully adapted to marginal land, limited genetic research has focused on these due to the lack of genomic resources. Here, we present the results of deep transcriptome sequencing of D.antarctica and C.quitensis. Total sequence reads were assembled into 90,446 unigenes (average length: 929bp) for *D.antarctica* and 95,010 unigenes (average length: 1,300 bp) for *C.quitensis*. Assembled sequences were annotated based on homology to genes in multiple public databases. Differential expression analysis revealed that the lists of genes with significant different transcripts levels between in field-grown plants and in laboratory-grown plants. In the differentially expressed gene groups, the gene ontology terms: response to stimulus, response to stress, photosynthesis and carbohydrate metabolic process, were significantly enriched. Through the comparative analysis with the abiotic stress-transcriptomes of other species, we identified that the UV-B responsive gene sets, previous reported in model plants, are highly expressed in the field-grown plants exposed to high solar UV-B radiation in Antarctica.

691A

Confirmatory evidence of U1-dependent definition in D. melanogaster

Gildas Lepennetier, Francesco Catania
Institute for Evolution and Biodiversity, Münster, Germany

U1-dependent definition is a model of splice site recognition that unifies exon and intron definition. This model specifies how the interplay between splicing factors and cleavage/polyadenylation factors, which compete for access to overlapping or neighboring binding sites along nascent mRNAs, may potentially contribute to gene expression and gene architecture in eukaryotes. Here, we use the model organism Drosophila melanogaster to test and verify a number of this model's predictions.

692B

Relaxation or selection? A study of coding sequence and expression polymorphism in testis-specific genes in primates

Ayshin Ghalichi, Ezgi Ozkurt, Gulsah Merve Dal-Kilinc, Mehmet Somel
Middle East Technical University, Ankara, Turkey

In mammals, male reproductive characteristics are strongly correlated with inter-male sperm competition. In species where females mate promiscuously, males have evolved large testes and fast sperm, as in chimpanzees. In contrast, in species where females generally mate with a single male, males have smaller testes and slower sperm, as in humans and gorillas. Large testes and fast sperm could be driven by positive selection, whereas small testes and slow sperm could be outcomes of relaxation of constraints. A recent study on human non-synonymous polymorphism among tissue-specific gene sets found an intriguing trend: significantly higher frequencies of functional (possibly damaging) mutations in testis-specific genes, compared to other tissue-specific genes. In line with the above described theory, this result was interpreted as indication for relaxation of constraint in the human testis. However, positive or balancing selection on testis-related traits could also potentially explain the observed pattern. To address this, here we analyze genome-wide protein coding sequence polymorphism in two other great ape species: chimpanzees and gorillas. We further supplement our analysis with within-species gene expression data in humans, chimpanzees, and macaques. We expect to resolve the question regarding the role of selection versus relaxation in the maintenance of high levels of functional polymorphism in human testis genes.

693C

Amylase gene copy number variation and carbohydrate digestion – A postprandial study.

Katherine D. Brown [1], Sarah Berry[2], Peter Ellis[2], Dallas Swallow[1], Frances Bligh[3], Mark G. Thomas[1]

[1] University College London, London, UK, [2] King's College London, London, UK, [3] Unilever Discover Colworth, Bedford, UK

Copy number variation (CNV) in amylase genes may impact the ability to digest complex carbohydrates such as starch, the rate of glucose absorption, and the total energy obtained from dietary carbohydrate. Persistently elevated levels of postprandial glycaemia and insulinemia can result in the development of insulin resistance which leads to weight-gain and increased body fat percentage, which in turn cause predisposition to type-2 Diabetes mellitus.

In this clinical study we test whether CNV in the salivary amylase gene (AMY1) affects postprandial glycaemia and insulinaemia in healthy individuals (n = 41), or is correlated to obesity markers including Body Mass Index (BMI), total percentage body fat, waist-to-hip ratio and waist circumference (n=71). Finally, we investigate whether AMY1 CNV affects habitual dietary consumption of carbohydrate (n=78). So far, significant positive correlations have been found between AMY1 copy number (CN) and BMI (kg/m^2) and AMY1 CN and consumption of carbohydrates (%Total Energy), and both starch (%Total Energy) and sugars (%Total Energy).

694D

A conserved role of the Piwi/piRNA pathway in mammalian gene regulation

Daniel Gebert [1], René Ketting[2], Hans Zischler[1], David Rosenkranz[1]
[1] Institute of Anthropology, Johannes Gutenberg-University, Mainz, Germany, [2] Institute of Molecular Biology IMB, Mainz, Germany

Piwi-interacting (pi-) RNAs constitute a class of metazoan germline-expressed small non-coding RNAs that associate with Piwi clade Argonaute proteins. Together they form the core of the piRNA-induced silencing complex, which selects target nucleic acids by sequence complementarity. Piwi proteins and their guiding piRNAs have been mainly implicated in silencing of transposable elements (TEs), especially in mammalian gametogenesis, during which global DNA demethylation leads to a reactivation of TEs that pose a threat to genome integrity. However, evidence for functions beyond transposon silencing is mounting. Recent studies, including our own, suggest a conserved role for the Piwi/piRNA pathway in mammalian gene regulation. We found that both sense and antisense piRNAs from testes derive from protein-coding genes in pig, mouse and human, while exhibiting features showing that they originate from the Piwi/piRNA-mediated post-transcriptional silencing pathway, commonly referred to as ping-pong cycle. Targeting and processing of homologous genes by piRNAs in these species, as well as a general bias for nucleic acid binding property and nuclear localization in target gene products, indicate a non-random selection of transcripts and a conserved function. Moreover, our results suggest a mechanism by which genes are specified for Piwi-mediated regulation. Genes and pseudogenes that lie on the opposite strand within piRNA clusters, genomic loci from which piRNA precursors are produced, provide the sequence information for the generation of primary antisense piRNAs directed against gene transcripts, initiating ping-pong cycle processing. These findings demonstrate that the mammalian Piwi/piRNA pathway is more versatile than commonly perceived.

695A

Evolutionary rewiring of the human regulatory network by waves of genome expansion

Davide Marnetto [1], Ivan Molineris[1], Federica Mantica[1], Elena Grassi[1], Igor Pesando[3,4], Paolo Provero[1,2]

[1] Department of Molecular Biotechnology and Health Sciences, Universita` degli Studi di Torino, Turin, Italy, [2] Center For Translational Genomics and Bioinformatics, San Raffaele Hospital, Milan, Italy, [3] Department of Physics, Universita` degli Studi di Torino, Turin, Italy, [4] I.N.F.N., Section of Torino, Turin, Italy

The emergence of de novo genomic regions through the insertion of transposable elements (TEs), or due to extensive sequence divergence from one phylogenetic lineage to another, is thought to be an extremely important evolutionary mechanism. Relying on genome wide alignments provided by UCSC and a parsimony algorithm, we estimated the age of each region of the human genome. We then studied the age distribution of several types of functional regions.

We confirm previous results on coding regions, such as correlations between evolutionary age, role of the gene product and gene expression. The age distribution of non-coding sequence and in particular regulatory elements reveals the extensive use of newly formed genomic sequence in the evolutionary rewiring of regulatory networks. The binding sites of many transcription factors have emerged in waves driven by the expansion of transposable elements. For some transcription factors it is possible to discern successive waves of binding site expansion, creating target genes associated to different biological processes. For example TCF3, SIX5 and HEY1 show old sites targeting genes enriched in chromatin organization processes and new sites involved in neural development categories, including axonogenesis and neuron differentiation.

696B

The coevolutionary dynamics of cell type transcriptome: explaining the so-called "species signal" in transcriptome data

Cong Liang [1,3], Jacob Musser[1,2], Gunter Wanger[1,2]

[1] Systems biology institute, Yale University, West Haven, CT, USA, [2] Department of ecology and evolutionary biology, Yale University, New Haven, CT, USA, [3] Program of computational biology and bioinformatics, Yale University, New Haven, CT, USA

Differential gene expression analysis has been a powerful tool for understanding cell type specific gene functions. In cross-species cell type transcriptome comparison, if a pair of cell types arose prior to the most recent common ancestor of the compared two species, one would expect that transcriptomes of homologous cell types are more similar to each other than cell types from the same species. However, in may studies it is observed that transcriptomes of cell types clustering more frequently by species (so-called species signal) than by homologous cell types even though the cell types are older than the respective species. We proposed that the so-called species signal arises from the co-evolutionary dynamics of cell types and thus is a consequence of the fact that cell types from the same species often are not fully individualized and experience correlated evolutionary change. These correlated changes will be interpreted as shared derived similarities leading to a phylogentic tree that is dominated by species signal. In this contribution, we developed a stochastic model for the co-evolution of cell type transcriptome within species. Our model illustrated a phase transition between the two grouping patterns in parameter space. The co-evolution parameter serving as a measure of individuality between cell types could be inferred from multiple transcriptome datasets with maximum likelihood methods. By modeling the dynamics of cell type transcriptome co-evolution, we probed how co-evolutionary signal of cell type transcriptomes influenced the phylogentic reconstruction of cell type history. This addressed an important question of individualization of phenotypic characters.

697C

The Caliphoridae family as a model to understand the evolution of parasitism

Gisele Cardoso [1], Raquel Monfardini[1], Marco Marinho[2], Ana Azeredo-Espin[3], Tatiana Torres[1]
[1] Sao Paulo University-USP, Sao Paulo, Sao Paulo, Brazil, [2] Sao Paulo University-USP, Ribeirao Preto, Sao Paulo, Brazil, [3] Campinas State University-Unicamp, CampiCampinas State University-Unnas, Sao Paulo, Brazil

Flies of the Calliphoridae family feed in different sources such as living tissues of vertebrates host (parasitism) or decaying organic matter (saprophagous). The presence of such contrasting habits allows the comparison among closely related species and study the evolution of parasitism. As a first step, we selected eight candidate genes and compared their expression in larvae and adult flies.

Comparisons within larvae demonstrated high expression conservation in seven genes, indicating a purifying selection signal. In adults, differences in expression correlated with the divergence among species possibly due to accumulation of neutral mutations over time.

Only the candidate *Malvolio* had a different expression pattern between habits in all comparisons. Variations in regulatory regions may contribute to the differences in its expression levels. Furthermore, topologies generated using expression data supported that feeding behavior was the major determinant of *Malvolio* expression divergence. Furthermore, we sequenced part of the coding region of five genes and submitted them to neutrality and selection tests. With this, we observed a signal of selection only in *Malvolio*.

Our results motivated us to pursue gene expression comparison on a genome-wide scale using RNA-seq data. Based on a phylogenetic method and searching for specific genes by blast, we found 3464 ortholog transcripts among four species. From this set, 55 transcripts were differently expressed between habits and are potential candidates. These genes are involved in pathways such as toxin metabolism and proteolysis. We will perform behavior assays to understand the role of these genes in feeding habit focusing mainly in *Malvolio*.

698D

Neutral and adaptive genetic characterization of Senegalese cultivated pearl millet germplasm [*Pennisetum glaucum* ssp. *glaucum* (L.) R. Br.]

Oumar Diack [1,2], Adeline Barnaud[3,4], Leila Zekraoui[3,4], Jésaelle Piquet[4], Marie Couderc[4], Hamidou Tall[5], Ousmane Sy[5], Amadou Fofana[5], Mame Codou Guèye Cissé[1,6], Cecile Berthouly-Salazar[3,4], Yves Vigouroux[3,4], Diaga Diouf[2], Ndjido Ardo Kane[1,3]

[1] National Research Laboratory for Plant Production, Senegalese Institute of Agricultural Research (ISRA), Dakar, Senegal, [2] Plant Biotechnology Laboratory, Department of Plant Biology, Faculty of Science and Technology, University Cheikh Anta Diop (UCAD), Dakar, Senegal, [3] Laboratory on the adaptation plant and associated microorganisms to environmental stresses (LMI LAPSE), Dakar, Senegal, [4] UMR DIADE, Institute of Research for Development (IRD), Montpellier, France, [5] National Center for Agricultural Research (CNRA), Senegalese Institute of Agricultural Research (ISRA), Bambey, Senegal, [6] Regional Study Centre for the Improvement and Adaptation to Drought (CERAAS), Senegalese Institute of Agricultural Research (ISRA), Thies, Senegal

Pearl millet is a widely grown cereal in Sub-Saharan Africa, where it is the staple food for 50 million individuals. For this purpose, understanding the structure of the genetic resources available is necessary to manage, conserve and implement breeding programs to sustain regional agriculture production for food security.

This study aimed at determining the diversity and genetic structure of cultivated pearl millet in Senegal and the contribution of their wild relatives. The variability of 465 landraces distributed across Senegal including 353 early flowering and 112 late flowering varieties, was assessed using 12 microsatellites and two genes, PgMADS11 and PHYC, linked to flowering traits. Our dataset of neutral markers was combined with previously published data of wild pearl millet populations across West Africa to estimate the importance of gene flows between cultivated and wild populations. Principal Components Analysis and Bayesian approaches revealed that the Senegal pearl millet germplasm has a relatively high neutral diversity. We discuss our results in light to early vs. late flowering varieties and the importance of wild-to-crop gene flow for pearl millet diversity.

The study provides the first assessment of cultivated pearl millet in Senegal using genetic markers and will be important for advancing breeding programs.

In prospect, we aim to assess the importance of wild-to-crop introgressions as potential source of adaptations in cultivated pearl millet. We assume that wild genes conferring adaptations to harsher conditions have introgress into cultivated populations in regard to ongoing climate changes and/or anthropic selection.

699A

Evolution of parasitism and insecticide resistance in Oestroidea (Diptera: Calyptratae)

Tatiana Torres, Gisele Cardoso, Sophie Tandonnet, Raquel Monfardini, Marina Deszo
University of Sao Paulo, Sao Paulo, SP, Brazil

Flies of the superfamily Oestroidea are characterized by the ability of their larvae to develop in animal flesh. While all species of the Oestridae family are obligate parasites, members of the Calliphoridae family can be divided into three groups based on their larval feeding habits: saprophagy, facultative ectoparasitism, and obligate parasitism. The range of life-history strategies and the appearance of obligate parasitism in at least five independent occasions in Oestroidea make it an ideal model system for the study of the convergent evolution of phenotypic traits. We tested the hypothesis that the same genetic mechanisms underlie the evolution of two phenotypic traits, feeding preference and insecticide resistance, by comparing gene expression profiles among two species of the Oestridae family and five species of the Calliphoridae family. In a candidate gene approach, we identified a single gene with expression differences among flies with different feeding preferences. We also identified that resistance to insecticides has repeatedly evolved by changes in amino acid residues in a single gene, the alphaE7 gene, encoding the esterase 3. The comparative analysis of genome-wide expression profiles and the characterization of the genome of one parasitic Calliphoridae species is still underway and will allow us to test if the convergent evolution of feeding preferences involves the same genes or similar metabolic pathways and molecular functions.

700B

Comparative Study of Gene Expression Patterns in Skins between Humans and Other Primates

Nami Arakawa, Yohey Terai, Yoko Satta
SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan

When humans are compared with other primates, skins show extensive differences in its morphology and physiology. For example, it is obvious that humans possess little hair on their body surface. One of major physiological roles of hair on a body surface is retaining moisture of the skin. But the human skin also keeps moisture, suggesting some factors other than hair to retain moisture. Aquaporins, cell membrane channel proteins through which water permeates the lipid bilayer, may be one of such factors. *Aquaporin-3* (*AQP3*) is expressed in epidermal cells and knock out experiment of this gene in mice implies that this gene is responsible for maintaining skin moisture. Based on these facts, we predict some specific feature in human *AQP3* sequence or gene expression pattern compared with other primates should be responsible for retaining water in the skin without hair.

Using the 53 vertebrate *AQP3* nucleotide sequences from NCBI database, the phylogenetic analysis revealed the human CDS (coding sequence) has no specific features in the amino acid substitution mode and rate. The results of quantitative PCR (qPCR) until now showed no significant differences in the expression level of *AQP3* between humans and the other primates (chimpanzees, gorillas, and orangutans), although differences among individuals within a species are large.

Now I have been comparing comprehensively the expression levels of genes in the skin between humans and the other primates. This approach may reveal genes responsible for maintaining skin moisture in humans, in spite of their little hair.

701C

Translational efficiency is enhanced by trans-splicing in nematodes

Yu-Fei Yang [1], Xiaoqing Zhang[1,2], Qiushi Sun[3], Wenfeng Qian[1]

[1] Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China, [2] University of Chinese Academy of Sciences, Beijing, China, [3] Beijing Jiaotong University, Beijing, China

About 70% genes are trans-spliced to a 22 nucleotide spliced leader (SL) in Caenorhabditis elegans, in which process the 5' end of the native 5' untranslated region (UTR) is replaced by the trans-spliced leader. Trans-spliced leader 1 (SL1) is used for the first gene in an operon or genes not in operons, whereas SL2 is spliced to downstream genes in an operon. While it is widely accepted that SL2 is used to produce functional monocistronic mRNAs from polycistronic transcripts, the function of SL1 is still unclear. Because 5' UTR is important for protein translational regulation, we hypothesized that SL1 could enhance translational efficiency by replacing 5' UTR. To test the hypothesis, we first calculated the average ribosome density (i.e., translational efficiency) for each gene from the whole genome ribosome profiling data, and found that translational efficiencies of SL1 trans-spliced genes are significantly higher than those of non-trans-spliced genes, even after excluding the confounding factors, such as Kozak consensus sequence and codon usage bias. We then compared the translation efficiencies of one-to-one orthologous genes in four nematode species (C. elegans, C. briggsae, C. remanei and C. brenneri), and found that a gene with SL1 trans-splicing tends to have higher translational efficiency than its orthologous gene wihout trans-splicing. Furthermore, combining with experimental stragies such as sucrose density gradient centrifugation and qPCR, we found that the proportion of SL1 trans-splicing transcripts of a gene is correlated with its translational efficiency. In conclusion, SL1 trans-splicing provides an additional mechanism to regulate protein translational efficiency.

702D

The genomic landscape of position effects on gene expression and expression noise in yeast

Xiaoshu Chen, Jianzhi Zhang
University of Michigan, Ann Arbor, MI, USA

Position effect is a term coined by Dobzhansky in 1936 to refer to the phenomenon that the expression of the same gene depends on its chromosomal location. Although the position effect has been extensively studied in relation to heterochromatin, the genomic landscape of position effect, especially on expression noise, is unknown. Using four different promoters and a small number of genomic positions, we first showed that position effect appears to be insensitive to the type of promoter used. We then created heterozygous yeast strains by individually replacing one allele of each of ~500 endogenous genes with a green florescence protein (GFP) gene controlled by a relatively strong promoter. High-throughput flow cytometry was used to examine the position effects on gene expression and expression noise. The coefficient of variation across all examined loci is 0.39 for expression level and 0.69 for expression noise. Several histone modifications are correlated with the position effects on gene expression and expression noise, and loci engaging in long-distance intrachromosomal interactions tend to have similar position effects. Supporting Batada and Hurst's hypothesis, we found that essential gene clusters are formed in regions of the genome that have intrinsically low noise. However, contrary to their hypothesis, these regions do not have low nucleosome occupancy. Instead, we found them to be enriched with H3K4me3 histone modification. Taken together, our results suggest the important role of histone modifications in shaping position effects and genome organization.

703A

Suboptimal transcriptional response in genetically perturbed yeast cells

Károly Kovács [1], Zoltán Farkas[1], Dorottya Kalapis[1], Patrick Kemmeren[2], Frank C. P. Holstege[2], Richard Notebaart[3], Csaba Pál[1], Balázs Papp[1]

[1] Biological Research Centre Hungarian Academy of Sciences, Szeged, Hungary, [2] Molecular Cancer Research, University Medical Center Utrecht, Utrecht, The Netherlands, [3] Radboud University, Nijmegen, The Netherlands

Microbial cells frequently change their gene expression programs as a response to both environmental and genetic perturbations. However, the extent to which transcriptional changes following a genetic perturbation are adaptive, neutral or deleterious is still poorly understood.

A recent large-scale dataset, created by the Holstege lab, allowed us to investigate this question in a systematic manner. This dataset contains information on genome-wide mRNA expression changes for 1484 viable single-gene knock-out strains of Saccharomyces cerevisiae. We found that the number of up- or downregulated genes show strong positive correlation with both the fitness contribution and the degree of pleiotropy of the deleted gene. Two patterns indicate that this is unlikely to reflect extensive adaptive reprogramming specific for the perturbed gene. First, functional connections are rare between the deleted and up- or downregulated genes. Second, by overlaying the gene expression data on a genetic interaction map of yeast, we demonstrate that most expression changes are unlikely to provide compensation following gene deletion. Taken together, gene-specific adaptive expression changes appear to be rare. Finally, we speculate that some of the expression changes are detrimental to fitness, implying that the fitness effect of gene deletion can be indirectly caused by subsequent expression misregulation.

705C

Expression of sex-linked *inhibitor of apoptosis* alleles in Mountain Pine Beetle (*Dendroctonus ponderosae*)

Linda Horianopoulos, Celia Boone, Brent Murray
University of Northern British Columbia, Prince George, BC, Canada

Mountain pine beetle (*Dendroctonus ponderosae*) is a major forest pest endemic to Western North America. In recent outbreaks, its range has expanded northeastwardly in Western Canada and it has been found in a novel host, the jack pine. An EST library was screened for polymorphic gene-linked microsatellites to identify candidate genes for adaptive variation. A polymorphic microsatellite was found in the coding sequence of the sex-linked gene for *inhibitor of apoptosis* (*IAP*). When compared to neutral microsatellite markers using outlier analysis of Fst values, the neo-X specific alleles of *IAP* showed a strong signature of positive selection. Thirteen different neo-Y alleles have been identified across the North American range compared to eight neo-X alleles. Sequence differences between neo-X and neo-Y alleles are consistent with neutral evolution suggesting that the neo-Y allele may not be under functional constraints. Based on a comparison of levels of genomic and cDNA from *D. ponderosae* larvae, expression of the neo-Y alleles is reduced compared to the neo-X alleles. There is no difference in the overall levels of expression of *IAP* between males and females suggesting that the neo-X allele in males is compensating for the reduced expression of the neo-Y alleles. Furthermore, the reduced function of the neo-Y alleles of *IAP* suggested by both sequence differences and its lower levels of expression may foster a highly selective environment for differences among neo-X alleles.

706D

Modular regulatory reorganization of the human brain transcriptome during the perinatal transition.

Jimena Monzón-Sandoval [2], Atahualpa Castillo-Morales[2], Araxi O Urrutia[2], Humberto Gutierrez[1]
[1] University of Lincoln, Lincoln, UK, [2] University of Bath, Bath, UK

During early development of the nervous system, gene expression patterns are known to vary widely depending on the specific developmental trajectories of different structures. On the other hand, the brain transcriptome is organized into discrete clusters or regulatory networks of co-expressed genes with related functions. Whether the observed variations in gene expression profiles during normal development reflect corresponding changes in the level of expression of existing gene networks or the potential reassembly of new regulatory interactions is not known. By examining existing expression data derived from the developing human brain, here we show that postnatal gene expression profiles do not correlate with those of foetal stages. This sharp division between foetal and postnatal profiles is not the result of sudden changes in level of expression of existing gene networks. Instead we demonstrate that the perinatal transition is marked by the widespread regulatory rearrangement within and across existing gene clusters, giving rise to the emergence of new functional groups. This rearrangement is organized into discrete blocks of genes targeting virtually unique sets of biological functions. Our results demonstrate an acute modular reorganization of the regulatory architecture of the brain transcriptome occurring at birth reflecting the reassembly of new functional associations potentially required during the transition from prenatal to postnatal brain development.

707A

Consequences of positive selection at lipid carrier *ABCA12.*

Roberto Sirica [1], Lucia De Martino[1], Dario Antonini[1], Gennaro Andolfi[1], Ombretta Guardiola[1], Gabriella Minchiotti[1], Caterina Missero[2], Qasim Ayub[3], Chris Tyler-Smith[3], Vincenza Colonna[1]
[1] National Research Council, Naples, Italy, [2] University of Naples Federico II, Naples, Italy, [3] Wellcome Trust Sanger Institute, Hinxton, UK

ABCA12 is a lipid carrier protein expressed in keratinocytes. It is essential for forming the skin lipid barrier and is down regulated by UV radiation. Mutations in *ABCA12* causes a rare recessive disorder characterized by accumulation of intracellular lipids, disruption of lipid barrier and formation of skin scales (Ichthyosis). We found a signal of positive selection in non-Africans associated with the derived allele of an intronic variant (rs10180970). We now aim to elucidate the environmental pressure determining the selective event and investigate possible connections with ichthyosis.

Data from 12 ancient *Homo sapiens* shows that the derived allele at rs10180970 is (i) observed for the first time 45 kya in the heterozygosis in an individual from Asia that shares a 27 kb haplotype with modern Asians and Europeans; (ii) in homozygosis in ten more recent (< 5.8 kya) individuals from Asia and Europe. Among other hominins, one Neanderthal and one Denisovan from > 50 kya have the ancestral homozygous genotype.

rs10180970 is the most differentiated variant between Africans and non-Africans in the whole gene. Preliminary data suggests an increasing gradient of gene expression between populations significantly associated with genotype at rs10180970. Ichthyosis can be seen as an extreme of this gradient therefore we are investigating a possible role of rs10180970 in the disease. Last ten years literature on coding region analysis indeed shows a variable penetrance for this disease and cases of ichthyosis in hemizygosis, suggesting a possible role of the non-coding part of the gene.

708B

Genome rearrangements in Drosophila are constrained by regulatory interactions

Raquel S. Linheiro, Dave T. Gerrard, Mar Marzo, Casey M. Bergman
University of Manchester, Manchester, UK

Understanding whether genome rearrangements impact function at the molecular or organismal levels remains an open question in genome biology. The genomic regulatory block hypothesis proposes that regions of conserved genome order between species (microsyntenic regions) are maintained because of long-range interactions between cis-regulatory modules (CRMs) and their target genes, however a formal test of the key assumption of this hypothesis is lacking. Extremely high rates of genome rearrangement and a wealth of comparative and functional genomic data in the genus Drosophila offer an excellent model system to test the hypothesis that cis-regulatory interactions constrain genome rearrangement. Here we create detailed maps of microsyntenic regions conserved across nine species in the genus Drosophila to assess the conservation >1,700 functionally-characterized CRMs with known target genes. We find a significant excess of CRMs to be located in the same microsyntenic region as the transcription start sites (TSSs) of their target gene, regardless of the definition of microsynteny used. We show that the distance from the edges of microsyntenic regions to the nearest TSS matches that of known CRM-target gene interactions, thus allowing us to define the "genomic radius" of CRM action in the Drosophila genome. Additionally, we show that breakpoints that disrupt microsyntenic regions are enriched at the boundaries of topologically associating domains (TADs) that have measurably lower levels cis-interactions. Our findings provide support for a crucial assumption of the genomic regulatory block hypothesis and establish that functional constraints on genome rearrangements are linked to the regulatory landscape of the genome.

709C

Finding enhancers and their target genes in zebrafish by looking for long range evolutionary constraints in fish evolution

Yves Clement, Hugues Roest Crollius
Institut de Biologie de l'ENS, Paris, France

While finding long-range regulatory regions (or enhancers) is trivial, finding which genes they regulate is more challenging. For example, considering the nearest gene as the target gene can be problematic as enhancers can regulate genes located several megabases away.

We previously developed a method to 1) identify putative enhancers in genomes by looking for conserved regions in multiple alignments and 2) identifying their target genes by looking for conserved enhancer - target genes association in multiple species by computing an association score. The hypothesis being that the enhancer - target gene link, if functional, will be conserved by natural selection. We applied this method to zebrafish, an ideal species as it is a model organism for for vertebrate development. We used custom-built genome-wide multiple alignments 7 well sequenced and reconstructed fish genomes. We found almost 150 000 putative enhancers evolutionary linked with at least one gene. While the majority are nearest genes, at least 17 000 of them are not the nearest gene, which demonstrate the value of such method. We found that enhancers overlap histone marks associated with development and that this overlap increases with the enhancer - target gene association score, hinting that these enhancers play an important role in fish development. Finally, comparing these results with others obtained in human provide a unique opportunity to study the evolution of gene regulation in vertebrates.

710D

Phylogenetic modeling of codon usage preferences.

Simon Laurin-Lemay [1], Hervé Philippe[1,2], Nicolas Lartillot[1,3], Nicolas Rodrigue[4]
[1] Robert-Cedergren Center for Bioinformatics and Genomics, Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, Canada, [2] Theorisation et de Modelisation de la Biodiversite, Station d'Ecologie Experimentale du CNRS, Moulis, Ariege, France, [3] Laboratoire de Biometrie et Biologie Evolutive, UMR CNRS 5558, Universite Lyon 1, Lyon, Rhone, France, [4] Department of Biology, Carleton University, Ottawa, Ontario, Canada

Mutation-selection models of codon substitution have attracted increased interest in recent years, as have the important roles of synonymous codon usage in molecular evolution. In a seminal paper from 1998, Halpern and Bruno provided the theoretical developments for a model that assigns site-specific fitness parameters to each sense codon; in practice, however, the authors simplified their model to one where fitness parameters are equivalent for all codons encoding the same amino acid, disabling the ability of the model to detect selection on synonymous mutations on a site-by-site basis. A variety of other mutation-selection models have since been proposed, some of which have been implemented, while other remain hypothetical. Here, we have implemented a large set of mutation-selection models, including all of those discussed in the literature to date. Some of these account for global codon fitness parameters with site-specific amino acid fitness parameters. Others have site-specific codon fitness parameters. Through simulations, and through the analysis of several real data sets, we report the behavior of these models, and discuss their strengths and weaknesses. Our current finding call for caution in the use of these models for measuring the strength of selection on codon usage.

711A

Paralogous genes in Teleost: Relationships among divergence, similarity and function

Elena Sarropoulou [1], Daniel Garcia de la serrana[2]

[1] Institute of Marine Biology, Biotechnology and Aquaculture, Heraklion, Greece, [2] University of St Andrews, St Andrews, UK

The present study investigates in paralogous genes within teleosts. Teleosts are of particular interest since they experienced an additional third-round (3R-) whole genome duplication (WGD), which is estimated to have occurred 320-400 million years ago. Genome duplication and thus paralogous genes are considered to be one of the major driving forces for speciation and diversification. Today, the study of paralogous genes in teleost is the focus of rising attention as steadily genomes of several teleost species are sequenced, which facilitate to a great extend retrieval and evaluation of paralogous genes. In order to distinguish true paralogs from alleles, splice variants and copy number variations paralogs can be evaluated by comparative genome mapping approach and by phylogenetic analysis as an ultimate proof. Further investigation by gene structure, synteny and gene expression analysis may shed light on to their evolutionary history. In the present study we show the power of database searches in order to identify putative paralogs, even though their genomic sequences are not available either because the genome is not sequenced yet, or because they are located in regions of unknown (gap) or not well resolved nucleotides (N). On the other hand we investigate paralogous genes with significant differences in expression which may be in turn important to specific physiological processes and / or adaption. Finally we systematically investigated the relation between their taxonomic position, their similarity degree among homologs as well as their functional divergence.

712B

Investigating genetics and epigenetics in the evolution of human adipocytes using iPSCs

Sasha Makohon-Moore, Greg Wray
Duke University, Durham, NC, USA

Understanding the evolutionary basis and molecular underpinnings of uniquely human traits has long been hampered by the limited ability to perform experiments on and access to samples from our closest relatives. The development of induced pluripotent stem cell (iPSC) lines provides us with virtually unlimited samples and allows, for the first time, precise experimental control in order to understand the relative roles of genetic, epigenetic and environmental effects. By using iPSCs from humans and chimpanzees, we can perform genome-wide gene expression measurements and regulatory-element sequencing. This facilitates the characterization and identification of genes whose expression has changed as a result of epigenetic modifications, genetic modifications or both. The ability to differentiate iPSCs into a variety of cell types allows for the characterization of specific cellular traits and responses to environmental stimuli. In particular, focusing on adipocytes, the key component cell of fat, allows for the identification of differences in molecular, cellular and metabolic traits. Changes in these traits may have been significant during human origins as shifts in diet and metabolic function were of particular importance.

713C

Early development and transgenesis in the Midas cichlid (Amphilophus spp.) - a new model system for evo-devo research

Maggie Sefton [1,2], Yipeng Liang[1], Claudius Kratochwil[1], Axel Meyer[1,2]
[1] University of Konstanz, Konstanz, Germany, [2] IMPRS for Organismal Biology, Konstanz, Germany

Cichlid fish of the Midas species flock (Amphilophus spp.), found in Central American crater lakes, are famous for their rapid rates of diversification and for being one of few definitive examples of sympatric speciation. There are numerous studies on the ecology and morphology of these species, and high-quality genomic resources are available; however, it was not until recently that a detailed description of early ontogeny in Midas cichlids was published. Recently, we established a developmental staging system in which we described the first seven days post-fertilization in A. xiloaensis (Kratochwil et al. 2015. BMC Dev.). We defined homologous stages both to teleost models and to other cichlid species, highlighting and discussing key differences such as early chromatophore patterns and the presence of adhesive glands. This staging system will allow researchers to perform comparative developmental studies and will serve as a resource for functional experiments, including transgenesis. Midas cichlids possess life history traits - substrate brooding, large amounts of eggs, and slow rates of early development - that make transgenic experiments feasible. As a first effort in this direction, we performed Tol2-mediated transgenesis in Midas cichlids. This method allows us to examine the activity of ubiquitous and cell-type specific enhancers and promoters, by tracing Green Fluorescent Protein (GFP) expression throughout early development. The Midas cichlid, with its available genomic resources, and well-studied ecology and evolutionary history, makes it an excellent novel model system for evolutionary-developmental research.

714D

EVOLUTION, EXPRESSION AND REGULATION OF THE METAL RESPONSE IN AMPHIOXUS

Lea Kablik [1], Christopher Materna[1], Paul Shin[2], Thomas Sorger[1]
[1] Roger Williams University, Bristol, Rhode Island, USA, [2] City University of Hong Kong, Kowloon, SAR, Hong Kong

BACKGROUND. Lancelets (*Branchiostoma* spp.) occupy a basal position in chordate evolution and their shallow, coastal habitat is highly susceptible to the effects of ocean warming, hypoxia and pollution. MTF1 (metal-activated transcription factor 1) mediates cellular responses to hypoxia as well as trace metals in vertebrates and invertebrates. We investigated (a) the evolution of this protein in three Amphioxus species and (b) its role in the metal response of the Asian lancelet, *B. belcheri*.

RESULTS. (a) MTF1 in *B. belcheri*, *B. lanceolatum* and *B. floridae* has 6 C2H2-type zinc fingers, the same domain architecture found in vertebrates. Bayesian phylogeny confirmed the basal position of cephalochordates in the chordate lineage and indicated that the MTF1 gene in the Asian lancelet is ancestral to the genes in the Mediterranean and Florida lancelets. (b) Relative to actin, mRNA levels for Mth were significantly increased by Cd (35-fold) and Ni (24-fold) ($p < 0.05$). MTF1 transcripts were also increased by Cd (75-fold) and Ni (22-fold) ($p < 0.005$). The order of activity of these metals, Cd > Ni > Cr = Control, correlates with the predicted ability of their hydrated ions to compete for a zinc binding site ($p < 0.01$).

DISCUSSION. A metal-induced increase in the expression of MTF1 itself, as well as Mth, has previously been observed in zebrafish cells (Cheuk et al 2008). Since neither the lancelet nor zebrafish MTF1 promoter has a conserved metal response element (MRE), the upregulation of this transcription factor may instead reflect an increase in mRNA stability.

715A

Decoding the global gene expression reprogramming in genetically perturbed yeasts

Djordje Bajic [1,2], Balázs Papp[1], Juan F. Poyatos[2]
[1] Biological Research Center, Szeged, Hungary, [2] National Biotechnology Center, CSIC, Madrid, Spain

According to the classical paradigm, the cellular response to perturbations consists of specific expression changes in a limited set of functionally related genes. This view has been challenged with the application of genome-wide technologies. Experiments monitoring responses to environmental perturbations revealed that expression changes usually involve a large number of functionally heterogeneous genes. Whereas such environmentally-induced global patterns are being characterized in recent years, global gene expression reprogramming upon mutations remains largely unexplained. Firstly, to what extent is reprogramming in response to deletions part of the problem (i.e. a global mutation-induced imbalance) or rather part of the solution, contibuting to compensate for the defect? To address this question, we firstly identify recurrent gene expression programs in response to 1500 single gene deletions in yeast. We integrate several computational and statistical approaches to characterize these changes. Our results indicate that biologically meaningful rationale can be found for a significant fraction of the observed reprogramming. Most of the observed changes seem to be due to imbalance cascades. However, the similarity in gene expression profiles between specific genetic and environmental perturbations also suggests a frequent evolution of mutational robustness as a by-product of environmental adaptation. Overall, our results anticipate complex gene-environment interactions and deepen our understanding of the evolution gene expression.

716B

The population genetics of the duffy locus in humans

Kimberly McManus [1], Omar Cornejo[2], Martin Sikora[3], Carlos Bustamante[1]
[1] Stanford University, Stanford, CA, USA, [2] Washington State Universty, Pullman, WA, USA, [3] Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark

The human DARC (Duffy antigen/chemokine receptor) gene encodes a membrane-bound chemokine receptor crucial to infection of red blood cells by Plasmodium vivax, a widespread and chronic form of malaria. This gene's three major allelic classes (FY*B, FY*A, FY*O) have been highly studied due to their clinical implications and spatial localization; however, little recent research has looked in-depth at the population genetics of this locus.

This is a particularly interesting opportunity, as there is strong biological evidence of a selective sweep at this locus, but very little evidence via typical genomic methods of inferring selective sweeps. The FY*O allele is thought to protect against P.vivax infection and is near fixation in sub-Saharan Africa, but absent from the rest of the world. The FY*A allele may protect slightly against P.vivax and is very common in Asia, while the FY*B allele is ancestral and located primarily in Europe.

We investigate the genetic structure of this locus in sequence data of diverse populations, estimate the TMRCA of the derived FY*O and FY*A mutations, estimate the selection coefficient of the FY*O mutation, and analyze this locus in great ape sequence data.

717C

Evolution of chromate resistance determinants in bacteria

Adriana Julian-Sanchez, Hector Riveros-Rosas
Depto. Bioquimica, Fac. Medicina, Universidad Nacional Autonoma de Mexico, Mexico, DF, 04510, Mexico

ChrA is a membrane protein that functions as a chemiosmotic pump that effluxes chromate ions from the cytoplasm using the proton motive force. Plasmid pUM505 from *Pseudomonas aeruginosa*, plasmid pMOL28 from *Cupriavidus metallidurans*, plasmid 1 from *Shewanella* sp. ANA-3, and transposon Tn*OtChr* from *Ochrobactrum tritici* 5bv11, all encode ChrA homologs which confer resistance to toxic chromate ions. We showed previously that ChrA proteins comprise the chromate ion transporter (CHR) superfamily, composed by two families: large bidomain proteins (comprising seven LCHR subfamilies), and short monodomain proteins (comprising three SCHR subfamilies). However, because the only few ChrA proteins characterized are associated to *chrB, chrC,* and *chrF* genes, we performed a phylogenetic analysis of these chromate resistance *(chr)*determinants to obtain additional insights about the origin and functional role of the correspondent gene products. Psi-Blast exhaustive search for each *chr* determinant was performed at the NCBI website. Multiple alignments were performed with Clustal-X and phylogenetic analysis with MEGA6. Results obtained show that ChrB, ChrC, and ChrF proteins possess a limited distribution inside bacteria, and are associated to *chrA* gen coding proteins from LCHR2 and LCHR5 subfamilies. Since the genomic context among CHR subfamilies is not conserved, it is concluded that ChrA proteins can be functional without the concurrence of any other *chr* determinant.

Supported by DGAPA-UNAM IN216513.

718D

Weak conservation of pathways in mouse and human aging tissues

Andrea Komljenovic [1,2], Marc Robinson-Rechavi[1,2]

[1] University of Lausanne, Lausanne, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland

Despite abundant experiments and diverse data available to study aging, the mechanisms of aging are still poorly understood. To tackle this, we are interested in evolutionarily conserved marks associated with aging, from short lived model organisms to long lived species such as human. We present the analysis of publicly available aging datasets from human and mouse tissues to analyze gene expression changes during aging. We characterized co-modules showing the level of gene expression conservation between homologous tissues. Meta-analysis across different tissues in mouse and human shows overall down-regulation of age-related gene expression profiles**.** We identified the biological processes using gene set enrichment analysis for these tissues, and found that changes associated with age-related gene expression in skeletal muscle and brain are involved in the mitochondrion pathways and inflammatory response, respectively. These tissues are known to be important to changes in aging. However, there is only a weak positive correlation between aging effects in the human and mouse homologous tissues. The co-module identification showed connection to immune response process in brain tissue between human and mouse. Our study provides a framework for further comparative analysis in aging across different species.

L6 B

Diversity of small non-coding RNAs controlling the dominance network among self-incompatibility alleles in *Arabidopsis halleri*

Nicolas Burghgraeve, Vincent Castric
University of Lille. France

The existence of genetic elements acting as modifiers of dominance/recessivity of other genes was a source of great dispute between R. Fisher and S. Wright, and that lasted through the 20th century. The recent discovery of a set of small non-coding RNAs modifying dominance/recessivity relationships between alleles at the self-incompatibility locus in Brassicaceae has provided a unique opportunity to study the evolution of dominance via modifier loci. Since the discovery of these small RNAs several questions concerning the action of natural selection on these systems have emerged: 1) Is selective pressure on these modifiers stronger in the case of dominant or recessive alleles? 2) What happens when two small non-coding RNAs of a given haplotype present the same target? In order to address these questions, we performed a sequencing campaign of small non-coding RNAs from a sample of S-alleles found throughout Europe. We measured the observed diversity between and within populations at this locus in order to characterize the footprints of natural selection.

L 9A

Mutation rate plasticity in bacteria dependent on population density

Rok Krasovec [1], Roman Belavkin[2] , Alastair Channon[3,] , Elizabeth Aston[3,] , Huw Richards[1] , Chris Knight[1]

[1] The University of Manchester, Faculty of Life Sciences, UK, [2] Middlesex University, School of Engineering and Information Sciences, London, UK, [3] Keele University, School of Computing and Mathematics, Keele, UK

The probability of mutational events, known as mutation rate, can vary between genotypes and locally within genomes. However, the degree of organismal control over mutation rates remains an open question. We recently demonstrated that in the bacterium Escherichia coli, mutation rates conferring resistance to the antibiotic rifampicin increase more than 5-fold with decreasing final population density. This mutation rate plasticity (MRP) is mediated by luxS-dependent cell-cell interactions. luxS is a central quorum-sensing gene known to regulate various others density-dependent phenotypes. These initial studies involved mutations at a single gene (rpoB giving rifampicin resistance) in simple media containing a single carbon source. We have gone on to characterise MRP at other loci (primarily gyrA, where mutation can confer resistance to nalidixic acid) and in complex (Luria-Bertani) media. We find that the rate of mutations in each circumstance is also density-dependent, extending the range of the plasticity. These findings suggest that in this MRP we are observing plasticity of general mutation rate and that this plasticity applies to a range of environmental conditions. The evolutionary consequences of such plasticity in mutation rate remain to be elucidated.

23 Reframing the demography vs. selection debate using 21st century models and data

23.1

The history of migrations and adaptations to high altitude in humans

Anna Di Rienzo [1], Choongwon Jeong[1], Buddha Basnyat[2], Geoff Childs[3], Sienna Craig[4], Maniraj Neupane[5], Joseph Marcus[1], David Witonsky[1], John Novembre[1], Cynthia Beall[6]
[1] University of Chicago, Chicago, USA, [2] Oxford University Clinical Research Unit, Kathmandu, Nepal, [3] Washington University, St Louis, USA, [4] Dartmouth College, Hanover, USA, [5] Mountain Medicine Society of Nepal, Kathmandu, Nepal, [6] Case Western Reserve University, Cleveland, USA

The ascent to high altitude presents formidable challenges to physiological processes, including hypoxia as well as cold stress and exposure to resource-poor habitats; these environmental features pose severe constraints on work capacity and reproduction. Therefore, the history of initial settlement at high altitude and subsequent population movements is intertwined with the history of adaptations. Studies of human populations from the Tibetan region have revealed a complex history of mixing among modern human populations as well as evidence that a key adaptive allele was introduced into the Tibetan genome through admixture with a Denisova-like population. This history provides novel opportunities for mapping locally beneficial alleles and for learning about the genetic architecture of high altitude adaptations. We used a large cohort of Tibetans to conduct local ancestry analyses to identify loci with excess high altitude ancestry and with signals of positive selection. These loci include the *EPAS1* and *EGLN1* genes as well as new signals at *PPARA, SLCO1B1/SLCO1B3, HFE,* and *PON1*, which are known to play a role in hypoxia response or in blood traits. We also analyzed whole genome sequence data using the MSMC approach to investigate further the population migration and admixture model and show that a simple model of divergence without major gene flow events is unlikely to explain sequence variation data for Tibetans, Sherpa and low altitude East Asians. Finally, by using the ABBA/BABA test on whole genome sequence data, we detect a significant contribution from a Denisova-like population to the genome of high altitude populations.

23.2

Robust inference of selection against maladaptive gene flow

Simon Aeschbacher, Graham Coop
University of Califorina, Davis, Davis, CA, USA

Spatially heterogeneous selection is a crucial factor in the context of local adaptation and speciation. Selection against maladaptive immigrant alleles causes a reduction in gene flow at sites under selection. Importantly, this effect spills over to neutral variation and makes it hard to separate demography from selection. Here, we exploit the fact that the footprint of such selection depends on recombination. Merging the concept of effective migration rate with the structured coalescent, we predict the aggregate signal of selection against maladaptive gene flow as a function of the recombination rate. Given a genetic map and estimates of divergence, we infer the intensity of selection against gene flow per basepair. As we incorporate the genome-wide effect of selection at unlinked sites, we are able to estimate the neutral baseline migration rate, and hence disentangle demography from selection. We can also date the onset of selection, which may inform us about the timing of the colonisation of a new environment or when speciation began. We test our approach against simulations and apply it to two datasets from the yellow monkeyflower (*Mimulus guttatus*). In the first, we detect the signal of local adaptation to serpentine soils, and in the second we quantify and date selection against introgression from the selfing sister species *M. nasutus*. Lastly, we combine our theory with the sequentially Markovian coalescent to explore variation in effective gene flow both over time and as a function of recombination. This may shed further light on the demographic and selective past of populations.

23.3

Robust estimation of the distribution of fitness effects (DFE) of new mutations from genome-wide patterns of polymorphism and divergence

Paula Tataru [1], Maéva Mollion[1], Marco A.P. Franco[1], Nicolas Galtier[2], Sylvain Glemin[2], Thomas Bataillon[1]

[1] Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark, [2] Institute of Evolutionary Sciences, University Montpellier 2, Centre national de la recherche scientifique, Montpellier, France

Inference of the DFE of new mutations from patterns of polymorphism and divergence remains an exciting challenge. Numerous datasets currently assembled enable us to learn about the DFE in more species. Reliable inference of the DFE requires modeling jointly the effects of selection and demographics on the polymorphism, through the site frequency spectrum (SFS), and divergence data.

We develop a hierarchical probabilistic framework to model counts in the SFS and divergence for a set of independent genome fragments. Our approach extends significantly available methods that either rely on hierarchical frameworks or model SFS under explicit demographic models. Our primary goal is to estimate the DFE together with the distribution of mutation rates throughout the genome. We furthermore account for the effects of unknown demographics through nuisance parameters that "distort" the SFS. Our framework also relaxes the strong assumption often made in other methods: all observed non-synonymous mutations are either neutral or deleterious.

We compare the performance of our approach with existing methods. We analyze a set of exome re-sequencing data from three subspecies of chimpanzees, and a set of recently acquired RNA-seq datasets spanning multiple non-model animal species. We discuss what biological factors may account for the differences in the inferred DFE across species. We present evidence that both slightly deleterious and beneficial mutations are likely shaping the patterns of polymorphism and divergence in genomes. We additionally discuss how these findings are consistent with recent models relying on phenotypic fitness landscapes for predicting DFE and dynamics of adaptation over time.

23.4

Selection and demography in triallelic sites: a diffusion approach and application to nonsynonymous sites

Aaron Ragsdale, Ryan Gutenkunst
University of Arizona, Tucson, AZ, USA

Contemporary population genetics is focused on loci with two segregating alleles, but deep sequencing within populations is revealing increasingly more loci with more than two segregating alleles. This has rekindled interest in modeling multiallelic loci, including a recently developed coalescent method that can calculate the allele frequency spectrum (AFS) of triallelic loci under arbitrary demography. However, this coalescent method, like most, does not incorporate selection. Here we present a diffusion method that can readily incorporate the simultaneous effects of selection and demography. Our numerical approach to the triallelic diffusion equation with selection provides a fast and accurate approximation to the triallelic AFS.

As an application, we consider triallelic sites in coding regions, which can have both derived mutations synonymous to the ancestral state, one synonymous and the other nonsynonymous, or both nonsynonymous. The case in which both mutations are nonsynonymous is of particular interest, because it allows us to assess the correlation between selection strengths of mutations at the exact same site.

Finally, the numerical framework we have developed to tackle the triallelic diffusion equation can be extended to obtain the joint frequency spectrum for pairs of segregating sites with arbitrary recombination rate between them.

23.5

Inferring the recent evolutionary history of the sibling species pair, *Solanum peruvianum* and *S. chilense* , using 21st century models and data.

Laura Rose [1], Aparna Reddy[1], Paul Staab[2], Dirk Metzler[2]
[1] *Heinrich Heine University, Duesseldorf, Germany,* [2] *University of Munich - LMU, Munich, Germany*

The availability of newer affordable genome technologies has allowed scientists to venture outwards from the classical model systems and to begin answering questions about evolutionary genetics and trait evolution in non-model systems. A first step in many of these evolutionary studies is the description of a species' demography. This is important because some demographic effects can leave similar signatures in the genome as natural selection. Our labs focus on the evolutionary history of two species of wild tomatoes, *Solanum peruvianum* and *S. chilense*. We have generated a large RNA-seq dataset from nearly 40 individuals of these species. For inferring population genetic parameters that capture these species' evolutionary history, we developed a method called Jaatha. Jaatha is a composite-likelihood method which incorporates recent divergence and can be applied even when intralocus recombination rates are high, conditions which both apply to the wild tomato species. Similar to ABC methods, our method uses summary statistics based on the frequency spectrum of mutations to compare observed data to data simulated under different parameter combinations. Earlier population genetic analyses of these species based upon a much smaller sample of genes indicated that the migration rate between these species was non-zero, despite strong present-day barriers to hybridization. Furthermore, according to these studies, gene flow between species stopped in the very recent past. Here we will present our newest results based upon this large RNA-seq dataset to understand the interplay between demography, selection and speciation in wild tomatoes.

23.6

Expansion Out of Africa Affected Accumulation of Deleterious Alleles in Diverse Human Genomes

Brenna Henn
SUNY Stony Brook, Stony Brook, NY, USA

The Out-of-Africa (OOA) dispersal ~50,000 years ago is characterized by a series of founder events as modern humans expanded into multiple continents. Population genetics theory predicts an increase in the proportion of deleterious alleles in populations that have undergone serial founder effects during range expansions. To test this hypothesis, we have sequenced full genomes and high-coverage exomes from over 50 individuals from 7 human populations, establishing a picture of genomic diversity in geographically divergent groups from Namibia, Congo, Algeria, Pakistan, Cambodia, Siberia and Mexico. We find that individual genomes vary modestly in the overall number of predicted deleterious alleles they carry, although at the population-level, signatures of purifying selection are detected. Specifically, OOA populations show, on average, higher frequencies of deleterious alleles than African populations. We show via spatially explicit simulations that the frequency distribution of deleterious alleles are consistent with the Out-of-Africa dispersal, particularly under a model where deleterious mutations are partially recessive. Analysis of variation in OMIM genes across our datasets suggests an enrichment of deleterious alleles in genes identified as conferring pathogenicity through a recessive model of inheritance. We conclude that there is a strong signal of purifying selection at conserved positions within Africa, but that most predicted deleterious mutations have evolved as if they were neutral during the expansion out of Africa. OOA populations are thus likely to have a higher mutation load due to increased allele frequencies of nearly neutral variants that are recessive or partially recessive.

23.7

The dark side of domestication: Deleterious variation in the dog genome

Clare Marsden[1], Diego Ortega-Del Vecchyo[1], Gary Johnson[2], Robert Schnabel[2], Robert Wayne[1], Kirk Lohmueller [1]

[1] UCLA, Los Angeles, CA, USA, [2] University of Missouri, Columbia, MO, USA

Domesticated dogs have undergone dramatic population bottlenecks associated with domestication and breed formation as well as recent inbreeding. To investigate the role of population history in shaping the patterns of deleterious mutations in canids, we analyzed 49 genomes from 37 breeds of domestic dogs, 24 village dogs, and 19 genomes from 13 grey wolf populations. We find that the ratio of amino-acid changing to silent variants increases with decreasing neutral genetic diversity. Overall, breed dogs have the highest ratio of amino-acid changing to silent variants, wolves have the lowest ratio and village dogs are intermediate. Interestingly, the Isle Royale island wolf population, which has a small population size, also shows an unusually high ratio of amino acid changing to silent variants. In principle, the disproportionate accumulation of deleterious variants in domestic dogs could be due to either bottlenecks or recent inbreeding. We find the same overall patterns when computing heterozygosity by sampling sequencing reads from distinct individuals within a population, suggesting that the observed increase in the ratio of amino acid changing variants in dogs was likely driven by population bottlenecks, rather than recent inbreeding. This conclusion is supported by forward simulations. While the average wolf is heterozygous for more deleterious variants than the average dog, the average dog carries significantly more deleterious variants in the homozygous state. In sum, these results suggest that the interplay between purifying selection and the domestication process has affected patterns of deleterious variants in dogs relative to wolves.

23.8

Shifts in selective pressure or random genetic drift? From balancing to positive selection out of Africa

Cesare de Filippo [1], Felix M. Key[1], Silvia Ghirotto[2], Andrea Benazzo[2], Philip Kleinert[1], Juan R. Meneu[1], Antje Weihmann[2], NISC Comparative Sequence Program[3], Genís Parra[1], Eric D. Green[3], Aida M. Andrés[1]

[1] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2] Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy, [3] National Human Genome Research Institute, National Institutes of Health, Bethesda, USA

The interest in balancing selection has grown vastly in recent years, as it becomes increasingly clear that it plays an important role in maintaining advantageous polymorphism in the human genome. Most studies in humans have focused on long-standing balancing selection, which persists over long time and is generally shared across populations. But balanced polymorphisms can also promote fast adaptation, especially when the environment changes. To better understand the role of previously balanced polymorphisms in novel adaptations we investigated loci that show hallmark signatures of long-term balancing selection in populations from Africa, but not from Eurasia. The disparity between populations is due to changes in allele frequencies, with many intermediate-frequency alleles in Africans (likely influenced by balancing selection) segregating instead at low- or high-frequency in Eurasians. These changes in allele frequency are among the most extreme in the human genome, but they could be explained by shifts in natural selection or by the effect of random genetic drift during the bottleneck accompanying the out-of-Africa. We identify three genes whose patters of diversity cannot be explained by demographic factors alone and are best explained by recent changes in selective pressure among populations. Specifically, we infer that alleles previously under long-term balancing selection, or linked alleles, were targeted by positive selection in Eurasian populations. Balancing selection thus likely served as a source of functional alleles that mediated subsequent adaptations to novel environments. Finally, based on novel methodological approaches, we quantify the relevance of this mechanism in recent local adaptations of human populations.

23.9

Measurement of selection in humans by the analysis of autozygous regions in 3,222 exome sequences from a highly endogamous population

Vagheesh Narasimhan [1], Konrad Karczewski[4], Hunt Karen[2], Yali Xue[1], Petr Danecek[1], Shane McCarthy[1], Tyler-Smith Chris[1], Chris Griffiths[2], Dan Mason[3], John Wright[3], Eamonn Maher[5], Daniel MacArthur[4], Trembath Richard[2], David van Heel[2], Durbin Richard[1]

[1] Wellcome Trust Sanger Institute, Hinxton, UK, [2] Queen Mary University of London, London, UK, [3] Bradford Institute for Health Research, Bradford, UK, [4] Massachusetts General Hospital, Boston, USA, [5] University of Cambridge, Cambridge, UK

In the past few years, there has been much debate on the impact of demography on the efficacy of selection. Here, we provide evidence for the impact of population structure on selection against deleterious variation using an exome dataset of 3,222 healthy adult individuals with high amounts of parental relatedness from the British Pakistani population, which has had a long term endogamous structure.

Although genetic diversity is relatively high in the sample, we show that there has been a significant reduction of severely deleterious variants in individuals from our dataset when compared to those from outbred 1000 Genomes cohorts. Using the long autozygous stretches in these samples, we can observe selection on human genomes acting within a single generation by quantifying the reduction in rare loss of function variants (LoFs) within compared to outside these stretches. As LoFs in homozygous state result in gene knockouts, this also gives a lower bound on the fraction of human genes for which knockouts are incompatible with healthy human life (15.6%). Aggregating the difference across allele frequencies, we obtain a new direct estimate of the number of recessive variants carried in a single individual that are incompatible with healthy human life when homozygous (1.6/individual). Our results show that using exome sequence data from a highly endogamous population we are able to address key questions of selection and demography in a strong quantitative fashion.

23.10

Pervasive long-range linkage disequilibrium in *D. melanogaster*

Nandita Garud, Dmitri Petrov
Stanford University, Stanford, CA, USA

The extent to which adaptation in natural populations of *D. melanogaster* occurs via classical selective sweeps versus non-classical selective sweeps such as soft sweeps and incomplete sweeps is yet to be fully understood. Linkage disequilibrium (LD), measured between pairs of sites or as haplotype homozygosity, can offer insight into the selective and demographic processes shaping genetic variation in a natural population. We find elevated long-range LD on the scale of 10Kb in samples of >100 fully sequenced strains from North Carolina and Zambia *D. melanogaster* populations as compared to LD measured in simulations of several complex demographic scenarios. This elevation in LD and haplotype structure remains even after controlling for many sources of LD in the data including genomic inversions, population substructure, close relatedness of individual strains, and recombination rate variation. In addition, we find that background selection and complete hard sweeps cannot explain the elevation in LD without completely depressing diversity to levels orders of magnitude lower than observed in data. We speculate that either soft sweeps or incomplete sweeps occurring at a very high rate might be required to generate the LD and haplotype structure observed in data. In addition, we observe substantial differences in LD at sites belonging to different functional categories, suggesting that different selective forces may act on polymorphisms of varying functionalities.

23.11

Determinants of adaptive evolution in great apes

Kiwoong Nam, Thomas Mailund, Great Apes Genome Diversity Project, Kasper Munch, Mikkel Heide Schierup
Bioinformatics Research Center, Aarhus University, Aarhus, Denmark

It is an open question whether adaptation is mainly driven by selection on standing variation or on denovo mutations. Along the same lines we do not know to which extent adaptation is mutation limited.

One way to approach these questions is to contrast closely related species with different, but partly known, demographic histories. We contrast inferred strength and of adaptive evolution in 9 different species and subspecies of great apes using 87 high coverage full genome sequences from the great apes genome diversity project. We focus on the decay of diversity in intergenic, non-functional sites (as estimated by Phastcons) as a function of distance to genes. We find that magnitude of the decay in diversity as well as the physical extent are very different and positively correlated with the estimated effective population size of the different great apes species. In orangutans and gorillas, diversity is reduced by 30-40% around genes significantly reduced up to 1Mb away from genes, whereas bonobo and humans have reduction around 15% and only up to 100-200 kb away from genes. We use extensive simulations of recombining sequences under background selection, and positive selection to show that the observed patterns are compatible only with hard selective sweeps playing a large role on diversity patterns and that these are more numerous in species with large population sizes. This suggests that the rate of adaptive evolution may be limited by the occurrence of new mutations in great apes species.

23.12

Background Selection Drives Population Specific Patterns of Diversity and Biases Demographic Inference in Humans

Raul Torres [1,2], Ryan Hernandez[1,3]

[1] University of California, San Francisco, San Francisco, CA, USA, [2] Biomedical Sciences Graduate Program, San Francisco, CA, USA, [3] Department of Bioengineering and Therapeutic Sciences, San Francisco, CA, USA

The influences that demography and linked-selection have on driving patterns of genetic variation globally and locally across the genome have both been well characterized in humans. However, it is less well known how these two forces jointly act in determining patterns of variation across human populations. Additionally, the biases that linked-selection can introduce to demographic inference procedures have not been well studied. Utilizing phase 3 of the 1000 Genomes Project (TGP) and population genetic simulations, we investigate the impact of background selection (BGS) on both of these questions. With forward simulations incorporating linked purifying selection and a three population demographic model, we show that relative loss of genetic diversity (measured as the ratio of diversity from simulations with linked selection versus no linked selection) occurs at a faster rate in Asians and Europeans compared to Africans. Measuring genetic diversity (π) across Asian, European, and African TGP populations as a function of BGS strength along the genome confirms these observations. Additionally, we also observe that BGS influences the SFS by generating an increase in the proportion of rare variants in TGP populations. To investigate the effects of this on demographic inference, we applied the software tool dadi across the high coverage YRI, CEU, and CHS Complete Genomics samples of the TGP and find that performing demographic inference across regions of the genome under strong BGS results in biases that affect both the inferred magnitude of population expansions as well as the timings of such events.

23.13

Novel probabilistically interpretable methods for localizing targets of selective sweeps

Lauren A. Sugden, Sohini Ramachandran
Brown University, Providence, RI, USA

Many methods have been introduced to locate targets of hard selective sweeps, most of which do not model demographic history and rely on single statistics that identify genomic signatures of a sweep: long-range haplotype blocks, deviations from the neutral site frequency spectrum, and locus-specific population differentiation. Recently, methods that combine multiple statistics into a composite score have shown increased power to detect sweep sites. However, these methods must artificially compensate when any component statistic is undefined (as often happens with long-range haplotype statistics) and rely on arbitrary thresholds for classification.

We developed a novel method for identifying and localizing targets of hard selective sweeps that reports the probability that a site in question is the site of an adaptive mutation, and inherently accounts for undefined statistics. Our method learns the distributions of multiple component statistics through simulations of neutral loci and sweep scenarios under a demographic model for the populations of interest. Based on comparisons using simulated data, our method outperforms current approaches, particularly in two biologically interesting scenarios: when identifying completed sweeps (where the frequency of the derived allele equals 1) and when identifying recent fast sweeps. Applying our method to data from the 1000 Genomes, we recover known sweep targets and identify new candidate causal variants in the LCT and ADH gene families. We also extend our framework to consider dependencies between component statistics and find that controlling for observed values of F_{ST} leads to an excellent classifier.

23.14

Habitat shift model supports demography as the main driver of genome evolution

Tristan Lefebure [1], Claire Morvan[1], Florian Malard[1], Lara Konecny-Dupre[1], Laurent Gueguen[4], Michele Weiss-Gayet[2], Andaine Seguin-Orlando[3], Luca Ermini[3], Clio Der Sarkissian[3], Pierre Charrier[1], David Eme[1], Clementine Francois[1], Laurent Duret[4], Cristina Vieira[4], Ludovic Orlando[3], Christophe Douady[1]

[1] LEHNA, Université Lyon 1, Villeurbanne, France, [2] CGphyMC, Université Lyon 1, Villeurbanne, France, [3] Center for GeoGenetics, , University of Copenhagen, Copenhagen, France, [4] LBBE, Université Lyon 1, Villeurbanne, France

Eukaryotic organisms exhibit striking variations in their genome size (GS). Even though the contribution of mechanisms such as polyploidization events or transposable elements to DNA gain or loss is now better understood, the evolutionary origin of GS variation still remains largely unexplained. The effective size of populations, which determines the magnitude of genetic drift and regulates the efficiency of natural selection, has been hypothesized as the key parameter controlling genome and GS evolution. This demographic hypothesis is much debated and has been challenged by mutational and adaptive alternatives. We developed a new comparative genomic model to test whether this debated hypothesis might account for changes in GS among asellid isopods that have undergone multiple independent evolutionary habitat shifts from surface water to low-energy groundwater. Here we show that most of these habitat shifts were associated with a long-term reduction in population size, as depicted by a higher transcriptome wide dN/dS ratio, and did not result in a higher rate of positive selection. These independent population size reductions were paralleled by a massive increase in GS (25% increase on average), an increase also confirmed in other non-asellid groundwater taxa. Genome sequencing showed that these GS inflations were triggered by the invasion of the genome by repetitive elements. While adaptive forces are probably at play during the colonization of new habitats, our model suggests that the long-term evolution of the genome and its size are mainly controlled by demographic forces.

160A

Genome-wide diversity patterns in two domesticated Prunus species (peach and cherry) using NGS data

Stephanie Mariette, Pauline Garnier-Gere
INRA, Cestas, France

Genetic diversity patterns observed in domesticated species result from complex evolutionary histories. The initial assumed or documented bottleneck generally brings about a strong reduction in diversity across the whole genome whereas artificial selection is expected to only affect a limited number of genes. Mating system changes have also probably played a role in domestication rates, predominant selfing favoring the fixation of domestication alleles.

Building on the peach draft genome, allelic resequencing of peach and cherry varieties allows exploring the genomic impact of domestication and breeding in fruit tree species. Peach has long been cultivated in eastern Asia (since 2000 years Before Present (BP) whereas cherry domestication is said to be more recent (a few centuries BP). The two species also differ for their mating system since peach is a selfing species, whereas cherry mating system is controlled by a self-incompatibility locus. We analysed 454 next-generation genomic re-sequencing data using different sampling strategies (Pool-seq and individual re-sequencing) and assembly

methods in order to compare genetic diversity patterns in ancient and modern varieties of peach and cherry, sampled in the European genepool.

161B

High diversity, cross-taxon gene flow and signals of SIV-related selection in vervet monkeys inferred from Africa-wide whole genome sequencing

Hannes Svardal [1], The Vervet Consortium[1,2], Anna Jasinska[2], Nelson Freimer[2], Magnus Nordborg[1]

[1] Gregor Mendel Institute, Vienna, Austria, [2] University of California Los Angeles, Los Angeles, USA

The characterisation of the evolutionary history and adaptive forces in a close human relative is relevant for understanding our own evolutionary past and susceptibility to disease. Here we use whole genome re-sequencing of 163 vervet monkeys of five sub-taxa sampled across the African continent to infer subspecies relationships and demonstrate cross-taxon gene-flow. Identifying more than 50 million single nucleotide polymorphisms, we find both high diversity within and differentiation across sub-taxa. Unexpectedly, 10% of the common variants are common in all sub-taxa, pointing to potential targets of balancing selection. A scan for balancing selection shows enrichment in un-annotated genes and identifies the MHC and other immune system related genes. A scan for positive selection that makes use of the allele frequency differentiation across sub-taxa is highly enriched in genes whose orthologs have been demonstrated to interact with HIV, pointing to candidate loci for the adaptation to SIV and other viral pathogens. The candidates are further examined using complementary methods such as dn/ds ratio and local LD patterns.

162C

How common is balancing selection in newt immune genes?

Anna Fijarczyk, Katarzyna Dudek, Wieslaw Babik
Jagiellonian University, Kraków, Poland

Balancing selection is potentially an important mechanism maintaining genetic diversity, but it remains unclear whether it is common within genomes. Balanced polymorphisms may be a source of adaptive variation that can be shared by hybridizing species bolstering their adaptive potential. However, the genomic signatures of balancing selection can be easily distorted or mimicked by various demographic processes, introgression or presence of paralogous sequences. We used targeted re-sequencing approach to scan more than 600 immunological genes in two closely related species of newts for signals of balancing selection. *Lissotriton vulgaris/montandoni* constitute a complex of recently diverged, morphologically distinguishable taxa, characterized by a complex history of hybridization. Within this system we focus on two distinct lineages with no apparent history of gene flow to avoid the confounding effects of introgression in identification of balanced polymorphisms. We used several population genetic approaches to identify plausible targets of balancing selection. To minimize the incidence of false positives and maximize power of the scan a sampling scheme accounting for the effects of population structure was applied and a demographic model estimated from the data was used to generate neutral expectations. We present a set of candidate genes and discuss how they can be used to test for adaptive introgression within *Lissotriton* newts system.

163D

Keeping It Local: Evidence for Positive Selection in Swedish Arabidopsis thaliana

Christian Huber[1,2], Magnus Nordborg[3], Joachim Hermisson[4,1], Ines Hellmann [5]
[1] Mathematics and BioSciences Group, Max F. Perutz Laboratories, Vienna, Austria, [2] Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria, [3] Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria, [4] Department of Mathematics, University of Vienna, Vienna, Austria, [5] Department of Human Genetics & Anthropology, LMU, Munich, Germany

Detecting positive selection in species with heterogeneous habitats and complex demography is notoriously difficult and prone to statistical biases. The model plant A.thaliana exemplifies this problem: In spite of the large amounts of data, little evidence for classic selective sweeps has been found. Moreover, many aspects of the demography are unclear, which makes it hard to judge whether the few signals are indeed signs of selection, or false positives caused by demographic events.
Here, we focus on Swedish A. thaliana and we find that the demography can be approximated as a two-population model. Careful analysis of the data shows that such a two island model is characterized by a very old split time that significantly predates the last glacial maximum followed by secondary contact with strong migration. We evaluate selection based on this demography and find that this secondary contact model strongly affects the power to detect sweeps. Moreover, it affects the power differently for northern Sweden as compared with southern Sweden. However, even when the demographic history is accounted for, sweep signals in northern Sweden are stronger than in southern Sweden, with little or no positional overlap. Further simulations including the complex demography and selection confirm that this is not compatible with global selection acting on both populations, and thus can be taken as evidence for local selection within subpopulations of Swedish A. thaliana. This study demonstrates the necessity of combining demographic analyses and sweep scans for the detection of selection, particularly when selection acts predominantly local.

164A

Efficient joint inference of demography and selection

John Kamm, Jonathan Terhorst, Yun Song
UC Berkeley, Berkeley, CA, USA

We consider the problem of jointly inferring demography and selection using the allele frequency spectrum (AFS). The two standard approaches of computing the AFS are the Wright-Fisher diffusion and the coalescent. We advocate a third approach, the Moran model, which has superior speed and numerical stability compared to diffusion and coalescent approaches. In this talk, we will describe the Moran model approach and present our inference method, called momi, that can efficiently compute the multi-population AFS under selection, and jointly estimate demographic and selective parameters via gradient descent.
We will also discuss the following subtle issue: Under neutrality, the Moran model can be constructed so that the AFS is exactly the same as under the diffusion and the coalescent. However, under directional selection, the Moran AFS is not exactly equal to the diffusion AFS, but only converges to it as the number of lineages goes to infinity. Fortunately, we find that only a low to moderate number ($N < 100$) of lineages is needed for the Moran AFS to be good approximation of the diffusion AFS. Hence, we conclude that the Moran model provides an appealing and computationally efficient alternative for population genetic inference of demography and selection.

165B

Joint likelihood-free inference of the distribution of fitness effects, locus-specific selection and demographic history.

Athanasios Kousathanas [1,3], Christoph Leuenberger[1], Matthieu Foll[2,3], Daniel Wegmann[1,3]
[1] University of Fribourg, Fribourg, Switzerland, [2] Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [3] Swiss Institute of Bioinformatics, Lausanne, Switzerland

We present a new likelihood-free Markov chain Monte Carlo (MCMC) method optimized for the analysis of multidimensional models and apply this new method to jointly infer selection and demography. The approach consists of two key innovations: Updating only one parameter per iteration of the MCMC chain and accepting or rejecting this update based on linear combinations of statistics specific for that parameter. This increases the acceptance rate of such chains dramatically, and in turn allows for likelihood-free inference under high dimensional models of almost arbitrary complexity, for instance models encompassing demographic parameters along with locus specific selection coefficients. We prove that our method approximates the true posterior distribution and that for linear models, a single combination of summary statistics per parameter is sufficient. We then present an application of our method to jointly infer the distribution of fitness effects of new mutations (DFE), selection coefficients per segregating allele and the effective population size based on time-series allele frequency data. We show an analysis of time-series data obtained from experimental evolution of the Influenza virus under antiviral drug treatment. Our analysis provides insights into the shape of the DFE and reveals several positively selected mutations close to the interaction center between key viral proteins and the drug. Finally, we discuss how our new method can be applied to the multidimensional problem of jointly estimating the DFE, selective sweeps and complex demographic history using genome-wide polymorphism data by exploiting combinations of genomic signals that are uniquely informative for each model parameter.

166C

Modelling the effects of background selection on differentiation patterns in subdivided populations

Kai Zeng, Pádraic Corcoran
University of Sheffield, Sheffield, UK

Understanding the roles of positive and negative selection in shaping patterns of genetic differentiation between demes in a subdivided population is an important, but challenging, question in population genetics, as can be seen from much recent discussion on how these forces may have caused “genomic islands” of elevated differentiation in various species. Given the ubiquity of deleterious mutations, a model that takes into account the effects of selection against these mutations is likely to be more realistic than the neutral model as a null hypothesis for detecting loci under positive selection. To this end, we analysed a model of background selection that incorporated population subdivision, migration, changes in population size, and recombination. Analytic equations for pairwise coalescence times and F_{ST} as a function of time since an ancestral population split into partially isolated subpopulations were derived. A coalescent simulator that can generate samples of large sizes was also written. These results have been verified by carrying out extensive forward simulations, from which the conditions under which they are accurate were also obtained. The new model should be of use to researchers working on topics such as local adaptation and speciation.

167D

Inferring the demographic history of populations with NGS data: the case of deer mice from the Nebraska Sand Hills

Vitor C Sousa [1,2], Brant K Peterson[4], Catherine R Linnen[5], Mathieu Foll[6,2], Stefan Laurent[6,2], Susanne Pfeifer[7], Jeffrey D Jensen[6,2], Hopi E Hoekstra[3,4], Laurent Excoffier[1,2]

[1] CMPG, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3] Howard Hughes Medical Institute, Department of Organismic and Evolutionary Biology, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA, [4] the Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA, [5] Department of Biology, University of Kentucky, Lexington, KY, USA, [6] School of Life Science, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [7] Center for Integrative Bioinformatics Vienna, Vienna, Austria

The study of natural populations has been revolutionized by Next Generation Sequencing (NGS), which enables us to obtain genome-wide data from multiple individuals and populations. In particular, we can learn about the past demographic history of populations based on the site frequency spectrum (SFS). We will exemplify the application of a flexible composite likelihood method based on coalescent simulations, which allows testing the fit of alternative demographic scenarios and inferring demographic parameters. We applied this method to study deer mice (*Peromyscus maniculatus*) exhibiting striking coat-colour variation between populations on and off the Nebraska Sand Hills. We sequenced 330 individuals sampled along an environmental transect for approximately 1500 anonymous 1-kb genomic regions, and two candidate pigmentation genes: the Agouti signalling protein (Agouti; 185kbp) and the melanocortin-1 receptor (Mc1r; 160kbp) genes. Preliminary results suggest that models with gene flow between populations on and off the Sand Hills are clearly better supported by the data. Estimates based on the anonymous regions point to moderately high symmetric gene flow between on and off the Sand Hills. Interestingly, for the Agouti locus we find evidence of reduced gene flow in both directions, consistent with the action of divergent selection acting against migrants. For Mc1r, estimates point to asymmetric migration, with reduced gene flow only from on to off the Sand Hills. Together these results suggest populations on and off the Sand Hills are diverging in the face of gene flow due to the action of selection.

168A

MS ^{3} : Fast Coalescent Simulation of Selection at Multiple Sites With Migration and Recombination.

Gregory Ewing [1], Derek Setter[2], Joachim Hermisson[2], Jeffrey Jensen[1]
[1] École polytechnique fédérale de Lausanne, Laussane, Switzerland, [2] MaBS, University of Vienna, Vienna, Austria

Previously, the simulation of selection at multiple sites, including effects such as epistasis and linkage, relied on forward simulation tools. While flexible, these approaches are computational expensive and thus slow. While Coalescent simulation can be significantly more efficient, it typically cannot model such selection scenarios. Here, we report about MS^3. A hybrid simulation program that has a fast forward simulation pass for selection at multiple sites and a conditional coalescent pass to model linked neutral variation. This advance results in greatly improved performance allowing for the practical use of ABC inference methods for estimating the number of, and parameters underlying, multiple selected loci under complex demographic models .We present simulation studies of such inference and discuss practical implications for empirical data analysis.

169B

Identifying selection and local adaptation in spatially structured populations

Veronika Siska [1], Anders Eriksson[1,2], Andrea Manica[1]

[1] Department of Zoology, University of Cambridge, Cambridge, UK, [2] Integrative Systems Biology Lab, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

The availability of large-scale, dense genomic data has facilitated the development of various statistical methods to detect signals of recent natural selection in the genome. However, selection scans using different methods and datasets often generate lists that have very few candidate regions in common, questioning whether many hits are false positives. Neutral demographic processes can leave signatures which are hard to distinguish from those of local selection, and might be a major confounding factor limiting our ability to detect selected loci. In this project, we use a spatially explicit model informed by past climate to reconstruct past population sizes, local movements, and range expansions of Anatomically Modern Humans, and thus provide a realistic null demographic model against which the signature of geographically-localised selection can be detected. We parameterise our spatial model using patterns of genetic variation in modern human populations and ancient genomic data, and then use its predictions to how different methods behave on neutral markers, as well as under various modes of selection under such realistic complex demographies.

170C

Detailed signatures of human evolutionary history in linguistic and genetic data

Nicole Creanza [1], Marcus Feldman[1], Sohini Ramachandran[2]
[1] *Stanford University, Stanford, CA, USA,* [2] *Brown University, Providence, RI, USA*

Human evolutionary history, including the peopling of the world, has left a strong signature on human genetic data. Do languages, which can change faster than genes and are not necessarily vertically transmitted, exhibit similar traces of human demographic history? We examine large databases of microsatellite alleles and phonemes, the minimal sound units that can distinguish meaning between words; this research indicates that worldwide genetic and linguistic variation show similar regional patterns and strong associations with geography. However, whereas genetic and geographic distances are correlated on a worldwide scale, spatial structuring in languages is only detectable within ~10,000 km.

By projecting geographic distance matrices in different directions within large regions, we calculate the geographic axes with the strongest signals of population differentiation. Genetic and phonemic data predict similar axes of differentiation at a continental scale; however, this result likely conflates different population processes and multiple migration events. To determine whether specific features of human demography left signatures on genes and languages, we compare small-scale axes of linguistic and genetic differentiation with published predictions of regional migration paths, gene flow and admixture events, and selection pressures via environmental covariates. We then analyze the circumstances under which linguistic and genetic data show signals of differentiation along similar versus different trajectories to those predicted by demographic history or changing selection pressures. Further, we address potentially contradictory genetic and linguistic signals of human migrations and gene flow. By integrating data types and approaches across disciplines, we aim for a more nuanced understanding of human history.

171D

Constraint and adaptation in newt Toll-like receptor genes

Wieslaw Babik, Katarzyna Dudek, Anna Fijarczyk, Maciej Pabijan, Michal Stuglik, Rafal Szkotak, Piotr Zielinski
Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland

Acute die-offs of amphibian populations worldwide have been linked to the emergence of viral and fungal diseases. Inter- and intra-specific immunogenetic differences may influence the outcome of infection. Toll-like receptors (TLRs) are an essential component of innate immunity and also prime acquired defenses. We report the first comprehensive assessment of TLR gene variation for urodele amphibians. The *Lissotriton* newt TLR repertoire includes representatives of 13 families and is compositionally most similar to that of the anuran *Xenopus*. Both ancient and recent gene duplications have occurred in urodeles, bringing the total number of TLR genes to at least 21. Purifying selection has predominated the evolution of newt TLRs in both long (~70 million years ago) and medium (~18 mya) timescales. However, we find evidence for both purifying and positive selection acting on TLRs in two recently diverged (2-5 mya) allopatric evolutionary lineages (*L. montandoni* and *L. vulgaris* graecus). Overall, both forms of selection have been stronger in *L. v. graecus*, while constraint on most TLR genes in *L. montandoni* appears relaxed. The differences in selection regimes are unlikely to be biased by demographic effects because these were controlled by means of a historical demographic model derived from an independent dataset of 62 loci. We infer that TLR genes undergo distinct trajectories of adaptive evolution in closely related amphibian lineages, highlight the potential of TLRs to capture the signatures of different assemblages of pathogenic microorganisms, and suggest differences between lineages in the relative roles of innate and acquired immunity.

172A

Ancestral GC-bias in four-fold degenerate sites of *Drosophila melanogaster* eroding, but maintained in *Drosophila simulans*

Juraj Bergman, Claus Vogl
Institute of Population Genetics, Vetmeduni Vienna, Vienna, Austria

The high GC content of four-fold degenerate sites in *Drosophila melanogaster* compared to putatively neutrally evolving short introns can be explained by selection acting to increase the efficacy of gene transcription or mRNA translation, or both. However, analysis of polymorphism in an African *D. melanogaster* population reveals no current directional selection operating on GC content in four-fold degenerate sites. Genomic data from this population were polarized with respect to the state in four other melanogaster subgroup species and compared to short introns. Conditional on the ancestral nucleotide base composition, the site frequency spectrum of *D. melanogaster* is similar to that of the short introns. This indicates no directional force towards increased GC on four-fold degenerate sites in the *D. melanogaster* lineage after the split from *D. simulans*. On the other hand, a similar analysis in *D. simulans* shows that the ancestral selection pattern is maintained within this lineage. This difference between the two sibling species could be due to a decreased selection coefficient or a historically smaller effective population size in *D. melanogaster*, or both. Furthermore, site frequency spectra of both species contain an excess of singletons. This deviation from neutral equilibrium expectation is indicative of recent population expansion.

173B

A flexible framework for demographic inference

Donna Henderson [1], Joe Zhu[1], Gemma Clucas[2], Tom Hart[1], Gerton Lunter[1]
[1] University of Oxford, Oxford, Oxfordshire, UK, [2] University of Southampton, Southampton, Hampshire, UK

Population structure is a crucial consideration in discoveries of disease associated variants and genes under selection. Our tool SMC^2 provides a flexible approach to demographic inference from whole genome sequence data. SMC^2 uses sequential Monte Carlo methods to create a sample of simulated ancestral graphs representing the history of the populations of interest. From this sample of graphs, we can infer historical population sizes and migration rates. In order to efficiently apply simulation methods to highly parameterized models, we have developed an adaptive version of Online Expectation Maximization, called Introspective Online Expectation Maximization (IOEM). IOEM allows the estimator of each parameter to learn at a rate that is determined by trends in its previous estimates. This increases computational efficiency when simultaneously inferring dozens of parameters. The flexibility of Monte Carlo techniques coupled with our novel IOEM algorithm allow for inference of not only effective population sizes, but also additional model complexities such as migration rates.

We have applied SMC^2 to samples from Gentoo penguins to elucidate the relationship between two subspecies. These subspecies reside in geographically close, but ecologically different habitats. We are interested in the complex population dynamics of these subspecies. Accounting for migration between the populations with SMC^2 enhances the investigation into the selection acting on these two subspecies.

174C

p { margin-bottom: 0.25cm; line-height: 120%; } Native Americans and their susceptibility to infectious diseases

Alex Mas Sandoval [1,3], Tábita Hünemeier[2], Francisco M Salzano[1], Maria Cátira Bortolini[1], David Comas[3]

[1] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [2] Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, São Paulo, Brazil, [3] Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain

Native American populations are characterized by complex evolutionary and demographic histories. For instance, they have experienced dramatic episodes of genetic drift, and successive bottleneck events. Furthermore, signals of positive natural selection associated to autochthonous environmental and cultural conditions have also been described. As a result, a unique genetic background emerged. The main objective of the present study was to test if, at least in part, the currently differential infectious disease susceptibility and other immunological conditions found in Amerindians and their descendants could be associated to their gene pool. Based on available and unpublished data, a total set of 250 Amerindians (Reich et al. 2012; Nature, 488:370-4) and 600 Brazilian Mestizos were analyzed. Data from 157 SNPs with a well-known association with infectious diseases and other conditions (GWAS Catalog, NHGRI; e.g malaria, hepatitis B, smallpox, as well as inefficient response to vaccines) were selected and analyzed with those from East Asians, Africans and Europeans. Our results revealed an excess of variants associated with infectious diseases in Amerindians when they were compared with the others ($p<0.0001$). Two SNPs (rs7574865 and rs9835973) present highly significant values ($p=1.36E-11$ and $p=1.71E-12$, respectively). The first is related to the prevalence of hepatocellular carcinoma in hepatitis B virus carriers; while the second is related to an enhanced response to the smallpox vaccine. This kind of knowledge, related to the unique evolutionary history of each population, can be used as an important instrument in the implementation of public health measures, including prevention.

175D

Inference of purifying and positive selection in three subspecies of chimpanzees (*Pan troglodytes*) from exome sequencing

J. Duan [1], T. Bataillon[1], C. Hvilsom[2,3], X. Jin[4], Y. Li[4], L. Skov[1], S. Glemin[5], K. Munch[1], T. Jiang[4], Y. Qian[1], A. Hobolth[1], J. Wang[4,6], T. Mailund[1], H.R. Siegismund[3], M.H. Schierup[1]
[1] Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark, [2] Science and Conservation, Copenhagen Zoo, Copenhagen, Denmark, [3] Bioinformatics, University of Copenhagen, Copenhagen, Denmark, [4] BGI Shenzhen, Shenzhen, China, [5] Institut des Sciences de l'Evolution, Universite Montpellier 2, Montpellier, France, [6] The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [7] The Department of Genetic Medicine, Faculty of Medicine and Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia, [8] Department of Biology, University of Copenhagen, Copenhagen, Denmark, [9] Macau University of Science and Technology, Macau, China

We study genome-wide nucleotide diversity in three subspecies of extant chimpanzees using exome capture. After strict filtering, SNVs and indels were called and genotyped for >50% of exons at a mean coverage of 35x per individual. Central chimpanzees (*P. t. troglodytes*) are the most polymorphic (nucleotide diversity, q_w= 0.0023 per site) followed by Eastern (*P. t. schweinfurthii*) chimpanzees (q_w = 0.0016) and Western (*P. t. verus*) chimpanzees (q_w = 0.0008). A demographic scenario of divergence without gene flow fits the patterns of autosomal synonymous nucleotide diversity well except for a signal of recent gene flow from Western into Eastern chimpanzees. The striking contrast in X-linked vs. autosomal polymorphism and divergence previously reported in Central chimpanzees is also found in Eastern and Western chimpanzees. We show that the direction of selection (DoS) statistic exhibits a strong non-monotonic relationship with the strength of purifying selection S, making it inappropriate for estimating S. We instead use counts in synonymous vs. non-synonymous frequency classes to infer the distribution of *S* coefficients acting on non-synonymous mutations in each subspecies. The strength of purifying selection we infer is congruent with the differences in effective sizes of each subspecies: Central chimpanzees are undergoing the strongest purifying selection followed by Eastern and Western chimpanzees. Coding indels show stronger selection against indels changing the reading frame than observed in human populations.

176A

Selective pressures acting on human polymorphic inversions

Carla Giner-Delgado [1,2], David Castellano[1], Magdalena Gayà-Vidal[1], Sergi Villatoro[1], David Izquierdo[1], Isaac Noguera[1], Marta Puig[1], Paul F O'Reilly[3], Mario Cáceres[1,4]
[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain, [2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain, [3] MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, [4] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Chromosomal inversions have been a model in evolutionary biology for decades. An important effect of inversions is that they suppress recombination in heterozygotes. This suppression may contribute to processes such as local adaptation, evolution of the sex chromosomes, and even speciation. However, very little is known about the evolutionary dynamics of these mutations, especially in humans. Here, we exploit the large-scale genotyping effort of the INVFEST project, where 45 inversions have been genotyped in 550 individuals from seven populations of the 1000 Genomes Project, to investigate the leading evolutionary forces acting on them and identify candidates to have a functional role. We adopted complementary strategies to interrogate the evidence for selection acting on the inversions. First, we estimated the expected frequencies and geographical distribution of the inversions under human demographic history, subject to ascertainment corresponding to our calling panel. 1000 Genomes Project SNPs and simulated data were used as null models. With this approach we detected inversions showing extreme population differentiation, suggestive of the initial phase of a selective sweep or local adaptation. Second, we estimated the age of unique inversions from nucleotide variation levels to try to narrow the expected frequency distribution and detect deviations from neutrality. Finally, we looked for signs of balancing selection in old inversions at intermediate frequency with low population differentiation in the present. Our study therefore contributes to the better understanding of this type of structural variants, and reports inversions that deserve further molecular and phenotypic characterization.

177B

BALANCING SELECTION IN HUMANS: INSIGHTS FROM A NOVEL SFS-BASED METHOD

Bárbara Bitarello [1,2], Cesare de Filippo[2], Aida Andrés[2], Diogo Meyer[1]
[1] University of São Paulo, São Paulo, Brazil, [2] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Detection of long-term balancing selection in humans has proven challenging, given that its footprints are restricted to short genomic segments. We present a method to identify these signatures, and its application genome-wide to identify targets of balancing selection. Our novel statistic, NCV (non-central variance), quantifies the deviation of allele frequencies from a given deterministic frequency equilibrium (feq), while including information on the number of fixed differences. Unlike classical methods such as Tajima's D, NCV it can be modified to maximize its power for feq other than 0.5. Through extensive simulations accounting for the demographic history of human populations, we show that NCV outperforms Tajima's D, HKA and T1/T2 tests for realistic demographic scenarios, and selective regimes feq 0.5-0.3, and has a similar performance to T2 for feq 0.5.

Simulations revealed this neutrality test can produce false positives in genomic with low numbers of informative sites (I.S). We have therefore both considered the number of I.S in a window to define significance, and restricted the scan to windows with at least 19 I.S in Africans. We then considered as candidate targets of balancing selection those for which NCV is lower than any observed NCV among 10,000 neutral simulations performed while conditioning on the observed number of I.S.

When applied to African and European populations from the 1000 Genomes low coverage data sets , we uncover targets including some previously identified loci – e.g. HLA-B, ABO and PROKR2 - as well as a considerable number of novel candidates of long-term balancing selection.

L 3C

Genome assembly and pooled population sequencing to identify genomic regions under selection in natural populations of the clam shrimp Eulimnadia texana

James G. Baldwin-Brown[1], Anthony D. Long[1], Stephen C. Weeks[2]
[1] *University of California, Irvine,* [2] *University of Akron*

Small invertebrates such as *Drosophila melanogaster* have historically been useful as model organisms quantitative and population genetics, but most are incapable of being archived in any practical way. Here, I present a set of results that establish the clam shrimp *Eulimnadia texana* as a practical model organism for quantitative and population genetics. Clam shrimp are attractive for genetics because of their small genome size (150Mb), large populations, short generations (~3 weeks), and their ability to lay eggs that can remain in diapause for decades without loss of viability. I generated a whole genome, hybrid de novo assembly of the clam shrimp genome using a combination of Illumina short read genomic data and PacBio long read data. I used Illumina RNA sequencing of both male and hermaphrodite clam shrimp in order to annotate the genome, then compared the genome to that of the closest sequenced relative, the water flea Daphnia pulex. In order to facilitate the use of this species in further genomics research, I generated allele frequency estimates from pooled sequencing data from 11 separate wild populations of clam shrimp and estimated average linkage disequilibrium, FST, and other classic population genetics statistics. A comparative analysis of FST between different populations should reveal sites with differential selection. I correlated these with a variety of environmental factors, including presence of predators, vernal pool dimensions, and ratio of clam shrimp males to hermaphrodites.

24 Short Tandem Repeats in the post-genomics age: Accurate typing, variability, evolution, and function

24.1

Genome-Wide Analysis of Expression Short Tandem Repeats in Human

Melissa Gymrek[1,2], Thomas Willems[1,2], Barak Markus[2], Stoyan Georgiev[5], Jonathan Pritchard[4], Yaniv Erlich [1,5]

[1] New York Genome Center, New York, USA, [2] Whitehead Institute for Biomedical Research, Cambridge, MA, USA, [3] MIT, Cambridge, MA, USA, [4] Stanford University, Stanford, CA, USA, [5] Columbia University, New York, USA, [6] University of Chicago, Chigaco, IL, USA

The involvement of repetitive elements in complex human phenotypes is largely unknown. We studied the contribution of polymorphisms in Short Tandem Repeats (STRs), one of the most abundant classes of repetitive elements, to expression of nearby genes in human lymphoblastoid cell lines. An initial genome-wide survey identified over 2,000 significant STR × gene pairs, termed expression STRs (eSTRs). Most of these eSTRs are replicable in orthogonal population groups and expression assays. A series of conservative statistical fine mapping techniques showed that approximately 500 eSTR associations remain after controlling for linkage disequilibrium with potential causal SNPs. Using linear mixed models, we show that eSTRs account for on average 25% of the gene expression heritability relative to that explained by all other common variants on the local haplotype. Further functional analysis shows that eSTRs are significantly enriched in conserved regions and disproportionately dwell in regulatory elements including promoters and certain transcription factor binding sites, corroborating their putative functional role. Taken together, our results show that eSTRs provide a novel set of regulatory variants and suggest that they may play a role in the genetic architecture of complex traits.

24.2

A first survey of tandem repeat instabilities and associated gene expression changes in 35 colorectal cancers

Tugce Bilgin Sonay [1,3], Malamati Koletou[1,2], Andreas Wagner[1,4]

[1] University of Zurich, Zurich, Switzerland, [2] ETH, Zurich, Switzerland, [3] The Swiss Institute of Bioinformatics, Lausanne, Switzerland, [4] The Santa Fe Institute, Santa Fe, New Mexico, USA

Colorectal cancer is a major contributor to cancer morbidity and mortality. Tandem repeat variation, a hallmark of colorectal cancer, and its effect on cancer phenotypes remain so far poorly studied on a genome-wide scale. Here we analyze the genomes of 35 colorectal tumors and their matched normal (healthy) tissues for two types of tandem repeat instability, de-novo repeat gain or loss and repeat copy number variation. Specifically, we study for the first time genome-wide repeat instability in the promoters and exons of 18,439 genes, and examine the association of repeat instability with genome-scale gene expression levels. We find that genes with a *de novo* repeat gain or loss are significantly enriched in tumors. Unstable repeats in promoters are almost twice as abundant in tumor genomes. Genes in tumor genomes with unstable repeats in their promoters are significantly less expressed and show higher levels of methylation. Genes in well-studied cancer-associated signaling pathways also contain significantly more unstable repeats in tumor genomes. Genes with such unstable repeats in the tumor-suppressor p53 pathway have lower expression levels, whereas genes with repeat instability in the MAPK and Wnt signaling pathways are expressed at higher levels, consistent with the oncogenic role they play in cancer. Our results suggest that repeat instability in gene promoters and associated differential gene expression may play an important role in colorectal tumorigenesis which is a first step towards the development of more effective molecular diagnostic approaches centered on repeat instability.

24.3

The evolution of short indels at meiotic recombination hotspots

Angelika Heißl, Barbara Arbeithuber, Irene Tiemann-Boege
JKU Linz, Institute of Biophysics, Single Molecule Genetics, Linz, Austria

Meiosis drives the rapid sequence evolution at recombination hotspot caused by the double-strand breaks that introduce *de novo* mutations or biased gene conversion at heteroduplexes during mismatch repair. Analyses on the rapid sequence evolution at hotspots have mainly focused on single nucleotide polymorphisms, but very little is known about the evolution of microsatellites at hotspots. Although the evidence correlating microsatellites with hotspots has been inconclusive, it has been reported that short indels are suppressors of recombination. Here, we have analysed the transmission of microsatellites at a recombination hotspot on chromosome 16 at two heterozygous mononucleotide repeats; (A7/A6) and (A22/A12). For this purpose, we have collected a large number of molecules of both crossover products by pooled sperm typing in several donors, and analyzed the crossover breakpoints with TaqMan–PCR and Sanger sequencing. We observed that the microsatellite (A22/A12) located 151 bp upstream from the hotspot center diminishes the breakpoint exchanges and show a reduced recombination frequency, while at the mononucleotide repeat (A6/A7) located 492 bp further upstream, we observed an overtransmission of the insertion over the deletion. It is possible that the position of the indels in relation to the hotspot or double-strand break center seems to influence the transmission of the repeats. Although more data is needed to verify these trends, the analysis of the transmission of a large number of single crossover products can provide important insights into human meiotic recombination biases.

24.4

Accurate typing of Short Tandem Repeats from genome-wide sequencing data and its applications

Arkarachai Fungtammasan[1], Guruprasad Ananda[1,3], Suzanne Hile[2], Marcia Shu-Wei Su[1], Chen Sun[1], Robert Harris[1], Paul Medvedev[1], Kristin Eckert[2], Kateryna Makova [1]

[1] Penn State University, University Park, PA, USA, [2] Penn State College of Medicine, Hershey, PA, USA, [3] The Jackson Laboratory of Genomic Medicine, Farmington, CT, USA

Short Tandem Repeats (STRs) are implicated in dozens of human genetic diseases and significantly contribute to genome variation and instability. Yet profiling STRs from short-read sequencing data is challenging because of their high sequencing error rate. Here we developed STR-FM, Short Tandem Repeat profiling using Flank-based Mapping, a computational pipeline that can detect the full spectrum of STR alleles from short-read data, can adapt to emerging read-mapping algorithms, and can be applied to heterogeneous genetic samples (e.g., tumors, viral populations, and genomes of organelles). We used STR-FM to study STR error rates and patterns in publicly available human, and in-house generated ultra-deep plasmid, sequencing datasets. We discovered that STRs sequenced with a PCR-free protocol have up to 9-fold fewer errors than those sequenced with a PCR-containing protocol. We constructed an error correction model for genotyping STRs that can assign genotypes correctly to 98-100% of STRs and can distinguish heterozygous alleles containing STRs with consecutive repeat numbers. Utilizing this pipeline, for the first time we determined the genome-wide STR germ-line mutation rate from a deeply sequenced human pedigree. More mutations originated in the male germ line. Additionally, we built a tool that recommends minimum sequencing depth for accurate STR genotyping, depending on repeat length of interest and sequencing read length. The required read depth increases with STR length and is lower for a PCR-free protocol. This suite of tools addresses the pressing challenges surrounding STR genotyping, and thus is of wide interest to researchers investigating disease-related STRs and STR evolution.

24.5

Microsatellite tandem repeats are abundant in mammalian and avian promoters and are associated with regulatory elements

Neil Gemmell, Hideaki Abe, Sterling Sawaya
University of Otago, Dunedin, New Zealand

Tandem repeats are genomic elements that are prone to changes in repeat number and are thus often polymorphic. These sequences are found at a high density at the start of human genes, in the gene's promoter. Increasing empirical evidence suggests that length variation in these tandem repeats can affect gene regulation. One class of tandem repeats, known as microsatellites, rapidly alter in repeat number. Some of the genetic variation induced by microsatellites is known to result in phenotypic variation. Recently, our group developed a novel method for measuring the evolutionary conservation of microsatellites, and with it we discovered that human microsatellites near transcription start sites are often highly conserved. New work in birds, particularly chickens, provides further evidence for the conservation of microsatellites near transcription start sites. Due to their intrinsic lability and their overlap with predicted functional elements, these results suggest that many promoter microsatellites have the potential to affect phenotypes by generating mutations in regulatory elements. We will discuss the potential functions of promoter microsatellites and their roles in rapid phenotypic shifts and disease.

719A

Conservation of low complexity regions through time: A study in Plasmodia

Sophia Chaudhry, Nyein Lwin, Fabia Battistuzzi
Oakland University, Rochester, Michigan, USA

Every genome analyzed so far has localized regions with lower diversity of nucleotides and amino acids compared to surrounding areas. Despite the widespread presence of these low complexity regions (LCRs), their evolutionary mechanisms are poorly understood although an association with fast rates of evolution is generally found. Homopolymeric (HP) repeats are a subset of LCRs that are composed of a single amino acid and evolve even more rapidly than LCRs. As a consequence of these rates, LCRs, and HP repeats in particular, are expected to be poorly conserved through time. We tested this prediction with a set of closely related lineages of Plasmodia, including three primate-infecting species one of which represented by five strains of the human pathogen Plasmodium vivax. Through a comparative analysis of orthologs of these species we identified the conservation level of low complexity regions in an effort to determine their decay time and their evolutionary paths from ancestor to descendants. Contrary to what was expected, we find at least 3% of HP repeats to be fully conserved among the three plasmodia and 30% conserved within P. vivax. We also find that HP repeats have variable evolutionary paths with ancestors that are either HP regions (20.5%) or LCRs (28.3%) These results suggest that HP repeats evolve by complex evolutionary mechanisms and encourage a comparative approach to study their evolution through time.

720B

Exploiting transcriptome data for the development and characterization of molecular markers related to storability of sugar beet tap roots

Eva Maria Sehr, Elisabeth Wischnitzki, Maria Berenyi, Silvia Fluch
AIT Austrian Institute of Technology GmbH, Tulln, Austria

Sucrose loss during post-harvest storage of sugar beet (Beta vulgaris L.) tap roots is a major concern of the sugar beet industry. During the storage period of up to two months, sucrose is mainly lost through respiration accompanied by the accumulation of invert sugar, the latter severely affecting sucrose processing. Thus, developing a variety selection / breeding program for storage could be of considerable economic benefit to the sugar beet industry, which can be achieved through the development and application of molecular markers.
Using high-throughput Illumina RNA sequencing, we analyzed the sugar beet tap root transcriptomes of two varieties reacting differently during storage (variety A: good storability, low sucrose loss; variety B: poor storability, high sucrose loss). Based on comparative transcriptomics, we identified 36 genes being expressed only in variety A and 57 genes only in variety B during a period of 12 weeks covered by 4 sampling time points. In addition, we characterized 5.728 simple sequence repeat (SSR) motifs within the pool of differentially expressed genes, where trinucleotide was the most common repeat unit (66%). After further validation, the identified expressed genes and SSRs might serve as potential molecular markers for the characterization of sugar beet tap root storability.

721C

Spatial Analysis and Genotype Specific Expression of Gene-linked Microsatellite Variation in Mountain Pine Beetle (Dendroctonus ponderosae): Evidence for expression driven election of inhibitor of apoptosis variation

Brent Murray
University of Northern British Columbia, Prince George, Canada

The mountain pine beetle (Dendroctonus ponderosae) is one of the most destructive forest insect pests in North America. Endemic to western North America, recent outbreaks have reached unprecedented levels and have resulted in range expansion northeastwardly into the oreal forests of western Canada. Current research has examined the spatial genetic variation (both traditional microsatellite and SNP) of mountain pine beetle populations in order to understand the genomic characteristics of the expansion population. Gene-linked microsatellites are a rich source of potentially adaptive variation. The transcriptome of D. ponderosae was screened for polymorphic microsatellites. The spatial variation of fifteen gene-linked microsatellites were compared to results of "neutral" microsatellite markers. Outlier Fst tests showed the microsatellite variation in the coding sequence of the neo-X inhibitor of apoptosis (IAP) gene to be consistent with positive selection. IAP is sex-linked in D. ponderosae having distinct neo-X and neo-Y specific alleles. Eight neo-X alleles have been identified and the variation among these alleles is primarily constrained to the microsatellite region corresponding to a variable number of serine repeats. In the northernmost range of D.
ponderosae, including the expansion population, an allele with 14 serine repeats predominates. In order to investigate functional differences among alleles, overwintering larvae were genotyped and IAP expression
levels measured using RT-qPCR. Genotype-specific expression in overwintering larvae showed reduced expression of IAP in males lacking the northern14-serine neo-X allele. These data suggest a possible mechanism
to explain the evidence for positive selection noted for the neo-X IAP alleles.

722D

Tandem repeat variation in human and great ape populations and its impact on gene expression divergence

Tiago Carvalho [1], Tugce Sonay[2,5], Mark Robinson[5,8], Maja Greminger[9], Michael Krutzen[9], David Comas[1], Gareth Highnam[3], David Mittelman[3], Andrew Sharp[4], Tomas Marques-Bonet[1,7], Andreas Wagner[2,6]

[1] Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain, [2] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [3] Department of Biological Science and Virginia Bioinformatics Institute, Virginia Tech, Virginia, USA, [4] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai School, New York, USA, [5] The Swiss Institute of Bioinformatics, Zurich, USA, [6] The Santa Fe Institute, Santa Fe, Santa Fe, USA, [7] Centro Nacional de Análisis Genómico (CNAG), PCB, Barcelona, Spain, [8] Institute of Molecular Life Sciences, Zurich, Switzerland, [9] Evolutionary Genetics Group, Anthropological Institute and Museum, University of Zurich, Zurich, Switzerland

Tandem repeats (TR) are stretches of DNA that are highly variable in length and mutate rapidly, and thus an important source of genetic variation. This variation is highly informative for population and conservation genetics, and has also been associated with several pathological conditions and with gene expression regulation. However, genome-wide surveys of TR variation have been scarce due to the technical difficulties derived from short-read technology.

Here, we explored the genome-wide diversity of TRs in a panel of 83 human and nonhuman great ape genomes, and their impact on gene expression evolution. We found that populations and species diversity patterns can be efficiently captured with short TRs (repeat unit length 1-5 base pairs) with potential applications in conservation genetics. We also examined the potential evolutionary role of TRs in gene expression differences between humans and primates by using 30,275 larger TRs (repeat unit length 2-50 base pairs). About one third of the 13,035 one-to-one orthologous genes contained TRs within 5 kilobase pairs of their transcription start site, and had higher expression divergence than genes without such TRs. The same observation held for genes with repeats in their 3' untranslated region, in introns, and in exons. Using our polymorphism data for the shortest TRs, we found that genes with polymorphic repeats in their promoters showed higher expression divergence in humans and chimpanzees compared to genes with fixed or no TRs in the promoters. Our findings highlight the potential contribution of TRs to recent human evolution through gene regulation.

723A

Direct observation of microsatellite germline mutations in *Schistocerca gregaria* suggests higher evolution rates than expected from arthropod model species

Marie-Pierre Chapuis [1], Christophe Plantamp[4], Réjane Streiff[2], Laurence Blondin[3], Cyril Piou[1]
[1] CIRAD, UMR CBGP, Montpellier, France, [2] INRA, UMR CBGP, Montpellier, France, [3] CIRAD, UPR B-AMR, Montpellier, France, [4] CNRS, Université Lyon 1, Villeurbanne, France

In this study, we used 3492 parent-offspring segregations, a validated pool PCR protocol and a newly developed hierarchical Bayesian model to estimate rates of germline mutations at 24 short tandem dinucleotide repeats in the desert locust, *Schistocerca gregaria*. The mean rate of mutations at STRs from genomic untranslated regions was estimated at 2.1e-4 per locus per generation, which compared with vertebrate estimates and was one order of magnitude higher than estimates known in *Drosophila melanogaster* and *Daphnia pulex*. Biological factors (such as numbers of germ cell divisions) were unlikely to explain our result. Conversely, because we found evidence for directional bias for expansions even for long alleles and exceptionally large ranges of allele sizes, the mismatch repair and its activity dependence to allele length may explain order variation in mutation rates of STRs. In addition, microsatellites derived from interspersed repetitive elements, abundant in Orthopteran genomes, were among those with the highest mutation rates in the desert locust. Finally, the mean rate of mutations at STRs derived from expressed libraries was two-fold lower and associated with shorter microsatellite repeat tracts than for loci from genomic untranslated regions. For this origin of loci only, the mutational model seemed to deviate from that usually considered for STRs, i.e. a generalized mutational model with a larger fraction of single-step mutations than multi-step mutations. These results can be explained by selective constraints on the mutation length in dinucleotide STR loci in order to not disrupt the reading frame.

724B

Genomic diversity in the captive population of the Puerto Rican parrot: is there a path to recovery?

Yashira M. Afanador [1], Jafet Velez-Valentin[2], Ricardo Valentin-de la Rosa[3], Juan Carlos Martinez-Cruzado[1], Bridget vonHoldt[4], Taras Oleksyk[1]

[1] University of Puerto Rico at Mayaguez, Mayaguez, PR, Puerto Rico, [2] US Fish and Wildlife, San Juan, PR, Puerto Rico, [3] Natural Resources State Department, Arecibo, PR, Puerto Rico, [4] Princeton University, Princeton, NJ, USA

The Puerto Rican parrot is one of the most endangered species in the world. The dramatic decline of this species to 13 individuals in 1975 lead to the establishment of two captive breeding programs. To date, these conservation efforts resulted in a significant population recovery with over 500 individuals between wild and captive birds at the two locations. But, despite the almost 40 years of the conservation efforts being running an accurate picture of the genetic diversity of the Puerto Rican parrot has not yet been assessed. We isolated 13 highly polymorphic microsatellite markers for the Amazona vittata with observed heterozygosity between 0.484 and 0.800. Using these novel loci, we were able to analyze individuals of three different generations (breeding years) from two of the main breeding facilities of the Puerto Rican parrot (Rio Abajo and El Yunque), and analyzed our data for population and demographics inferences. A panel with combination of only five of our markers can identify a single individual by a simple genetic test, and can then be used for identification and verification of paternity among captive as well as wild population of the species. The gene diversity maintained in both populations was estimated to be 64%. Allelic richness was estimated to be 3.7. Our results will eventually aid the conservation breeding efforts by U.S. Fish and Wildlife Service and Departamento de Recursos Naturales of Puerto Rico.

725C

The evolution of satellite DNA in heterochromatin of Drosophila

Kevin Wei, Daniel Barbash, Andrew Clark
Cornell, Ithaca, NY, USA

Tandmely repeating satellite DNA elements in heterochromatin occupy substantial portions of eukaryotic genomes. While characterized as genomic parasites, they are often crucial for processes like chromosome segregation. Adding to the perplexity, they turnover at remarkably high rates between closely related species and their divergence has been implicated in reproductive isolation between species in the form of hybrid incompatibility. First, to understand their evolution at a short time scale, we characterized natural variation in simple-sequence repeats from inbred *Drosophila melanogaster* lines derived from multiple populations, using k-Seek which analyzes unassembled short reads. Many of the repeats identified show strong population differentiation, including two that are present in only some populations. Curiously, the population structure inferred from satellite quantities fails differ from the expected population relationships. The abundance of many satellites are correlated across lines, revealing concerted evolution. Surprisingly, we also identified negative correlations, suggesting antagonistic interactions. Second, we applied k-seek to Illumina sequencings of 12 Drosophila species to determine how satellite DNA contribute to the divergence of species. For many repeats, the divergence pattern is incongruous with the phylogeny. Many repeats are lost and/or gained within short evolutionary time. These results further underscore that satellite DNA evolve extraordinarily fast. Lastly, to explore one possible mechanism for their rapid evolution, we test whether population-specific satellites bias segregation of chromosomes during meiosis causing meiotic drive.

25 Ancient genomes: A Time machine for investigating natural selection

25.1

Ancient *Mycobacterium tuberculosis* genomes suggest re-adaption to pre-Columbian human populations

Johannes Krause [1], Kirsten Bos[1], Alexander Herbig[1], Sebastien Gagneux[2], Jane Buikstra[3], Anne Stone[3]

[1] Max Planck Institute for the Science of Human History, Jena, Germany, [2] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland, [3] School of Human Evolution and Social Change, Arizona State University, Tempe, USA

Tuberculosis has had a significant influence on human history in the Old as well as the New World. Here we discuss the evolution of the causative agent of the disease, *Mycobacterium tuberculosis*, by exploring the phylogeography of both modern and ancient bacterial strains. We focus in particular on ancient mycobacterial genomes isolated using hybridization capture of DNA from 1000-year-old human remains found in the Americas. These strains are most closely related to *M.pinnipedii*, a variant of tuberculosis that is adapted to seals and sea lions of the Southern hemisphere, and is a member of the *Mycobacterium tuberculosis* complex (MTBC). This finding emphasizes the potentially important role of sea mammals in disseminating the disease to human populations across ocean waters. We discuss various hypotheses about potential zoonotic transmission, re-adaptation, as well as replacement of ancient pre-Columbian mycobacterial strains in the New World. Using ancient mycobacterial genomes as tip-calibration points for the molecular clock we calculate in two independent dating analyses a most recent common ancestor for all MTBC strains of less than 6000 years, implying a Holocene dispersal of the pathogen and a rapid host adaption to various mammalian species including humans.

25.2

A paleogenomic perspective on horse domestication.

Ludovic Orlando
Centre for GeoGenetics, University of Copenhagen, Copenhagen, Denmark

The domestication of the horse had a far-reaching impact on the socio-political and economic trajectories of human societies. Horses provided, for the first time, a means of rapid transportation on land, which revolutionized communication and the circulation of ideas, languages and religions. Horse chariotry and cavalry also changed warfare and equestrian technologies greatly stimulated agricultural productivity. However, the 5,500 year long history of horse domestication, which redirected the natural evolution of wild horses into the hundreds of modern breeds living today, is difficult to reconstruct from archeozoological data and modern genetics. Yet, with archeogenetics, one can catch evolution red-handed and access the genetic information from past individuals and populations. Our group makes use of state-of-the-art methods in ancient DNA research to characterize ancient horse genomes across a full geographic and temporal range, and track the genomic changes underlying horse modifications, from the early stages of domestication to the recent formation of industrial breeds. Comparative approaches allow us to reconstruct the population context in which domestication took place and reveal the genetic basis for the behavioral, physiological, and other biological traits differentiating wild and domesticated horses. Finally, the extraordinary adaptations of Yakutian horses, which survived the coldest environment in the whole Northern hemisphere, give use the opportunity to quantify the strength of natural selection in the presence of limited human pressure.

25.3

Ancient genomes improve our understanding of modern human population differentiation

Felix M. Key [1], Qiaomei Fu[2], Frédéric Romagné[1], Michael Lachmann[3], Aida M. Andrés[1]
[1] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2] Harvard Medical School, Boston, USA, [3] Santa Fe Institute, Santa Fe, USA

Modern human colonization of our vast planet was accompanied by independent allele frequency changes in each population. As a consequence of this process, the frequencies of alleles at polymorphic sites harbor patterns of population differentiation that largely agree with geography. Some loci show nevertheless high population differentiation, and they are enriched in genic variants. This signature can be explained both by positive selection and by background selection that increases random genetic drift. We aim to improve our understanding of these processes with high-quality genomes of ancient modern humans, which provide a snap shot of the frequency of each allele in the past. We developed and applied a statistic that jointly analyzes modern and ancient genomes to investigate changes in the frequency of functionally relevant alleles. Using archaic genomes we observed changes in Eurasia that cannot be explained by neutral forces or background selection alone, and that are best explained by the action of recent positive selection. The genomes of several ancient European individuals allowed us to determine the contribution of different European founder populations to these signatures, and their effect in the genetics of present-day European groups. We also identified genes and functions that have experienced fast allele frequency changes at different time points through European history, which provide relevant information about the adaptive processes taking place in ancient, and very recent, European history.

25.4

Quantifying selection against Neanderthal introgression

Ivan Juric, Graham Coop, Simon Aeschbacher
University of California Davis, Davis, CA, USA

Admixture between diverged populations leads to an exchange of genetic material and the potential for selection against introgressing alleles. Over time, selection will reduce the frequency of introgressed deleterious alleles, while also indirectly affect linked neutral variation. We can learn about selection against introgression by studying how the frequencies of linked neutral alleles change as a function of their distance from selected loci. However, in practice we usually do not know the number, locations, and selection coefficients of deleterious alleles. In addition, genetic drift, demography, and population structure may confound the signal of selection at individual loci. Here, we describe a new approach for robustly estimating the proportion of functional loci under selection, and the genome-wide average selection coefficient. Our approach relies on modeling the expected level of introgression across the genome as a function of recombination rates and functional annotation, and fitting model parameters to observed introgression levels.
We use our method to estimate the extent of selection against Neanderthal ancestry in modern humans. We also discuss the role of selection during incipient speciation, as well as the differences between X chromosome and autosomes.

25.5

Ancient Exomes of Individuals from the Pacific Northwest Coast Reveal Immune-Based Adaptations to the Americas

John Lindo [1], Michael DeGiorgio[2], Ripan Malhi[1]
[1] University of Illinois, Urbana, USA, [2] Pennsylvania State University, University Park, USA

The genomic background of ancient Native Americans has been postulated as a factor for the disease-related population declines after European colonization. However, this concept has not been explored on a genome-wide level, utilizing both living and ancient individuals. Instead, most of this evidence is either indirect or based on genetic studies of living populations, making it difficult to examine the extent genetics contributed to this presumed susceptibility.

In this study, we scanned contemporary and ancient whole-exomes for signatures of positive selection to examine relevant genetic differences in a First Nation population, both pre- and post-European colonization. We identified significant changes in allele frequencies between the ancient and contemporary populations, involving genes that have been correlated with colonial-era disease. These changes may help explain aspects of the historical experiences of indigenous peoples with European-borne pathogens and help illuminate the dynamics of human adaptation to new environments, in both the context of isolation and rapid merging of populations.

25.6

Disentangling natural selection from drift using forward simulations based on ancient genetic data

Karola Kirsanow [1], Adam Powell[1], Mark Thomas[2], Joachim Burger[2]
[1] Johannes Gutenberg University Mainz, Mainz, Germany, [2] University College London, London, UK

Forward simulations have emerged as a robust method for directly testing hypotheses of natural selection using aDNA datasets. These methods simulate the trajectory of allele frequency change between ancient populations and an appropriate modern dataset, and can accommodate allele frequency uncertainty in ancient samples as well as other forms of parameter uncertainty, such as sampling biases in ancient and modern datasets, and uncertainty in absolute sample dates. Additionally, the strength of natural selection can be estimated for a variety of plausible demographic parameters.

Here we present new forward-simulation based strategies for investigating hypotheses of natural selection in different ancient human populations at various time depths, focusing on a suite of loci associated with human pigmentation, disease susceptibility, and metabolism identified as targets of natural selection from modern genetic data. We describe various means of incorporating models of drift plus natural selection into the simulation parameters, and discuss how estimating selection coefficients over a defined window of time can be used to better understand past selective regimes and population history.

25.7

Testing directional selection on polygenic traits using ancient DNA

John Novembre [1], Charleson Chiang[2], Joseph Marcus[1]

[1] University of Chicago, Chicago, IL, USA, [2] University of California-Los Angeles, Los Angeles, CA, USA

In many cases, understanding the evolutionary basis of trait variation requires the ability to distinguish between models where a trait has recently undergone directional selection as opposed to stabilizing selection or simply has been neutrally evolving. Methods to distinguish amongst such models are still poorly developed, especially for highly polygenic traits. Recent progress in understanding the polygenic basis of trait variation, using genome-wide association studies (GWAS), and the increasing availability of ancient DNA (aDNA) samples provide new opportunities for understanding selection on polygenic traits. Here, we develop a Bayesian inferential procedure that intersects putative quantitative trait loci discovered via GWAS with aDNA data to test whether a phenotype has experienced directional selection on the basis of allele frequency change. The underlying model incorporates uncertainty in allele frequencies from ancient and modern samples, as well as a stochastic model of genetic drift. We show that power exists even with small aDNA sample sizes when the signature of selection is distributed across many loci, and we apply this approach to investigate signatures of selection on height in humans, where directionality of selection may vary across populations. While aDNA studies in some species are rapidly scaling upwards in sample size, even small samples will be helpful for shedding light on the evolution of polygenic traits.

25.8

Detecting ancient selective sweeps in modern humans using the genomes of our closest extinct relatives

Stéphane Peyrégne [1], Christoph Theunert[1], Michael Dannemann[1], Michael Lachmann[2], Mark Stoneking[1], Kay Prüfer[1]

[1] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2] Santa Fe Institute, Santa Fe, USA

The genome sequencing of archaic humans (Neandertals and Denisovans) to high-coverage has opened up new avenues to investigate the genetic basis of modern human traits that are not shared by the archaics. Natural selection may have played a role in fixing these traits on the modern human lineage. However, the selection events driving the fixation would have had only a small timeframe to act, ranging from the split between archaic and modern humans ca. 650,000 years ago to the split of modern human populations from each other around 100,000 years ago.
We previously implemented a method to identify genomic regions that are likely to have undergone ancient selective sweeps. The method is based on a hidden Markov model that identifies regions in the genome where Neanderthals and Denisovans fall outside of the present-day human variation. Regions that are unusually long are candidates for ancient selective sweeps.
So far our test relied on polymorphisms within modern human populations. However additional information is present in variants that reached fixation during the sweep. We present a refined version of our method that uses fixed differences between archaic and modern human genomes as an additional source of information. To evaluate the power of our method to detect ancient selective sweeps we tested its performance under various scenarios of background selection and neutrality. Finally, we present an updated list of candidates that likely underwent positive selection on the modern human lineage since the split from Neandertal and Denisova.

298A

Comparative genomics between a mid-colonial *Capsicum* sample, wild chili pepper and modern domesticates

Maria Guizar [1,2], Angelica Cibrian[1,2]
[1] *CINVESTAV, Irapuato, Guanauato, Mexico,* [2] *LANGEBIO, Irapuato, Guanauato, Mexico*

After the onset of domestication, cultivated species are subjected to a heterogeneity of environmental and human-driven selective pressures; whereby phenotypic differences arise leading to new varieties during a process of diversification. We are interested in understanding the underlying genetic changes associated to this process in Capsicum (chili pepper). Presumed to be domesticated in Mexico c.a. 7,000 years ago, pepper was rapidly dispersed and adopted by different cultures worldwide soon after the discovery of America (1492), resulting in over 2,000 different cultivars with an array of phenotypic diversity, including gradients of pungency. Our goal is to identify and characterize loci and allelic variants involved in the generation of post-domestication diversification traits. To achieve this, we compared the patterns of genomic diversity in public genomes of Chinese and Mexican modern domesticated varieties of Capsicum annuum annuum, and the genomes of wild C. chinense and C. annuum var. glabriusculum to a genome we sequenced de novo from a Mexican archaeological sample corresponding to a period preceding the diversification of Chinese and most of Mexican modern varieties (360 +/- 30 BP, measured radiocarbon age). We discuss the presence of genomic variation that has been lost, changed or maintained in these samples, and provide insights of its role in the diversification history of Capsicum in two independent geographic regions (Mexico and Asia).

299B

Phasing of genotype data from ancient Nubian barley

William Nicholson, Robin Allaby, Neil Pearson, Sarah Palmer, Oliver Smith, Andrew Clarke, Roselyn Ware, Rafal Gutaker
Warwick University, Coventry, UK

Archaeobotanical samples of barley found at Qasr Ibrim display a two-row phenotype that is unique to the region of archaeological sites upriver of the first cataract of the Nile, characterised by the development of distinctive lateral bracts. The phenotype occurs throughout all strata at Qasr Ibrim, which range from 3000 to a few hundred years in age. Ancient DNA were extracted from barley samples from the entire range of occupancy of the site. Previous work by the group, showed a discord between the genotype and phenotype with regard to the Vrs1 gene, responsible for row number in extant barley - indicating a six-row ancestry for the Qasr Ibrim barley, followed by a reassertion of the two-row condition. In the current work, SNPs and indels were called on data from sequencing using DNA capture arrays. The current work will be described on approaches used for phasing the genotype data from the ancient barley samples for a number of different genes.

300C

Prehistoric Mitochondrial Genomes of Arabian Camels (*Camelus dromedarius*) Unravel the Genetic Variation in Ancestral Population and Possible Origin of Domestication

Elmira Mohandesan [1], Pauline Charrau[1], Michael Hofreiter[2], Joris Peters[3], Hans-Peter Uerpmann[4], Pamela Burger[1]

[1] Institute of Population genetics, Vienna, Austria, [2] The University of Potsdam, Potsdom, Germany, [3] -Ludwig Maximilian University, Munich, Germany, [4] -Eberhard-Karls University, Tübingen, Germany

Arabian camels (*Camelus dromedarius*), or dromedaries, are well-adapted to harsh, desert environments and provide a variety of goods and services to millions of people inhabiting these conditions. Archaeological data suggest that the domestication of dromedaries occurred approximately 3000 years ago in the Arabian Peninsula. However, the wild ancestor of dromedaries remains un-described and the timing and origin of dromedary domestication remains unclear. In other livestock species, an increasing number of genetic studies have taken advantage of ancient and historical samples from both extant and extinct species to investigate the historical domestication process. To examine these questions in dromedaries, we examined the genetic diversity in the ancestral population by integrating ancient DNA from up to 7000 year old wild (extinct) and early domestic dromedaries in a dataset of 759 modern mtDNA (dloop). In early-domestic camels, we recovered the existing modern haplogroups (A, B), which indicate the early and complete presence of both haplogroups. In wild specimens from UAE, unique haplotypes and the ones clustered with major haplogroup (B) were recovered. We suggest this region as possible location of dromedary domestication. Furthermore, we report 14 nearly complete mtDNA from early-domestic dromedaries dated to 1000 – 2000 ybp. Because the environmental conditions of the desert drastically reduce the chances of DNA surviving from poorly preserved specimens, DNA extraction and sequencing remain a challenge. This study highlights one of the few successful recoveries of genetic materials from specimens belonging to hot and arid environments, and reports the first mtDNA recovery from early domestic dromedaries.

301D

Holocene history of European white oaks based on subfossil wood from across the range

Stefanie Wagner [1,2], Frédéric Lagane[1,2], Emilie Chancerel[1,2], Andaine Seguin-Orlando[3], Willy Tegel[4], Yves Billaud[5,6], Ludovic Orlando[3], Christophe Plomion[1,2], Antoine Kremer[1,2]
[1] INRA, UMR1202, BIOGECO, 33610 Cestas, France, [2] Univ. Bordeaux, BIOGECO, UMR1202, 33170 Talence, France, [3] Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark, [4] Institute for Forest Growth IWW,University of Freiburg, 79106 Freiburg, Germany, [5] MCC / DRASSM, 13016 Marseille, France, [6] UMR5204, Edytem, Univ. Savoie, 73376 Le Bourget-du-Lac, France

Understanding evolutionary responses of plant and animal species to environmental changes is paramount in the face of the ongoing global warming. Long-lived species, which include tree species from all forest ecosystems, are of particular concern as their generation time might limit their ability to adapt to changing environmental conditions. In the last hundreds of thousand years, our planet experienced major climate changes, which can be used to explore the past resilience of forest ecosystems and adaptability of forest trees. This could, in turn, help us anticipate their potential responses to future changes, which is essential to future decision-making. Taking European white oaks as a model, our project aims at tracking past demographic, migratory and selective trajectories in the face of major climatic changes using ancient DNA spanning a full temporal and geographic range. We restrict our investigations to the Holocene where oak fossils are abundant. However, since ancient DNA studies on tree material are still not common, we first investigated subfossil wood material from different taphonomical and temporal contexts to optimize and maximize access to authentic ancient DNA. Authenticated qualitative extracts were then screened for diagnostic *Single Nucleotide Polymorphisms* to infer major chloroplast lineages related to last glacial refugia and to discriminate between the two major species (*Quercus robur* and *Q. petraea*) that are characterized by different ecological requirements, but show no distinctive anatomical features. Comparative approaches between ancient and modern oak genomes will enable us to estimate genome-wide rates of evolution.

302A

Reconstruction of an ancestral genome and its comparison with an ancient genome of Yersinia pestis

Wandrille Duchemin [1], Eric Tannier[1,2]
[1] UMR CNRS 5558 - LBBE, Villeurbanne, France, [2] Institut National de Recherche en Informatique et en Automatique, Montbonnot, France

Ancient genomes are sequenced from archeological remains, while ancestral genomes are computed using the genomes of their extant descendants. Ancient genomes allow the study of extinct organisms without descendants, while ancestral reconstruction methods enable the potential reconstruction of much older genomes. These two complementary means of exploring the past forms of life on earth can benefit from the comparative study of one another.

For instance, ancestral genomes can be used as a template for ancient genome assembly [1], or for the design of better baits for ancient genetic material. Conversely, ancient genomes are good tools for the validation of ancestral genomes (usually restricted to robustness analysis).

We propose a method to reconstruct an ancestral genome in order to compare it to an extinct descendant. We illustrate our method with the example of the bacteria Yersinia pestis and its late medieval ancient genome [2]. Because of their internal position in the phylogenetic tree, ancient and ancestral genomes can be expected to show greater similarity than with any of the other extant species. We compare the Yersinia pestis ancestral and ancient sequences both in terms of local mutations (SNPs, small insertions and deletions), but also and particularly in terms of structural rearrangements (losses and duplications, inversions,
rearrangements), in order to have insights on their adaptive potential.

[1] Rajaraman et al., 2013, Bioinformatics, doi:10.1093/bioinformatics/btt527
[2] Bos et al., 2013, Nature, doi: 10.1038/nature10549.A

303B

Ancient DNA suggests a massive migration from the steppe during the Late Neolithic in Europe

Wolfgang Haak [1,2], Iosif Lazaridis[3,4], Nick Patterson[4], Johannes Krause[2,5], David Anthony[6], Alan Cooper[1], Kurt Alt[7,8], David Reich[3,4]

[1] Australian Centre for Ancient DNA, The University of Adelaide, Adelaide, Australia, [2] Max Planck Institute for the Science of Human History, Jena, Germany, [3] Department of Genetics, Harvard Medical School, Boston, USA, [4] Broad Institute of Harvard and MIT, Cambridge/Boston, USA, [5] Institute of Archaeological Sciences, University of Tuebingen, Tuebingen, Germany, [6] Anthropology Department, Hartwick College, Oneonta, NY, USA, [7] Danube Private University, Krems, Austria, [8] Institute for Prehistory and Archaeological Science, University of Basel, Basel, Switzerland

We generated genome-wide ancient DNA data from 69 prehistoric Europeans by using a novel enrichment method targeting 390,000 polymorphisms. This allowed us to obtain new insights about migrations in Europe's prehistory with implications for the spread of Indo-European language groups. At the beginning of the Neolithic period in Europe, ~8,000-7,000 years ago, closely related groups of early farmers appeared in Germany, Hungary, and Spain, where different from indigenous hunter-gatherers. At the same time, a distinctive population of eastern hunter-gatherers, which had a high affinity to Paleolithic Siberians, inhabited the region of today's Russia. During the Middle Neolithic (~6,000-5,000 years ago) we observe a resurgence of hunter-gatherer ancestry throughout much of Europe, while the contemporaneous Yamnaya herders of the Russian steppes shared ancestry with the preceding eastern European hunter-gatherers but also show signs of influence from a different stream of ancient Near Easterners that had reached the steppes. Western and Eastern Europe came into contact during the Late Neolithic ~4,500 years ago, as the Corded Ware people from Central Europe traced ~75% of their ancestry to the Yamnaya. This steppe ancestry documents a massive (second) migration into Europe from its eastern periphery, persisted in Bronze Age Europeans and is ubiquitous in present-day Europeans. These results have important implications for the proposed steppe origin of at least some of the Indo-European languages in Europe.

304C

The evolution and functional impact of human structural variants shared with archaic hominin genomes

Yen-Lung Lin[1], Pavlos Pavlidis[2], Jerry Ajay[1], Emre Karakoc[3], Stefan Ruhl[1], Omer Gokcumen [1]
[1] University at Buffalo, Buffalo, NY, USA, [2] Foundation of Research and Technology-Hellas, Heraklion, Crete, Greece, [3] Max Planck Institute for Evolutionary Biology, Plon, Germany

In this talk, we will describe our recent findings regarding polymorphic human *structural variants* (deletions, duplications, inversions and translocations of large segments of DNA) that are shared with archaic hominin genomes. We first studied polymorphic human deletions that are shared with archaic hominin genomes. Of the 427 such deletions we identified, ~87% of which originated before the Human-Neandertal divergence (*ancient*) and only ~9% of which have been introgressed from Neandertals (*introgressed*). Recurrence, incomplete lineage sorting between human and chimp lineages, and hominid-specific insertions constitute the remaining ~4% of allele sharing between humans and archaic hominins. As expected, our analyses indicate that the genomic landscapes of these variants were primarily shaped by negative selection, eliminating large and exonic variants. However, we found 17 deletion variants that are shared with archaic hominin genomes, including those leading to 3 fusion transcripts, including deletions of *DMBT1* and *LCE3BC* genes. The affected genes are involved in metabolism of external and internal compounds, growth and sperm formation, as well as susceptibility to psoriasis and Crohn's disease. In addition, we found 10 intra-exonic tandem repeat variation in humans, most of which are also defined by ancient, divergent haplotypes also predating Human-Neandertal divergence. These variants were also associated with autoimmune-diseases, such as asthma and atopic dermatitis. Overall, our findings reveal functional genomic structural variants that emerged before Human-Neandertal divergence with immunity related phenotypes. We argue that these *exonic* structural variants have been maintained in the population through complex adaptive forces, including balancing and population specific positive selection.

305D

Comparing methods to estimate contamination for ancient DNA high-throughput sequencing data based on haploid chromosomes

Víctor Moreno-Mayar [1], Thorfinn Korneliussen-Sand[1], Rasmus Nielsen[2,1], Philip L. F. Johnson[4], Eske Willerslev[1], Anna-Sapfo Malaspinas[1], Anders Albrechtsen[3]

[1] Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark, [2] Center for Theoretical Evolutionary Genomics, Department of Integrative Biology and Department of Statistics, University of California, Berkeley, California, USA, [3] The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark, [4] University of Maryland, Maryland, USA

Ancient DNA (aDNA) research has recently entered a new area. Previously limited to rather uninformative short segments of mitochondrial DNA, whole nuclear genomes have now become available for several extinct species, thus providing access to genome-wide SNP information at the species level and delivering new insights into deep evolutionary time. Nevertheless the quality of the aDNA data owing to post-mortem DNA damage as well as to contamination by microbial and contemporary DNA remains a challenge. This is especially relevant to the field of human aDNA where it is particularly difficult to distinguish endogenous molecules from contaminants. Some methods have been developed relying on haploid chromosomes. Yet, little is known about their potential limitations. In this work we compare two recently published methods based on the mitochondrial genome and the X chromosome.

Through simulations, we investigate the conditions that affect the inferred parameters for each method. We assess the effect of 1.the depth of coverage, 2.the contamination fraction, 3.the choice of reference panel, 4.the genetic distances between the sample and the contaminant and the reference panel, respectively and 5.the error rates (post-mortem damage or sequencing error) on the estimates for each method. These results allow us to issue recommendations regarding the limitations of each method for a particular dataset and potential contamination sources. Secondly, we re-analyze publically available data and investigate the relationship between the X-based and the mtDNA-based estimates. Finally, we propose an extension to the methods to allow for ancestry assignment.

306A

Whole genome analyses of contemporary and historic populations of the heavily exploited Atlantic cod

Bastiaan Star [1], Sissel Jentoft[1], Michael Matschiner[1], Olav S. Kjesbu[1,3], Paul R. Berg[1], Ian R. Bradbury[2], Christophe Pampoulie[4], Marianne HS Hansen[1], Morten Skage[1], Kjetill S. Jakobsen[1], Nils Chr. Stenseth[0]

[1] University of Oslo, Oslo, Norway, [2] Fisheries and Oceans Canada, St John's, Canada, [3] Institute of Marine Research, Bergen, Norway, [4] Marine Research Institute, Reykjavik, Iceland

Ancient genomics provides a powerful tool for detecting natural selection in wild populations. Here, we use this tool to explore the evolutionary effects of human exploitation of Atlantic cod (Gadus morhua), an important marine resource. In this species, significant phenotypic changes affecting life history characters have been recorded over the last 100 years. While these changes coincide with the onset of intense industrial fishing, it remains controversial whether selective harvesting is to blame. One factor complicating evolutionary inference is the existence of multiple populations that differ in behavior or phenotype, and nonetheless have ample opportunity to hybridize during spawning. We sequenced the genomes of a mix of historic and contemporary individuals (n = 76, approximately 10x coverage each) from the spawning grounds of the Lofoten in the North East Atlantic ocean. Of these, 24 individuals belong to a stationary coastal population, 24 to a migratory oceanic population, and 28 to a historic population sampled well over 100 years ago. We obtained several million SNPs, which are analyzed in a spatiotemporal context. Despite low overall divergence, a limited number of clearly distinct genomic regions separate the two contemporary populations. This separation allows us to readily identify the provenance of the historic samples. The direct temporal comparison provides excellent opportunities to detect allele frequency changes that can be attributed to selection.

307B

Preliminary insights into the genetic diversity of pre-contact Puerto Rican populations

Maria A. Nieves-Colon [1], Meredith Carpenter[2], Alexandra F. Adams[2], William J. Pestle[3], Carlos D. Bustamante[2], Anne C. Stone[1]

[1] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA, [2] Department of Genetics, Stanford University, Stanford, CA, USA, [3] Department of Anthropology, University of Miami, Coral Gables, FL, USA

Human groups initially populated Puerto Rico approximately 4,000 years ago. However, due to the demographic shifts that occurred after European contact, information regarding the origins and connections of pre-contact groups with continental indigenous populations, as well as their contribution to the ancestry of modern Puerto Ricans, has been limited to the analysis of historical, archaeological or modern DNA data. Here we present the results of pilot research focused on retrieving ancient mitochondrial (mtDNA) and nuclear DNA from 27 human skeletal remains (dated 590 to 1280 cal AD) from three pre-contact Puerto Rican sites: Paso del Indio, Punta Candelero and Tibes. Through the use of target enrichment capture and Illumina sequencing we have recovered low-coverage mtDNA genomes (0.2x to 12.4x) from nine individuals, and nuclear SNP data from three individuals. Preliminary analysis indicates that Native American mtDNA haplogroups A2 and C1 are found in high frequencies in these samples (n=4 each). All mtDNA haplotypes identified in the ancient remains have also been found among modern Puerto Ricans of Native American mtDNA ancestry. However, principal components analysis comparing genomewide SNP diversity between one sample from Paso del Indio and HGDP populations demonstrates that the ancient individual does not cluster closely with modern Native American groups in the HGDP panel. This preliminary data suggests that there may be at least some continuity of mtDNA lineages between pre-contact groups and modern Puerto Ricans. However, a clear assessment of the genetic relationships between these ancient groups and continental Native American populations requires additional analysis.

308C

Ancient genomes without ancient DNA : molecular evolutionary reconstruction of ancestral species traits.

Emeric Figuet [1], Jonathan Romiguier[2], Benoit Nabholz[1], Nicolas Galtier[1]
[1] ISE-M (Université Montpellier 2, CNRS UMR 5554—Institut des Sciences de l'Evolution), Montpellier, France, [2] Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

Reconstructing ancestral character states is an essential step to understand the processes driving the evolution and adaptation of genomes and phenotypes. Fossils are an obvious source of information, both morphological and molecular, but fossil ancestors are scarce. Here we show that ancestral life-history and ecological traits can be reconstructed thanks to molecular data even when the fossil record is lacking. This is because species traits - such as body mass or mating system - influence key aspects of the molecular evolutionary process - such as dN/dS and base composition - through mutation biases and genetic drift, so that by reconstructing ancestral sequences one can approach ancestral traits.
We validated this approach in a group of large mammals, the cetartiodactyls (ruminants and whales), in which DNA-aided reconstructions correctly inferred a small ancestor in agreement with the well-documented fossil record (Figuet et al. 2013, JEB). Applying the method to a large sequence data set across all mammalian orders, we reconstructed a relatively long-lived (>20 years) placental ancestor, questioning the classical picture of mouse-like early placentals (Romiguier et al. 2013, MBE). Surprisingly, the recently sequenced 45 bird genomes did not reveal any influence of species traits on the dN/dS ratio (Nabholz et al 2013, GBE; Weber et al. 2014, Genome Biology). Sequencing various non-avian sauropsids, we show that this is a bird-specific exception, and we clarify the reasons of this anomaly through polymorphism data analyses.
This work demonstrates that modern molecular data carry valuable information about former species traits, thus appropriately complementing fossil-based inferences.

309D

Nuclear and mitochondrial DNA sequences from two Denisovan individuals

Susanna Sawyer [1], Gabriel Renaud[1], Bence Viola[1], Jean-Jacques Hublin[1], Marie-Theres Gansauge[1], Michael Shunkov[2], Anatoly Derevianko[2], Kay Prüfer[1], Janet Kelso[1], Svante Pääbo[1]
[1] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2] Palaeolithic Department, Institute of Archaeology and Ethnography, Russian Academy of Sciences, Novosibirsk, Russia

Denisovans are a sister-group of Neandertals that were described based on a nuclear genome sequence from a finger phalanx (*Denisova 3*) found in Denisova Cave, Altai Mountains. A molar (*Denisova 4)* found at the same site, has a mitochondrial (mt) DNA sequence similar to *Denisova 3*. Here we present nuclear DNA sequences from *Denisova 4*, with the morphological description and the mitochondrial and nuclear DNA sequences from another molar (*Denisova 8)* from Denisova Cave. Like *Denisova 4,* this molar is very large and lacks traits typical of Neandertals and modern humans. Nuclear DNA sequences from the two molars form a clade with *Denisova 3*. The nuclear DNA sequence diversity among the three Denisovans is comparable to that among six Neandertals but lower than that among present-day humans. The mtDNA of *Denisova 8* is more diverged from and has accumulated fewer substitutions than the mtDNAs of the other two specimens suggesting that Denisovans were present in the cave over an extended period of time.

310A

Degraded DNA from historic barkcloth fabrics: getting the most out of NGS of non-model organisms

Nathan Wales [1], Jazmín Ramos Madrigal[1], Adrienne Kaeppler[2], Greta Hansen[2], Michele Austin-Dennehy[2], M. Thomas P. Gilbert[1,3]

[1] University of Copenhagen, Copenhagen, Denmark, [2] Smithsonian Institution, Washington, D.C., USA, [3] Curtin University, Perth, Western Australia, Australia

Although great advances have been made in the recovery and analysis of ancient DNA (aDNA) from hominins, iconic mammals, and domesticated crops, many archaeological and evolutionary research questions pertain to species that have not undergone the same level of genetic study. For the moment, archaeological and historic samples derived from such organisms are effectively "off-limits" from state-of-the-art aDNA methodologies, and therefore cannot benefit from numerous advantages offered by high-throughput sequencing, including the recovery of ultrashort DNA (<50bp). Yet, if cost-effective methods were developed to explore such remains, a new avenue of research would be opened for answering relevant archaeological and evolutionary questions. Therefore, we explored the feasibility of analyzing degraded DNA from historic barkcloth specimens prepared from one or more Polynesian plant species with uncharacterized genomes. In addition to the fact that plants tend to have challenging genomes with many repetitive elements, these barkcloth fabrics were pulverized during processing, fermented in tropical environments, and curated for over 150 years. Not surprisingly, they yielded DNA that was greatly fragmented and contaminated from many sources. Nonetheless, the recovered DNA was tested against chloroplast genes in NCBI's GenBank to infer the taxonomic relationships of the samples. To better reach species identifications and investigate if intraspecies variation can be characterized, reference chloroplast genomes for candidate species were generated via shotgun sequencing. This project provides a new perspective on the degree to which existing database biases plastome-based analyses, and whether it is necessary to generate new reference data for aDNA research on non-model organisms.

311B

Identification of dosage balance-sensitive gene families in angiosperms using a probablistic model for gene family size evolution

Setareh Tasdighian [1], Yves Van de Peer[2], Steven Maere[1]
[1] University of Ghent, Ghent, Belgium, [2] VIB, Ghent, Belgium

Whole genome duplications (WGDs) are believed to play a major role in angiosperm evolution. Previous studies have found that some functional categories of genes, including regulatory and developmental categories, expand almost exclusively through genome duplication, likely because their expansion through small-scale duplications is counteracted by dosage balance effects. However, the duplication dynamics of individual gene families have not been studied in detail. We developed a stochastic birth-death model to study the size evolution of gene families across a species phylogeny, taking into account both small-scale and large-scale duplication events. We use this model on a set of angiosperm species with known WGD history to assess the dosage balance sensitivity of individual gene families, and we interpret the results in the context of the potential impact of WGDs on evolutionary innovation in angiosperms.

312C

Autosomal and Mitochondrial DNA of Ancient Coastal Californians

Christiana Scheib [1], Tori Randall[2], Zuzana Faltyskova[3], Jay Stock[1], Toomas Kivisild[1]
[1] University of Cambridge, Cambridge, Cambridgeshire, UK, [2] San Diego Museum of Man, San Diego, California, USA, [3] University College London, London, UK

Nine full high-coverage mitogenomes and eleven ultra-low coverage nuclear genomes were sequenced from eleven skeletons exhibiting "paleoindian" morphology that were excavated from early to middle Holocene sites in coastal San Diego, California and Baja California, Mexico. Novel Haplotypes within the B and C lineages reflect the pattern of high frequencies of these two haplogroups in the American Southwest and suggest possible maternal genetic continuity over the past five thousand years in the region. The autosomal data clustered closest to modern Pima and Tepehuano groups except for one sample that clustered closest to the Surui with heavy migration into the root of modern Pima and Tepehuano populations. The success of this project shows that museum collections can provide excellent resources for population-scale ancient genomic studies.

313D

Loss of MHC genetic diversity in the last population of the woolly mammoth

Patrícia Pecnerová [1,2], Sergey Vartanyan[3], Love Dalén[1]

[1] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden, [2] Stockholm University, Stockholm, Sweden, [3] North-East Interdisciplinary Scientific Research Institute, Russian Academy of Sciences, Magadan, Russia

The last population of the woolly mammoth (Mammuthus primigenius) survived on Wrangel Island for more than five thousand years after it became extinct in continental Eurasia and North America. A lower variation in selectively neutral genetic markers has previously been observed in Holocene Wrangel population compared to the ancestral Beringian population. However, it is unclear whether the isolation on Wrangel Island had any effect on adaptive variation. Here, we examine if the isolation affected genetic diversity in the DQA locus of the major histocompatibility complex, which is a locus known to be under balancing selection. We analyse partial sequences from two exons and one intron in the DQA locus from samples dated to before and after the isolation on Wrangel, and find a reduction in number of alleles as well as observed heterozygosity in the isolated Holocene population. This suggests that balancing selection was unable to maintain diversity in the DQA locus, and consequently that genetic drift may have had a negative impact on disease resistance in this last surviving population of woolly mammoth.

314A

Evolution of DNA methylation patterns in the hominine lineage: modern methods for old bones?

Giulia Lizzo [1], E. Andrew Bennett[1], Pierre-Antoine Defossez[2], Eva-Maria Geigl[1], Thierry Grange[1]
[1] Institut Jacques Monod - University Paris Diderot, Paris, France, [2] Epigenetics and Cell Fate Department - University Paris Diderot, Paris, France

Epigenetic modifications are important modulators of gene expression that can be associated to phenotypic changes and used to track the evolution of cis-regulatory elements. Among the different types of epigenetic marker, DNA methylation is conserved over time and can be measured in ancient samples. We aim at performing an in-depth comparative study of the evolution of DNA methylation patterns in mineralized tissues of the hominine lineage. We are thus establishing reference evolutionary methylation maps using post-mortem samples of human and chimpanzee bones up to 100 years old, to ensure that they have experienced sufficient diagenetic transformations to mimic the taphonomic situation encountered in ancient bones. Furthermore, this study includes different types of bones in order to reduce noise due to inter-bone variability. Different methylation mapping approaches were used to identify those best suited to such samples. Whole genome bisulfite sequencing (BS) or reduced representation BS (RRBS) are not suitable for ancient samples due to the frequent presence of a vast excess of environmental DNA. We thus explored both high-throughput targeted BS using Bisulfite Patch-PCR, and a methylation-based enrichment method (MBD-seq). Both techniques require adaptations to ancient sample characteristics, including low quantity of endogenous DNA, high environmental DNA contamination and DNA fragmentation. We present preliminary results illustrating strengths and drawbacks of the chosen strategies.

315B

Natural selection at the brush-border: recent and ancient adaptations to carbohydrate diets in humans and other mammals

Chiara Pontremoli [1], Alessandra Mozzi[1], Diego Forni[1], Rachele Cagliani[1], Uberto Pozzoli[1], Giorgia Menozzi[1], Nereo Bresolin[1,2], Mario Clerici[3,4], Manuela Sironi[1]

[1] Scientific Institute IRCCS E.MEDEA, Bosisio Parini, Italy, [2] Dino Ferrari Centre, Department of Physiopathology and Transplantation, University of Milan, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, Milan, Italy, [3] Department of Physiopathology and Transplantation, University of Milan, Milan, Italy, [4] Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

Adaptation to food resources is a driver of molecular evolution in mammals and a major transition in human history, the development of agriculture, determined increased carbohydrate consumption. We investigated the evolutionary history of genes encoding brush-border proteins involved in carbohydrate digestion/absorption. Results indicated widespread adaptive evolution in mammals, with several branches experiencing episodic selection, particularly strong in bats. Many positively selected sites involved positions of fundamental importance to protein function (e.g. within glucosidase catalytic crevices), with parallel evolution at SI and MGAM. In human populations five genes were targeted by positive selection. Analysis of ancient DNA samples indicated that most selected alleles were already present in the Paleolithic, thus predating the emergence of agriculture. Thus, agriculture determined no major selective event at carbohydrate metabolism genes in humans, with implications for susceptibility to metabolic disorders. Our work also provides a list of candidate functional variants to be prioritized in functional studies.

316C

Temporal dynamics of adaptation in a selfing plant

Laurène Gay [1], Julien Dhinaut[2], Jean-Marie Prosperi[1], Joëlle Ronfort[1]
[1] Equipe GE2POP, UMR AGAP, Montpellier, France, [2] UMR CNRS 5561 BioGéoSciences, Dijon, France

Time series data are useful to understand the dynamics of adaptation in populations and the evolutionary mechanisms involved. In plants, this is made possible thanks to seed collections that can be preserved for a large number of generations. Using seeds collected over 25 years across France and Spain, we show that the selfing plant *Medicago truncatula* has shifted towards earlier flowering times in Corsica. This is further confirmed by a temporal analysis at the intra-population scale, where we found a shift of about two days in 23 years in a population in Corsica. The selection gradient measured in this population on modern plants suggests that there is still on-going selection towards earlier flowering. The analysis of genetic changes through time showed that this population is composed by a set of multilocus genotypes persistent through time, with varying frequencies that can jointly reflect the effects of drift and local adaptation. These results also provided evidence for substantial migration, which contributes to the observed change in genetic value for flowering time. Finally, we used the polymorphism at 2000 SNP targeted in candidate genes for flowering time to search for signatures of selection, using a method adapted to temporal samples.

317D

Genomic signals of migration and continuity in Roman Britain

Rui Martiniano [1], Russell McLaughlin[1], Malin Holst[2], Anwen Caffell[2], Kurt Hunter-Mann[3], Gundula Müldner[4], Janet Montgomery[5], Matthew Teasdale[1], Steve Roskams[6], John Oxley[7], Colleen Morgan[6], Maureen Carroll[8], Jan H Veldink[9], Wouter van Rheenen[9], Leonard H van den Berg[9], Mark Thomas[10], Ian Barnes[11], Christine McDonnell[3], Matthew J Collins[6], Daniel G Bradley[1]

[1] Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, [2] York Osteoarchaeology Ltd, University of York, York, UK, [3] York Archaeological Trust, York, UK, [4] Department of Archaeology, University of Reading, Reading, UK, [5] Department of Archaeology, Durham University, Durham, UK, [6] Department of Archaeology, University of York, York, UK, [7] City of York Council, York, UK, [8] Department of Archaeology, University of Sheffield, Sheffield, UK, [9] Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands, [10] Research Department of Genetics, Evolution and Environment, University College London, London, UK, [11] Department of Earth Sciences, Natural History Museum, London, UK

York (Eboracum) was a provincial city at the edge of the Roman Empire where a number of Roman cemeteries have been excavated. One of these at Driffield Terrace is unusual in a regional and national context, with a large predominance (70.8%) of decapitated young males buried there. They show frequent evidence for trauma consistent with violent life histories and have been alternately speculated as gladiators, soldiers and slaves with potentially foreign origins. Here we report the ~1X genome sequences of seven of these individuals from the Roman period. While six of the Roman burials show affinity with modern British populations, one sample, although indistinguishable in funerary ritual from the other skeletons in the cemetery, shows a clear signal of exogenous origin, with modern affinities pointing towards the Eastern Mediterranean, a clear indication of the cosmopolitan impetus of the Roman empire, even at its western fringe.

318A

P { margin-bottom: 0.08in; } Investigation of mitochondrial DNA diversity in early medieval Southern Italy

Maria A. Spyrou [1], Åshild J. Vågene[1], Luca Bondioli[2], Alessandra Sperduti[2], Kirsten I. Bos[3], Johannes Krause[1,3]

[1] Department of Archeological Sciences, University of Tübingen, Tübingen, Germany, [2] Museo Nazionale Preistorico Etnografico Luigi Pigorini, Rome, Italy, [3] Max Planck Institute for the Science of Human History, Jena, Germany

P { margin-bottom: 0.08in; }

Assessing the genetic make-up of past populations is key in understanding the changes in human genetic diversity through time. Due to its central location in the Mediterranean sea, the Italian peninsula has hosted various populations during prehistoric and historic times. Of particular interest to the present study, are the Roman and Byzantine imperial times (27 BC – 1453 AD) when extensive human movements, for colonization and trading purposes, likely played an important role in the contemporary genetic composition of the region. Here, we have focused on the analysis of 22 human mitochondrial genomes from a mass burial site, dated to the 8th-10 centuries AD, and located at the city of Venosa, in Southern Italy. Even though mitochondrial DNA analysis only yields information on maternal inheritance, it can still be used to demonstrate population dynamics and history. We used an in-solution hybridization capture technique coupled with next generation sequencing to compare the maternal lineage diversity of the Venosa population with present-day European populations. The mitochondrial data obtained, showed remarkable ancient DNA preservation, and permitted the reconstruction of 22 complete mitochondrial genomes of high coverage. Mitochondrial haplogroup assignment of every individual revealed a rather genetically diverse population. Further analysis proposed that the genetic diversity found in early medieval Venosa is comparable to the composition of most of Europe today. Our results suggest that the mitochondrial diversity in Europe today could potentially be traced back as far as the 8th century AD.

319B

Quantitative Evaluation of ancient DNA Losses in Fossil Bones irradiated using Synchrotron X-rays

Alexander Immel [1], Marion Bonazzi[2], Adeline Le Cabec[3,4], Paul Tafforeau[3], Verena J. Schuenemann[1], Alexander Herbig[5], Bence Viola[4], Jean-Jacques Hublin[4], Johannes Krause[1,5]

[1] Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Baden-Wuerttemberg, Germany, [2] Institute of Clinical Molecular Biology, Kiel University, Kiel, Schleswig-Holstein, Germany, [3] European Synchrotron Radiation Facility, Grenoble, Dauphiné, France, [4] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Saxony, Germany, [5] Max Planck Institute for Science and History, Jena, Thuringia, Germany

Ancient remains such as sub-fossilized bones and teeth may still contain DNA. Due to taphonomical processes, this ancient DNA (aDNA) is highly degraded and only present in minute amounts. Nevertheless, specific techniques have been optimized to recover and analyse degraded and damaged aDNA. Yet certain palaeontological studies, particularly those involving X-ray micro-computed tomography imaging, have been presumed to potentially compromise ancient DNA analyses. Micro-CT scanning is regarded as structurally non-destructive and precious remains are commonly scanned to preserve the morphological information prior to other analyses involving a destructive sampling. Although the effects of X-ray radiation on DNA in living organisms are well-documented, their impacts on aDNA (in dry conditions) remain largely unexplored.
We investigate the effects of synchrotron X-ray irradiation conducted at the European Synchrotron Radiation Facility (ESRF) on the retrieval of aDNA from fossil megafauna bones. The study was designed to compare the effects of different scanning conditions in order to identify the X-ray dose threshold below which no degradation of aDNA can be observed. All authenticity criteria of aDNA analyses were employed. We analysed: aDNA quantities, aDNA fragment length and C to T substitution frequencies typical of aDNA, both on taxa specific sequences and on total aDNA present in the samples. Although we observed a correlation between aDNA degradation and X-ray dose, the effect of the normal scanning conditions used at the ESRF for complete fossil imaging at 30 microns is not significant. We propose guidelines for X-ray imaging that should allow both, aDNA analysis and micro-CT scanning.

320C

Genetic diversity and evolution of cave bear from Central and Eastern Europe

Danijela Popovic[1], Mateusz Baca[2], Krzysztof Stefaniak[3], Grzegorz Lipecki[4], Adam Nadachowski[3,4], Martin Sabol[5], Bogdan Ridush[6], Martina Roblickova[7], Vlastislav Kana[8], Theodor Obada[9], Anna Stankovic[10], Pawel Mackiewicz [11]

[1] Centre of New Technologies, University of Warsaw, Warsaw, Poland, [2] Center for Precolumbian Studies, University of Warsaw, Warsaw, Poland, [3] Faculty of Biological Sciences, University of Wroclaw, Wroclaw, Poland, [4] Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Cracow, Poland, [5] Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia, [6] Department of Physical Geography and Natural Management, Yurii Fed'kovych National University, Chernivtsi, Ukraine, [7] Anthropos Institute, Moravian Museum, Brno, Czech Republic, [8] Faculty of Science, Masaryk University, Brno, Czech Republic, [9] Institute of Zoology, Academy of Sciences of Moldova, Chisinau, Republic of Moldova, [10] Faculty of Biology, University of Warsaw, Warsaw, Poland, [11] Faculty of Biotechnology, University of Wroclaw, Wroclaw, Poland

Significant progress in techniques of ancient DNA isolation, amplification and sequencing enable reconstruction of extinct species evolution. One of them is cave bear, which was widely distributed in the Late Pleistocene from Western to Eastern Europe, but also inhabited Asia. Genetic analyses of mitochondrial DNA recognised two main haplotype groups described in the species rank as *Ursus spelaeus* and *U. ingressus*. The former dominated in Western Europe, whereas the latter inhabited mainly Eastern and Central Europe. *U. ingressus* characterized by more 'expansive' character, which was manifested by the colonization of Alps and Swabian Jura, which resulted in the replacement of native from, *U. spelaeus*. To study genetic diversity and evolution of *U. ingressus* in the whole area of its distribution, including the understudied region of its north-eastern range, we analysed mitochondrial control region from more than 70 samples representing excavation sites from Poland, Czech Republic, Slovakia, Moldavia and Ukraine. The studies showed that specimens from Romania with one Polish site were placed at the base of phylogenetic tree, which suggests an early migration of the bear from South-Eastern Europe beyond the Carpathian arch. A huge clade in the tree represented samples from Poland localities, which grouped with single samples from Germany and Ural. It suggests that the territory of Poland was a migration centre to Western and Eastern Europe. Our results indicate that *U. ingressus* was in fact the more expansive form of cave bear. This work was supported by National Science Centre grant no. 2012/07/B/NZ8/02845 to P. M.

321D

Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens

Clemens L. Weiss[1], Verena J. Schuenemann[2,3], Jane Devos[1], Gautam Shirsekar[1], Billie A. Gould[4], John R. Stinchcombe[4,5], Johannes Krause[2,6], Hernán A. Burbano [1]

[1] Max Planck Institute for Developmental Biology, Tuebingen, Germany, [2] Institute of Archaeological Sciences, University of Tübingen, Tuebingen, Germany, [3] Departments of Paleoanthropology and Archaeogenetics, Senckenberg Center for Human Evolution and Paleoenvironment, University of Tübingen, Tuebingen, Germany, [4] Departament of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, [5] Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Canada, [6] Max Planck Institute for the Science of Human History, Jena, Germany

Herbaria archive a record of changes of worldwide plant biodiversity harboring millions of specimens that contain DNA suitable for genome sequencing. To profit from this resource it is fundamental to investigate in detail the properties of DNA retrieved from these specimens and to estimate rates of DNA degradation. By analyzing 86 herbarium samples spanning the last 300 years using Illumina shot-gun sequencing, we investigate patterns of DNA fragmentation -length and base composition at breaking points-, and nucleotide misincorporation. We found an exponential decay relationship between DNA fragmentation and time, and estimated a per nucleotide fragmentation rate that is ten times faster than the rate estimated employing fossilized bones. Although strand breaks occur specially before purines, depurination-driven DNA breakage cannot explain the negative relation between fragment length and time, and other processes linked to DNA degradation remain to be identified. Reminiscent of what has been found using animal bones, we found a strong correlation between the deamination-driven accumulation of cytosine (C) to thymine (T) substitutions and time, which reinforce the importance of substitution patterns to indicate the ancient/historical nature of DNA fragments. Accurate estimations of DNA degradation through time will allow taking informed decisions about laboratory and computational procedures in order to take advantage of the vast collection of worldwide herbarium specimens.

322A

Optimizing the recovery of ancient DNA sequences from highly degraded material

Isabelle Glocke, Matthias Meyer
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Optimizing the recovery of ancient DNA sequences from highly degraded material
Isabelle Glocke, Matthias Meyer

Even though it has long been known that ancient DNA survival is inversely correlated with fragment size, current sample preparation techniques are not fully suited to recover DNA sequences from DNA fragments shorter than 35 bp. Recent genetic analyses of the Middle Pleistocene remains from Sima de los Huesos have shown that the recovery of such short fragments can prove critical for the successful retrieval of DNA sequences from particularly strongly degraded ancient biological material (Dabney et al., 2013; Meyer et al. 2014). We show that the loss of short molecules predominantly occurs in DNA extraction and that it can be minimized when adjusting the binding buffer composition in silica-based DNA extraction, allowing for efficient recovery of molecules as short as 25 bp. Furthermore, an analysis of damage-induced substitutions as well as the base composition of sequences generated with this protocol indicate that molecules with single-strand breaks are easily lost under non-optimal conditions. We also provide strategies for monitoring the loss of DNA in extraction and library preparation, which can be severe if inhibitory substances are co-extracted with DNA from ancient bones and teeth. We hope that the methods presented here will enable the recovery of DNA sequences from ancient specimens that failed to produce results in earlier screening attempts.

323B

aMPlex Torrent, a metabarcoding strategy relying on multiplex PCR and next generation sequencing adapted to the high-throughput analysis of ancient remains.

Thierry Grange, Silvia Guimaraes, Julien Daligault, Mélanie Pruvost, Eva-Maria Geigl
Institut Jacques Monod, Paris, France

Genotyping of multiple ancient skeletal remains with poor DNA preservation requires a sensitive high-throughput method to generate sufficient data. Here, we present a cost-effective metabarcoding approach, aMPlex Torrent, relying on multiplex PCR adapted to highly degraded DNA, thus targeting very small amplicons using thoroughly optimized primer pair combinations. Independent barcoding of sample replicates and next-generation sequencing allows to analyze simultaneously many samples that can even be heterogeneous in a time- and cost-efficient manner. We demonstrate the power of this approach through genotyping of ancient rodent remains dated up to 70,000 years and originating from a warm region that adversely affects DNA preservation, Morocco a recognized biodiversity "hot spot". We were able to reconstruct the phylochronology of Meriones preserved in a cave whose stratigraphy covers the last ~120,000 years, thus linking the Pleistocene to the present. Our current approach is fully operational for the genotyping of rodents (Rodentia), which are important for epidemiology, agronomy and ecological investigations and are bioindicators for human- and/or climate-induced environmental changes. We also applied the aMPlex Torrent approach successfully to several other taxonomic groups showing that it should have general utility in ecology, conservation biology, palaeontology, and archaeology.

324C

Optimisation of DNA extraction protocols from historic formalin-fixed soft tissues for retrospective genomic analyses

Giada Ferrari, Gülfirde Akgül, Frank J. Rühli, Abigail S. Bouwman
Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland

Formalin-fixed wet specimens from museums and pathologic-anatomical collections represent an extensive and disease-specific archive for retrospective molecular investigations. Their precise dating and available diagnoses make them an ideal source for the reconstruction of historic human and pathogen genomes. However, while formalin fixation successfully preserves tissue integrity, it degrades DNA into short fragments, similar to ancient DNA. Furthermore, formaldehyde creates cross-links between nucleic acids and proteins, which results in DNA being unavailable for downstream applications. Protocols that apply high thermal energy during tissue lysis to reverse cross-linking have been reported to successfully isolate DNA from formalin-fixed tissues. However this results in a trade-off between cross-link reversal and further DNA fragmentation. The aim of this study is to develop a methodology to obtain DNA suitable for high-throughput sequencing from archival formalin-fixed tissues while balancing cross-link reversal and thermal-induced DNA fragmentation. To this end we evaluated ten protocols for DNA extraction from 19th and early 20th century museum formalin-fixed wet specimens. The samples were obtained from the pathologic-anatomical collection in the Narrenturm at the Vienna Museum of Natural History and consist of autopsy material from individuals that had suffered from, or died of, tuberculosis, leprosy or anthrax. Quantitative PCR assays, as well as length, quality and mappability to the human and pathogen genome references of Illumina shot-gun reads are used to evaluate the different extraction protocols. The development of a reliable DNA extraction protocol is of crucial importance to unlock the potential of formalin-fixed wet specimens for retrospective genomic investigations.

325D

Mitochondrial genome diversity in European Upper Paleolithic and Mesolithic hunter-gatherers

Cosimo Posth [1], Alissa Mittnik[1], Dorothée G. Drucker[2], Hervé Bocherens[2,3], Katerina Harvati[1,3], Nicholas J. Conard[1,4], Johannes Krause[1,5]

[1] Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany, [2] Institute for Geosciences, University of Tübingen, Tübingen, Germany, [3] Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany, [4] Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Tübingen, Germany, [5] Max Planck Institute for the Science of Human History, Jena, Germany

Studies of ancient human mitochondrial DNA (mtDNA) suggest genetic continuity between Upper Paleolithic and Mesolithic hunter-gatherers in Europe, followed by an almost complete replacement with limited genetic admixture by Neolithic farmers. The analyses of European pre-Neolithic mtDNAs revealed a predominance of mitochondrial haplogroups belonging to clade U. By contrast, early Neolithic European farmers were found to belong to a wider range of different mitochondrial clades. However the mitochondrial genome variation during the Paleolithic and Mesolithic across Europe is currently poorly understood, as only a limited number of hunter-gatherers from those periods have been genetically analyzed. Environmental changes in the Late Pleistocene could have influenced human migration patterns and changed the European genetic population structure. In this study high-throughput sequencing technologies are adopted to reconstruct the complete or almost complete mtDNA of Late Pleistocene and Early Holocene modern humans from different archaeological sites in southwestern Germany. DNA is extracted from skeletal remains, converted in genetic libraries, the mtDNA is enriched through a bait capture technique and sequenced on a next generation sequencing platform. Ancient mtDNA authenticity is verified by analyzing typical ancient DNA damage patterns as well as establishing a single source of mtDNA for the studied individuals. Ten newly reconstructed hunter-gatherer mitochondrial genomes are here co-analyzed with previously published ancient and modern complete mtDNA sequences and used to address questions about European genetic diversity through time and space.

326A

Tracing signatures of cattle improvement in Portugal

Luciana Simões[1], Emma Svensson[2,3], Simon Davis[4], Ana Elisabete Pires[1,5], Cleia Detry[6], José Matos[5], Anders Götherström[7], Catarina Ginja [1,8]

[1] CE3C, Centro de Ecologia, Evolução e Alterações Ambientais, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal, [2] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [3] Department of Evolutionary Biology and Science for Life laboratory, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden, [4] Laboratório de Arqueociências, DGCP, Lisbon, Portugal, [5] Grupo de Biologia Molecular, Instituto Nacional de Investigação Agrária e Veterinária, Oeiras, Portugal, [6] UNIARQ, Faculdade de Letras, Universidade de Lisboa, Lisbon, Portugal, [7] Stockholm University, Archaeological Research Laboratory, Stockholm, Sweden, [8] CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal

Cattle are useful to illustrate changes in inheritable traits associated with animal husbandry practices of different cultures. On the basis of size increase, several authors suggest that the Romans improved cattle (*e.g.* in England). In Portugal, however, we have no evidence for cattle improvement in Roman times. In order to register a size increase in Portuguese cattle we have to wait until the Christians re-invaded. We could determine the gender of a number of cattle metacarpals from 15th century Beja (Portugal, Post-Medieval), using molecular methods, and which corroborated osteometric information, *i.e.* a size increase of cattle occurred between Moslem and Christian periods (not due to a change of sex-ratio). We applied a multi-genetic marker approach to recover signatures of human-driven selection. The major aims were to determine whether Post-Medieval Christians *improved* cattle locally or if new stock were introduced from elsewhere, and to investigate in which way inheritable traits were modified. We sub-sampled 26 cattle metacarpals from Moslem Alcáçova de Santarém (9th-12th century AD) and 21 cattle metacarpals from 15th century Beja. High-throughput Illumina sequencing (HiSeq 2500) was used to investigate mitogenomic variation. Whereas, targeted 454-GS Junior sequencing (Roche) was used to examine the patterns of diversity around candidate genes associated with phenotypic traits. Genetic analyses of cattle bones from Moslem and post-Moslem period sites revealed genetic continuity which suggests local improvement. We present and discuss preliminary results of this innovative zooarchaeogenetics research regarding animal improvement in association with husbandry practices imposed by human cultures that inhabited Iberia.

327B

The Genomic connections between early farmers from Iberia and central Europe, and their relationship to populations from modern Spain

Torsten Günther [1], Cristina Valdiosera[2,1], Helena Malmström[1], Irene Ureña[3], Ricardo Rodriguez-Varela[3], Óddny Sverrisdóttir[1], Evangelia Daskalaki[1,4], Pontus Skoglund[1,9], Thijessen Naidoo[1], Emma Svensson[1,8], José María Bermúdez de Castro[5], Eudald Carbonell[6], Jan Storå[4], Eneko Iriarte[7], Juan Luis Arsuaga[3], José Miguel Carretero[3,7], Anders Götherström[3], Mattias Jakobsson[1]
[1] Uppsala University, Uppsala, Sweden, [2] La Trobe University, Melbourne, Australia, [3] Universidad Complutense de Madrid, Madrid, Spain, [4] Stockholm University, Stockholm, Sweden, [5] Centro Nacional de Investigación sobre la Evolución Humana, Burgos, Spain, [6] Institut Català de Paleoecologia Humana i Evolució Social, Tarragona, Spain, [7] Universidad de Burgos, Burgos, Spain, [8] Swedish University of Agricultural Sciences, Uppsala, Sweden, [9] Harvard Medical School, Boston, USA

The neolithization process swept over Europe after the advent of farming lifestyle in the Near East approximately 11,000 years ago. However, the mode of transition and its impact on the demographic patterns of Europe remains an area of open questions. Ancient genomics allow us to analyze individuals involved in these transitions directly and to make comparisons between populations over time. Previous studies have shown close relationships between early Scandinavian farmers and contemporary southern Europeans as well as strong differences between hunter gatherers and early farmers. However, mtDNA composition on the Iberian peninsula and different migration routes suggest a dissimilar history of southwestern Europe. We genomic sequences of eight between 4,000 and 5,600 year old early Iberian farmers from El Portalon, 15 km East of Burgos, Spain. In contrast to a 7,000 year old hunter gatherer from the near-by area La Brana, but similar to the pattern observed for central and northern European farmers, these individuals all show genetic similarities to modern-day southern Europeans, especially to Sardinians and Basques.

26 Genomic and epigenomic evolution of sex chromosomes: Broad patterns and intriguing cases

26.1

Genomic evolution of two Y chromosomes in papaya

Ray Ming
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Papaya has a male heterogametic sex chromosome system with two slightly different Y chromosomes, Y controlling males, and Y^h controlling hermaphrodites. We sequenced the X, Y, and Y^h chromosomes of papaya, yielding 8.1 megabase (Mb) each for male and hermaphrodite specific region of the Y chromosome (MSY and HSY), and a 3.5 Mb sequence for the corresponding X region. The larger size of the MSY and HSY is mostly due to retrotransposon insertions. The MSY and HSY regions have highly similar gene content and structure, and 99.6% sequence identity. The MSY sequences from wild males include three distinct haplotypes, associated with the populations' geographic locations, but gene flow is detected for other genomic regions. The Y^h sequence is highly similar to one Y haplotype (MSY3), found only in wild dioecious populations from the north pacific region of Costa Rica. The low MSY3-Y^h divergence supports the hypothesis that hermaphrodite papaya is a product of human domestication. We estimate that Y^h arose about 4,000 years ago, after crop domestication in Mesoamerica more than 6,200 years ago, but coinciding with the rise of the Maya civilization. The Y^h chromosome has lower nucleotide diversity than the Y and the pseudo-autosomal regions (PAR), consistent with a domestication bottleneck. Nucleotide diversity in the domesticated X chromosomes is half that in the wild Xs and one tenth that of PARs. The identification of the ancestral MSY3 haplotype will expedite investigation of the mutation leading to the domestication of the hermaphrodite Y^h chromosome.

26.2

Faster-X effects and gene expression bias in *Drosophila melanogaster*

Brian Charlesworth, Victoria Avila, Jose Campos
University of Edinburgh, Edinburgh, UK

Genes located on the X chromosome in *Drosophila* may experience faster rates of adaptive evolution relative to comparable autosomal genes, either because of their higher effective recombination rate, or because of the exposure of favourable X-linked partially recessive mutations to selection in males. Comparisons of species of the *melanogaster* subgroup indicate a faster rate of protein sequence evolution for X-linked relative to autosomal genes, but evidence for this is largely lacking in comparisons among *D. pseudoobscura* and its relatives. Analyses of polymorphism and divergence suggest that the "Faster-X" effect in the *melanogaster* subgroup is due to a higher rate of fixation of favourable nonsynonymous mutations, and that this effect is not caused by a higher effective X chromosomal recombination rate, or by differences in patterns of sex-biased gene expression.

26.3

GENETIC AND EPIGENETIC MECHANISMS TO TRIGGER OFF SEXUAL DIMORPHISM, Example of intriguing dioecious Schistosomatidae parasite.

Marion Picard, Christoph Grunau, David Roquis, Jean Francois Allienne, David Duval, Jérome Boissier, Céline Cosseau
(1) Univ. Perpignan Via Domitia, Perpignan, France

Among the 24.000 species of classically hermaphroditic trematodes, the hundred species of Schistosomatidae are intriguing because they are gonochoric.

Sex of Schistosomes is genetically determined, but there is no phenotypic dimorphism in the larval stages: sexual dimorphism appears only in adult worms. Female Schistosomes are ZW heterogametic and lack global dosage compensation. The W heterochromosome is devoid of W specific genes and instead, it presents blocks of satellite repeat sequences. Moreover, the chromatin structure of these repeat sequences changes throughout the parasite life cycle.

In order to connect the apparition of phenotypic sexual dimorphism to molecular mechanisms, we present RNA-Seq and ChIP-Seq data on specific histone marks on larval stages, differentiating stages (schistosomula) and adults. We observe a change of the dosage compensation mechanism during the developmental stages: global compensation is observed in larvae, whereas gene specific compensation occurs during sex differentiating stages. In short, schistosomes seem to reproduce on an ontogenic level in every life cycle what is thought to happen during evolution of autosomes into sex chromosomes in other species. RNA-seq data suggest an intense remodeling of the chromatin in male schistosomula, and ChIP-seq data reveals differences in chromatin structure between male and female but also, from larvae to adults. Altogether, these elements argue in favor of the epigenetic commitment to modulate gene dosage compensation during the parasite development and we hypothesize that this change in dosage compensation is necessary to trigger the sexual dimorphism in Schistosomes.

26.4

Unprecedented support for Ohno's hypothesis of dosage compensation from an ancient reptilian sex chromosome system

Ray Marin [1,2], Diego Cortez[1,2], Shantha Mahadevaiah[3], Philippe Julien[4], Angélica Liechti[1], Jean Halbert[1], Delphine Valloton[1], Kelly Ascencao[1], Halie N. Kerver[5], Juli Wade[5,6], Patrick Tschopp[7], James Turner[3], Henrik Kaessmann[1,2]

[1] Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3] Division of Stem Cell Biology and Developmental Genetics, Medical Research Council, National Institute for Medical Research, London, UK, [4] EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation, Barcelona, Spain, [5] Neuroscience Program, Michigan State University, East Lansing, USA, [6] Department of Psychiatry, University of Michigan, Ann Arbor, USA, [7] Department of Genetics, Harvard Medical School, Boston, USA

Sex chromosomes emerged from different ancestral autosomes in various vertebrate lineages. However, sex chromosomes have so far mainly been scrutinized in mammals and birds, and the functional evolution of XY systems outside of the mammalian clade remains unexplored. Here we trace the evolution of the XY chromosomes of the green anole lizard (Anolis carolinensis), on the basis of high-throughout genome/transcriptome sequencing and complementary cytogenetic data, and revisit dosage compensation evolution in representative mammals and birds with substantial new expression data. Our analyses reveal that Anolis sex chromosomes represent an ancient XY system that emerged ≈155 million years ago in the ancestor of Iguania lizards, shortly after the separation from the snake lineage, and was only recently replaced by a ZW (female heterogametic) system in some agamids. The Anolis Y retained only seven protein-coding genes, the smallest number of any sex-specific chromosome described so far, and four of them evolved testis-specific expression. Remarkably, to compensate for the almost complete Y chromosome degeneration, X-linked genes have overall become twofold upregulated specifically in males, thus restoring ancestral expression levels in this sex and superseding the need to evolve female X inactivation (absence of X inactivation was confirmed by cytogenetic experiments). The highly efficient dosage compensation mechanism of Anolis is reminiscent of that of Drosophila melanogaster and represents the so far only vertebrate case to support Ohno's original dosage compensation hypothesis. Overall, our study lends support to the hypothesis that patterns of dosage compensation mainly depend on properties of ancestral sex chromosome precursors.

26.5

Sexual selection drives short- and long-term evolution of the avian Z chromosome

Alison Wright [1], Peter Harrison[1], Fabian Zimmer[1], Stephen Montgomery[1], Marie Pointer[2], Judith Mank[1]

[1] University College London, London, UK, [2] University of Oxford, Oxford, UK

Comparisons between sex chromosomes and autosomes can reveal the magnitude of sexual conflict acting throughout the genome. We investigated the role of sexual selection in shaping gene expression and coding evolution of the Z chromosome in several species within the Galloanserae, a clade of birds with a remarkable range of sexual selection. First, we mapped out the fine-scale evolutionary history of the sex chromosomes using genomic, transcriptomic and phylogenetic data to reveal that independent inversions spread progressively across the sex chromosomes (Wright et al. 2012; Wright et al. 2014). Using this map, we show that over long evolutionary periods, the Z chromosome has been convergently and successively masculinized (Wright et al. 2012). This is consistent with theoretical predictions that Z-linked genes should be more often selected for male-benefit alleles due to their unequal inheritance pattern. We next examined the role of mating system in driving masculinization of the Z chromosome over shorter evolutionary timespans, using phylogenetically controlled comparisons among species with high and low levels of sexual selection. Our results, combining coding sequence, polymorphism and gene expression with phenotypic measures of sexual selection, indicate that variance in male reproductive success in promiscuous species reduces the effective population size of the Z chromosome, leading to relaxed purifying selection acting on the coding content (Wright et al. in press). This in turn limits the adaptive role of the Z chromosome in general, and its potential role in encoding sexually selected traits in particular.

26.6

A recent bottleneck of Y chromosome diversity coincides with a global change in culture

Toomas Kivisild
University of Cambridge, Cambridge, UK

It is commonly thought that human genetic diversity in non-African populations was shaped primarily by an out-of-Africa dispersal 50-100 kya. While the male to female effective population size ratio has been estimated as being below one throughout much of human evolutionary history the factors affecting its dynamics are still poorly understood. Here, we present a study of 459 geographically diverse high coverage sequences of Y chromosome, including 299 newly reported samples. Applying ancient DNA calibration we date the Y-chromosomal Most Recent Common Ancestor (MRCA) in Africa at 254 (95% CI 192-307) kya and detect a cluster of major non-African founder haplogroups in a narrow time interval at 47-52 kya, consistent with a rapid initial colonization model of Eurasia and Oceania after the out-of-Africa bottleneck. In contrast to demographic reconstructions based on mtDNA, we infer a second strong bottleneck in Y-chromosome lineages dating to the last 10 ky during which the male to female ratio of effective population size dropped below 1/15th. We hypothesize that this bottleneck is caused by cultural changes affecting variance of reproductive success among males.

26.7

Diversity varies across recombining and non-recombining regions of the human sex chromosomes

Melissa Wilson Sayres
Arizona State University, Tempe, AZ, USA

Although the human sex chromosomes, X and Y, do not recombine over most of their lengths, there are two small regions that do actively recombine during male meiosis. These pseudoautosomal regions (PARs) are sex-linked, but undergo recombination in males and females and are not sex-specific. PAR1 spans the distal 2.6 Mb on the short arm of sex chromosomes; a smaller second pseudoautosomal region, PAR2, spans the distal 320 kb on the long arm of each sex chromosome. It is well established that both nucleotide diversity and recombination rates are substantially higher in PAR1 than in the sex specific regions. Although they comprise only a small percentage of the genome, the PARs house essential genes for both sexes and provide a unique opportunity to explore the dynamics of sex chromosome evolution. We measure diversity along the entire length of the X in twenty-six unrelated females, observing differences between PAR1, PAR2, and nonPAR regions. First, consistent with previous observations, we find that that average values of diversity in PAR1 are greater than the sex-specific nonPAR sequence. Second, curiously, we observe that diversity in PAR2 is not significantly different from diversity in the nonPAR regions. Finally, in light of recent studies suggesting the X transposed region (XTR) is a third PAR, we also investigate whether diversity is increased in this region of the X chromosome and find that the XTR does, indeed, exhibit consistently higher diversity than nonPAR sequences, suggesting it may still be actively recombining in humans.

26.8

Comparative Analysis of the Y Chromosome Genomes from Greater Apes

Monika Michalovova [1], Marta Tomaszkiewicz[1], Samarth Rangavittal[1], Rebeca Campos Sanchez[1], Howard W. Fescemyer[1], Oliver Ryder[2], Malcolm Ferguson-Smith[3], Rayan Chikhi[4], Paul Medvedev[5,6], Kateryna D. Makova[1]

[1] Department of Biology, Pennsylvania State University, University Park, PA 16802, USA, [2] San Diego Zoo Institute for Conservation Research, CA, USA, [3] Department of Veterinary Medicine, University of Cambridge, Madingley Rd, Cambridge, CB3 0ES, UK, [4] University of Lille 1/CNRS, Lille, France, [5] Department of Computer Science and Engineering, University Park, PA 16802, USA, [6] Department of Biochemistry and Molecular Biology, University Park, PA 16802, USA, [7] Genome Sciences Institute of the Huck, Pennsylvania State University, University Park, PA 16802, USA

The female genomes of the four sequenced great ape species — human, chimpanzee, gorilla, and orangutan — have diverged from each other by less than 3%. What about the male-specific part of the genome, the Y chromosome? Even though chimpanzee and human have been shown to be highly identical at the whole-genome nucleotide level (98.8% identity), their Y chromosomes have been found to be highly divergent with more than 30% of non-homologous sequences. Furthermore, gorilla Y chromosomal X-degenerate genes appear to be more similar in sequence to those of human than to the ones of chimpanzee.

The gorilla and orangutan Y chromosome sequences have been the missing pieces for comparative analysis of the four great ape Y chromosomes. In this study, we sequence whole genome amplified flow-sorted gorilla and orangutan Y chromosome DNA with both short-read (Illumina) and long-read (PacBio) technologies, using a strategy previously established in our lab. We develop new methods for assembling Y-chromosome specific sequences and combine them with existing tools to generate draft gorilla and orangutan Y chromosomes. We estimate the divergence level, evaluate copy number of ampliconic genes, and detect rearrangements among great ape Y chromosomes. Our results indicate that great ape Y chromosomes are remarkably different in size, repeat content, and gene variation. We demonstrate the utility of the Y chromosome sequences to conservation genetics of great apes.

26.9

Wolbachia bacterial endosymbionts and the evolution of sex determination in the isopod Armadillidium vulgare

Sébastien Leclercq, Julien Thézé, Isabelle Giraud, Amine Chebbi, Bouziane Moumen, Lise Ernenwein, Pierre Grève, Clément Gilbert, Richard Cordaux
CNRS / Université de Poitiers, Poitiers, France

In the isopod Armadillidium vulgare, sex determination (SD) follows female heterogamety (ZZ males and ZW females). However, many A. vulgare populations harbor maternally-inherited Wolbachia bacterial endosymbionts which can convert genetic males into phenotypic females, leading to populations with female-biased sex ratios. The W sex chromosome has been lost in lines infected by Wolbachia and all individuals are genetic males. Female sex is determined by Wolbachia infection of the A. vulgare individual, thereby shifting from chromosomal to cytoplasmic SD. Surprisingly, some A. vulgare lines exhibit sex ratio biases despite the lack of Wolbachia. In these lines, female individuals are genetic males carrying an unknown feminizing factor. To elucidate the genetic basis of female SD in these lines, we sequenced the genome of a female. We identified a large piece of the Wolbachia genome transferred to the A. vulgare nuclear genome. The transferred genomic fragment co-segregates perfectly with female sex in pedigrees. Our results indicate that SD in these A. vulgare lines is under control of nuclear gene(s) of bacterial origin. More generally, they emphasize that bacterial endosymbionts are powerful sources of evolutionary novelty, e.g. by driving shifts in SD mechanisms in their animal hosts. Funded by an ERC Grant to RC.

349A

A transcriptome derived female-specific marker in the invasive Western mosquitofish Gambusia affinis

Dunja Lamatsch [1], Sofia Adolfsson[2], Stefan Müller[3], Alistair Senior[4], Manfred Schartl[5], Shinichi Nakagawa[6]

[1] University of Innsbruck, Research Institute for Limnology, Mondsee, Austria, [2] Uppsala University, Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala, Sweden, [3] LMU Munich, Institute of Human Genetics, Munich, Germany, [4] University of Sidney, Charles Perkins Centre / School of Biological Sciences, Sidney, Australia, [5] University of Würzburg, Physiolocial Chemistry, Würzburg, Germany, [6] University of Otago, Department of Zoology, Otago, New Zealand

Sex-specific markers are a prerequisite for understanding the reproductive biology, the genetic factors involved in sex differences, the mechanisms of sex determination and the evolution of sex chromosomes. The Western mosquitofish, Gambusia affinis is of particular interest as a model for sex chromosome evolution due to its female heterogamety (ZW/ZZ), but also because of its ecological relevance as world-wide invasive species.
Here, deep-sequencing transcriptomes were used to identify sequences that were highly transcribed in females but not in males. 136 primer pairs spanning 78 contigs were tested by PCR to identify sex-specific amplification products. From those one female-specific DNA marker was developed and validated in 115 fishes. Sequence analyses revealed a high similarity to the 3´ UTR of the aminomethyl transferase (AMT) gene of the relatively closely related platyfish (Xiphophorus maculatus). The gene was sequenced in G. affinis and its location identified on the W chromosome by fluorescent in-situ hybridization.
This is the first time that such a methodological approach was successfully used in a fish species, and the identified sex-specific marker represents one of a handful of such markers in fish species.

350B

Sex chromosomes in plants - are they comparable with animals?

Roman Hobza, Boris Vyskot
Institute of Biophysics, AS CR, v.v.i., Brno, Czech Republic

In contrast to animals, plant species with separate sexes and sex chromosomes are very rare. As our knowledge of the architecture of sex chromosomes in individual plant models increases, the question arises whether there is degeneration of the non-recombining part of the Y chromosome, as is found in many animal species. Is reduced gene expression followed by dosage compensation in plants? What is the impact of transposable elements in plant sex chromosome evolution? Much data have been published but they lead to ambiguous conclusions. We speculate that although molecular analyses indicate degeneration, at least in some plant species, the evolutionary processes forming sex chromosomes in plants may differ from those in animals.

351C

Speciation by Rapid Sex Chromosome Evolution

Heini Natri, Takahito Shikano, Juha Merilä
University of Helsinki, Helsinki, Finland

Different sex determination and sex chromosome systems are often found in closely related species in lower vertebrates. It has been hypothesized that transitions in these systems are associated with the formation of new species. However, it is unknown whether such transitions lead to, or result from, the evolution of reproductive isolation and speciation. We address this issue by investigating rapidly evolving sex chromosomes and reproductive isolation in ninespine sticklebacks (*Pungitius pungitius*).

We found that both XX/XY and ZZ/ZW systems have evolved in globally distributed lineages that diverged within the last 0.7 million years. Genetic analyses indicated that the morphologically and genetically differentiated X and Y chromosomes were formed by historical hybridization between two ancestral lineages approximately 0.3 million years ago. Recombination between the X and Y was largely suppressed most likely due to genetic incompatibility, associated with a chromosomal re-arrangement in the ancestral lineages. Male-specific sterility was found in F_1 hybrids of particular lineage combinations, and genetic mapping revealed sex chromosomal conflict as a cause of hybrid sterility.

Our study demonstrates rapid evolution of sex chromosomes within a species, and provides evidence for a significant role of sex chromosome evolution in the formation of reproductive isolation, and thus speciation.

352D

Great ape Y chromosome diversity reflects social structure and sex-biased behaviours

Pille Hallast [1], Pierpaolo Maisano Delser[1,2], Chiara Batini[1], Daniel Zadik[1], Werner Schempp[3], Mariano Rocchi[4], Chris Tyler-Smith[5], Mark Jobling[1]

[1] University of Leicester, Department of Genetics, Leicester, UK, [2] Museum National d'Histoire Naturelle, Ecole Pratique des Hautes Etudes, Paris, France, [3] University of Freiburg, Institute of Human Genetics, Freiburg, Germany, [4] University of Bari, Department of Biology, Bari, Italy, [5] Wellcome Trust Sanger Institute, Hinxton, UK

The diversity of the male-specific region of the Y chromosome (MSY) in humans has been widely exploited to shed light on population history, sex-biased processes, and social selection. For humans a robust MSY phylogeny of haplogroups exists based on thousands of slow-mutating SNPs defining the relationships between different lineages. Analysis of the same locus in great apes, however, has been scanty and has mostly been based on MSY-specific STRs rather than SNPs. This is partially due to the fact that among great apes the Y chromosome has been sequenced only for human and chimpanzee.

Here, we used a custom enrichment approach based on the human reference and sequenced ~3.9 Mb of DNA from the genomes of 19 male great apes to high coverage. Combining our data with MSY sequences extracted from the whole genomes of 20 additional individuals (Prado-Martinez et al. 2013) yields a total sample of four bonobos, 19 chimpanzees, 10 gorillas, and two Bornean and four Sumatran orangutans. Analysis of these great-ape sequences retained between 2 and 3.6 Mb of human-orthologous MSY material per species, identified species-specific deletions and duplications and thousands of novel Y-chromosome SNPs defining highly-resolved MSY phylogenies. Comparison with mtDNA phylogenies from the same individuals, as well as autosomal and X-chromosomal SNP data provides insights into the social structure and sex-specific behaviors of each species.

353A

A shift in sex determination mechanisms in an experimental hybrid copepod population?

Barret C. Phillips, Eric T. Watson, Suzanne Edmands
University of Southern California, Los Angeles, CA, USA

Chromosomal sex determination (CSD) has evolved repeatedly, and interactions between sex chromosomes and autosomes frequently underlie postzygotic isolation (PZI). However, little is known about how PZI evolves in species lacking heteromorphic sex chromosomes. Where multiple, unlinked loci are involved in sex determination (polygenic sex determination, PSD), the invasion of alleles of strong effect, coupled with sexual antagonists, offers a potential route to the evolution of new sex chromosomes. Hybridization in species with PSD potentially involves the interaction of novel combinations of sex-determining alleles-as well as genetic incompatibilities-and, if coupled with subsequent inbreeding, may provide a flash point for transitions to CSD. Under this scenario, we expect family sex ratio variance to increase in early generation hybrids, followed by a reduction in extra-binomial variance, as family sex ratios in organisms with sex chromosomes are expected to follow a binomial distribution. We tested for a transition from PSD to CSD in the copepod Tigriopus californicus using an experimental hybrid swarm composed of two highly divergent populations. Population-diagnostic SNPtyping revealed the presence of sex-specific heterozygosity on chromosome 10, which was maintained for >4 years (~50 generations), despite genetic swamping of one population's alleles across most of the remainder of the genome. Association tests and PCA identified this chromosome as strongly associated with sex. However, after >7 years (~90 generations) of admixture, family sex ratios were not binomially distributed in the hybrid swarm. Experimental evolutionary studies such as this are instrumental in understanding connections between sex determination mechanisms and reproductive isolation.

354B

The hunt for the sex chromosomes of *Salix viminalis*

Pascal Pucholt, Sofia Berlin
Swedish University of Agricultural Sciences, Uppsala, Sweden

To understand the broad pattern of evolution of sex determination systems and sex chromosomes it is necessary to study the sex determination mechanisms of a large variety of systems. Our model is the angiosperm dioecious plant genus *Salix* in the Salicaceae family or more specifically the species *Salix viminalis*. Dioecy is in this family highly predominant but different sex determination systems exist with likely independent origin. We showed in previous work that *S. viminalis* is female heterogametic and that the sex determining locus is located on chromosome XV.
We have now used a denser *S. viminalis* linkage map and refined the position of the sex determining locus and are analysing the genomic sequence in this region in order to identify and separate sequences of the incipient Z and W chromosomes. With this insight we aim to be able to identify female specific genomic regions and candidate key regulatory elements involved in sex determination in *S. viminalis*. Analysing the divergence between Z and W chromosomes will tell us more about the evolutionary history of sex determination mechanisms.
Understanding the sex determination system in the *Salix* genus will give insights in the evolution of the sex determination systems of the Salicaceae family and will contribute to answer the general question of how exactly separate sexes and sex chromosomes evolve.

355C

tba

356D

Characterization of the Z-Chromosome of the Puerto Rican Parrot *(Amazona vittata*)

Edwin Ramírez-Aponte, Jonathan Acevedo-Nieves, Yásen Bonew-Rivera, Jenelle Bonilla-Perez, Christian Cardona-Feliciano, Zahudy Figueroa-Vázquez, Alfredo Reyes-Oliveras, Stephanie Sánchez-Maldonado, Ghildres Tejedor, Isamar Torres-Hernández, Guillermo Valero-David, Edgardo Vélez-Ruiz, Olga Vera-Colón, Jenniffer Hernández-Muñiz, Michelle Vega-Medina, Adriana Gordon-De Jesús, Gustavo Rivera-Álvarez, Ingrid Rivera-Pagán, Taras Oleksyk, Juan Martínez-Cruzado
University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico

Puerto Rican parrot (*Amazona vittata*) genome data became publicly available thanks to the community effort of the people of Puerto Rico. To advance the description of rearrangements, conserved regions, protein-coding genes and other important protein and gene features, we are focusing on sex chromosome Z from the latest genome assembly (http://genomes.uprm.edu/parrot). Some annotation resources and tools used for this work are: Stand-alone Blast, Blast from NCBI, UCSC Genome Browser, Gene Model Checker, Ensembl, MEGA 6, and Repeat Masker. Fluorescence in situ hybridization of chicken's chromosomes against Puerto Rican Parrot shows that chromosome Z from both species hybridize completely and do not present any translocation with other chromosomes. We started the annotation by identifying 937 scaffolds of the parrot genome matching to the chicken (*Gallus gallus*) chromosome Z. Chicken and budgerigar (*Melopsittacus undulatus*) sequence files were used as major templates for this work, but occasionally it was necessary to use other genomes such as that of the zebra finch (*Taeniopygia guttata*), turkey (*Meleagris gallopavo*), collared flycatcher (*Fidecula albicolis*), saker falcon (*Falco cherrug*) and peregrine falcon (*Falco peregrinus*). We analyzed the larger 66 scaffolds and found in them 211 genes, most of them with a molecular function for binding or catalytic activity . The total length of these scaffolds adds to 23.3 Mb, which represents approximately 28.4% to 29.1% of Z chromosome. Most of the analyzed scaffolds are in synteny with other bird species, but Scaffold 605 shows a unique inversion for *A. vittata*, and corresponds to the beginning of the chromosome.

357A

Ancient male recombination shaped genetic diversity of neo-Y chromosome in *Drosophila albomicans* .

Kazuhiro Satomura, Koichiro Tamura
Tokyo Metropolitan University, Tokyo, Japan

The origin of sex chromosomes is thought to be a pair of autosomes, from which Y chromosome is drastically differentiated from X chromosome following the cessation of meiotic recombination between X and Y chromosomes. Non-recombining Y chromosome is largely degenerated with decrease in genetic diversity due to the influence of natural selection such as hitchhiking effect. Indeed, the genetic diversity of neo-Y chromosome of *Drosophila miranda* is extremely low. However, it is also expected that the neo-Y chromosome has had no chance to acquire genetic diversity since originated from a single chromosomal fusion event between a canonical Y chromosome and an autosome under the condition of recombination-free male germline in *Drosophila* species. Here, we report another case of neo-Y chromosome evolution in *D. albomicans*. We analyzed neutral nucleotide variations in 53 protein-coding genes, demonstrating that the neo-Y linked genes have as large genetic diversity as the neo-X linked genes in contrast to the case of *D. miranda*. Our cross experiment confirmed no male recombination in the current *D. albomicans* population but in its sibling species, *D. nasuta*, whereas the extensively shared polymorphism between the neo-X and the neo-Y chromosomes suggests ancestral recombination between them, which introduced genetic diversity of the neo-X chromosome into the neo-Y chromosome. After the recombination was ceased about a quarter million years ago estimated by a method of molecular clock, the high genetic diversity of the neo-Y chromosome has been maintained until now, suggesting no natural selection has worked to reduce the genetic diversity.

358B

Comparative Analyses of Y Chromosomes Across the Order Diptera

Shivani Mahajan, Doris Bachtrog
University of California, Berkeley, Berkeley, USA

Across the order Diptera, which includes the model organism *Drosophila melanogaster*, gene content and karyotype is largely conserved. However, recent studies have revealed that there have been numerous independent losses and gains of sex chromosomes within this order. There is a considerable amount of variation in the presence and nature of the sex chromosomes and in the percentage of genome that is sex-linked: Both male and female heterogamety have been observed; in some species the sex chromosomes are homomorphic, whereas others have heteromorphic sex chromosomes showing varying degrees of Y degeneration. These species, therefore, provide an interesting framework to study Y chromosome evolution in a comparative manner, and can shed light on the evolutionary forces shaping it. In most species with heteromorphic sex chromosomes, the Y chromosome is generally gene poor and the few genes that remain on it are thought to be important for male fertility and development. However, one major obstacle for evolutionary studies, is that the Y chromosome is usually heterochromatic and repeat-rich which makes both its sequencing as well as assembly extremely difficult. Recently, methods have been developed that aim to extract coding sequences from the Y, without necessarily having to assemble the entire chromosome. Combining elements from such studies, we use whole genome sequencing data as well as RNA-seq data from 24 species, sampled broadly across the Order, to identify candidate Y-linked transcripts for each of them. Our results indicate that there is little conservation of Y-linked genes between different species in our study.

359C

The evolution of suppressed recombination in Silene latifolia and S. dioica

Jose L Campos [1], Suo Qiu[2], Roberta Bergero[1], Sara Guirao-Rico[3], Deborah Charlesworth[1]
[1] University of Edinburgh, Edinburgh, UK, [2] University of Sheffield, Sheffield, UK, [3] Center for Research in Agricultural Genomics, Cerdanyola del Vallès, Barcelona, Spain

The plant genus Silene includes several species with separate sexes (dioecious species). Silene latifolia and its close relative S. dioica belong to a group of 5 species that have sex chromosomes (with XY males and XX females). These sex chromosomes evolved < 10 My, and are a good system for studying in detail the evolution of suppressed recombination between sex chromosomes.

Previous genetic mapping in Silene latifolia identified multiple genes in a fully sex-linked region (where recombination between the Y and the X chromosomes is suppressed) and others in a pseudo-autosomal region (or PAR) where recombination still occurs. Population genetic analysis of several genes located close to the boundary between the PAR and the fully sex-linked region shows a mixture of male-specific variants (indicating that these variants are Y-linked), and variants that show recombination between the Y and the X. In these genes, recombination between the X and Y chromosomes probably stopped very recently. To test when the suppression of recombination in this region occurred, we performed a similar analysis in S. dioica, focusing on these genes. In S. dioica, we found that most sites that behave as sex-linked in S. latifolia do not appear fully sex-linked. We conclude that recombination is not yet suppressed in these genes in S. dioica, but occurred after the speciation event between S. latifolia and S. dioica.

360D

Studying the Evolution of Young Sex Chromosomes in Drosophila *albomicans*

Qi Zhou, Dat Mai, Doris Bachtrog
University of California Berkeley, Berkeley, California, USA

In many species, sex chromosomes have become very divergent from each other, making it difficult to study the processes driving their evolution. Drosophila *albomicans* is a unique model organism where young sex chromosomes (dubbed neo-sex chromosomes) formed approximately one hundred twenty thousand years ago, through the fusion of their ancestral sex chromosomes to a pair of autosomes comprising roughly 40% of the entire genome. This provides us with an opportunity to study the very initial steps of the evolution of diverging sex chromosomes.
By comparing the neo-X and neo-Y genome sequences of D. *albomicans* to each other and to the homologous autosome sequence in a sister species, Drosophila *nasuta*, we found that the neo-Y has an excess of non-functional genes relative to any other chromosome and that the rate of protein evolution of the neo-Y is significantly increased relative to the neo-X. This indicates weaker purifying selection acting on the neo-Y after it lost the ability to recombine. We also assayed expression pattern of genes on the neo-sex chromosomes in male and female D. *albomicans* tissues, and in homologous D. *nasuta* tissues, to study the impact of sex-specific or biased transmission on gene expression. Finally, we profiled heterochromatic and euchromatic epigenetic markers of the neo-Y and neo-X in D. *albomicans*, to reveal the epigenomic evolution of this pair of young sex chromosomes.
Overall, we report the evolution of newly formed sex chromosomes relative to their autosomal counterparts as well as various ways that they differentiate over time relative to each other.

361A

A Y-chromosomal reference panel to detect single copy nucleotide polymorphisms for paternal line tracking in horses

Barbara Wallner [1], Nicola Palmieri[2], Claus Vogl[1], Christian Schlötterer[2], Gottfried Brem[1]
[1] Insititute of Animal Breeding and Genetics, Vetmeduni, Vienna, Austria, [2] Institute of Population Genetics, Vetmeduni, Vienna, Austria

Polymorphic Y-chromosomal markers provide useful information for tracing male genealogies. Modern domestic horses exhibit very low Y-chromosomal sequence diversity. Therefore comparatively large regions of the non-recombining region of the Y (NRY) need to be screened to obtain a sufficient number of polymorphisms to characterise individual male lineages. A further obstacle to screening of polymorphisms is the poor quality of the NRY reference sequence: due to the high content of repetitive DNA, the NRY sequence is fragmented into small, unmapped scaffolds.

Here we use horse male and female short read NGS data and the comparison of de-novo assembled contigs to identify Y-linked sequences that are suitable as a reference for SNP calling. We found 2500 short contigs (totalling 1,7 MB) located on the horse NRY in a unique single copy. Using these sequences as a reference to screen whole genome NGS data from 46 horses from 21 modern breeds revealed 53 SNPs. No multiply mutated sites were observed. The polymorphisms result in 26 different haplotypes and the inferred haplotype tree reflects modern horse breeding practices. Based on high-quality pedigree data we estimate a de-novo mutation rate on the horse NRY to be about equal to that of humans. Thus we could assemble a NRY reference panel for polymorphism screening in horses and the method can be easily adapted to other species.

27 The origins of multicellularity under the light of functional genomics

27.1

S. rosetta as a simple model for animal multicellularity

Nicole King
HHMI/University of California, Berkeley, Berkeley, CA, USA

All animals are multicellular, yet little is known about how animal multicellularity first evolved. The choanoflagellate *Salpingoeca rosetta* is one of the closest living relatives of animals and its study promises to illuminate the origin of animal multicellularity. Just as animals develop from a single cell – the zygote – a founding *S. rosetta* cell can produce a multicellular rosette colony through multiple rounds of cell division coupled with stable cell adhesion. Remarkably, this morphological transition is regulated by specific small lipid signals produced by environmental bacteria. In my talk, I will describe our studies of intrinsic and extrinsic factors that regulate the switch to rosette development and discuss the implications of our findings for understanding animal origins and development.

27.3

Insights into evolution of eumetazoan regulatory developmental networks from the sea anemone Nematostella vectensis

Ulrich Technau
University of Vienna, Vienna, Austria

Cnidaria, the sister phylum of Bilateria, lacks a number of key bilaterian traits, such as mesoderm, a central nervous system and a clear bilaterality, yet the underlying genetic basis for these crucial differences is unknown. Genome sequencing projects have revealed that the anthozoan Nematostella vectensis displays a stunning ancestral complexity in gene repertoire, gene structure and genome organisation. Therefore, differences in the cis-transcriptional or post-transcriptional regulation as well as protein interactions may account for differences in functions of conserved genes. To this end, we mapped cis-regulatory elements (promoters and enhancers) on a genome-wide level using a combination of histone modifications and binding of Pol II and of transcriptional cofactor p300. We found that the basic logic of enhancers and promoters marked by conserved and specific combinations of histone modifications predates the origin of eumetazoans. To assess the evolution of specific gene regulatory networks we also determined all target genes of the conserved transcription factor Brachyury by ChIP-seq on a genome-wide scale and compared it with similar datasets from sea urchins and frogs. Our data reveal the evolutionary conservation and divergence of a the GRN of a conserved developmental regulator.

27.4

Origin of animal transcriptional regulation: insights from a sponge with both unicellular and multicellular genomic characteristics

Selene Fernandez-Valverde, Bernard Degnan
University of Queensland, Brisbane QLD, Australia

The complex genomic regulatory system that underlies the formation and maintenance of animal body plans must have evolved prior to the emergence of the crown Metazoa. Although early-branching animals possess most developmental transcription factor families, the nature of their regulatory DNA is currently unknown. Here we characterise core promoter sequences, splice sites (5'SSs) that bind U1 and polyadenylation sites (PASs) in the genome of an early-branching metazoan, the sponge *Amphimedon queenslandica*. With a median intergenic distance of 0.59 kb and genes largely comprised of tiny introns, the genome organisation of *Amphimedon* is more similar to unicellular opisthokonts than to other metazoans. In contrast, its genes possess metazoan-like core promoters that include features that have been deemed previously to be vertebrate-specific, including Nrf-1 and Kruppel-like binding motifs. *Amphimedon* often initiates bidirectional transcription from a single promoter. Although there does not appear to be differences in the composition of bidirectional and unidirectional promoters, there is a differential enrichment of PASs and depletion of 5'SSs in non-coding antisense DNA strand upstream of unidirectional promoter. This is consistent with these transcripts being short-lived, as observed in vertebrates. The dual nature of the *Amphimedon* genome – unicellular organisation with metazoan *cis*-regulatory elements – suggests that sponges have maintained genomic features that emerged early in metazoan evolution. This includes promoters populated by metazoan binding sites that can drive bidirectional transcription initiation, and the prevalence of PASs upstream and 5'SSs downstream to promote greater longevity of coding strand transcripts.

27.5

Functional evolution of a bacterial small RNA that controls multicellular fruiting body development

I-Chen Kimberly Chen [1,2], Gregory Velicer[1,2], Yuen-Tsu Nicco Yu[2]
[1] *Indiana University, Bloomington, Indiana, USA,* [2] *ETH Zurich, Zurich, Switzerland*

A rapidly growing body of evidence has shown that non-coding small RNAs (sRNAs) regulate a variety of important biological processes across all life domains, including bacteria. However, little is known about the functional evolution of sRNAs in bacteria, which might occur via changes in structures and stability, or interactions with associated regulatory networks. The sRNA Pxr in the model myxobacterial species *Myxococcus xanthus* functions as a developmental gatekeeper that prevents the initiation of multicellular fruiting body development until nutrients have been depleted. Here, we introduced the common ancestor of *pxr* and its extant alleles from different species into an *M. xanthus pxr* mutant to examine the functional divergence of Pxr. Our results showed that the regulatory interactions of Pxr observed in *M. xanthus* has been established since this ancestor. Also, all *pxr* alleles from species with only one copy of this gene controlled development in *M. xanthus* in a manner qualitatively similar to that of the native *M. xanthus* allele. Nevertheless, two paralogs found in the genus *Cystobacter* failed to control development. Our results therefore illustrate both that Pxr may play a common fundamental role in developmental gene regulation across diverse species of myxobacteria and that the specificity of its function may be evolving in some lineages. The effects of nucleotide changes on the positions conserved across *pxr* homologs were also examined. Future work will include constructing *pxr* mutants in species other than *M. xanthus* to elucidate the effects of their own *pxr* alleles in their native genomic background.

187A

Functional genomics of the protist Creolimax fragrantissima sheds light on pre-metazoan genome regulation and the convergent origins of fungal traits

Alex de Mendoza [1,2], Hiroshi Suga[2,3], Manuel Irimia[4], Iñaki Ruiz-Trillo[2]
[1] The University of Western Australia, Perth, Western Australia, Australia, [2] Institute of Evolutionary Biology, Barcelona, Spain, [3] Prefectural University of Hiroshima, Hobara, Japan, [4] Centre Regulacio Genomica, Barcelona, Spain

Understanding how animals became multicellular from a unicellular ancestor is one of the major questions in evolutionary biology. Recently, genome sequence analyses have revealed that many important genes involved in animal multicellularity are encoded in the genomes of their living unicellular relatives, suggesting that complex genome regulation pre-dated animal origins. To address this issue we have sequenced the genome of the animal-related protist *Creolimax fragrantissima* and profiled its transcriptome developmental dynamics. This species undergoes coenocytic development, which involves the transition from a single celled amoeboid stage to a multinucleate (coenocytic) growth stage. Strikingly, differential expression analysis shows richer signalling, transcription factor and cell adhesion activities in the single-celled amoeboid stage. Additionally, we describe regulated alternative splicing patterns in both coding genes and a population of animal-like long non-coding RNAs. Moreover we find a clear transcriptional co-regulation pattern of the core components of the integrin adhesome, which is a key animal cell adhesion complex involved in the attachment to the extracellular matrix (ECM). To address whether this species secretes any kind of ECM, we profiled its secretome through proteomics approaches. We find no evidence of proteins containing ECM domains, however the secretome of *Creolimax* reveals a clear adaptation to a specialized osmotrophic lifestyle similar to fungi, including convergent expansion of proteolytic enzymes and secreted genes acquired through lateral gene transfer from bacteria. Overall, functional comparative genomics indicates that despite not being as tightly regulated as animal development, *Creolimax* development shows key features of complex transcriptional regulation pre-dating the origins of multicellularity.

188B

Modular Evolution of Transcription Factors During the Rise of Multicellularity

Jonathan Schmitz [1], Fabian Zimmer[2], Erich Bornberg-Bauer[1]

[1] Institute for Evolution and Biodiversity, Westfälische Wilhelms-Universität Münster, Münster, Germany, [2] Department of Genetics, Evolution and Environment, University College London, London, UK

Transcription Factors (TFs) are proteins that modulate gene expression,

mainly by binding specific DNA sequences. TFs are important regulators of cellular processes and play an important role in the life cycles of all organisms. Here, we conducted a systematic analysis of TF family sizes in a range of metazoans to assay the evolution of these TF families in metazoans. We find a correlation of TF family sizes and organismal complexity as measured in number of cell types. This correlation hints at an importance of TF family expansion during the evolution of more complex multicellular organisms. Through ancestral reconstruction we find that in metazoans the evolution of TF families mainly occured via family expansion. We also analysed the role of domain arrangements in the expansion of TF families. For most TF families we find a correlation between the number of genes and domain arrangements in a TF family. This correlation hints at a possible role of the importance of domain rearrangements in the expansion of TF families. Possible explanations for this role are domain repetitions that elongate the specific target sequence of the TF (as seen in C2H2 zinc fingers) and additional domains that facilitate new functionality, e.g. by modifying the interaction of a TF with other proteins. Consequently we propose an important role of domain rearrangements in TF family expansions that coincides with an important role of TFs during the evolution of multicellular organisms.

189C

Differential evolution of non-coding DNA across eukaryotes could lead towards different pathways of multicellular complexity.

Irma Lozada-Chávez [1,2], Peter Stadler[1,2], Sonja Prohaska[1]

[1] Evo-Devo & Bioinformatics Group, Department of Computer Science (IZBI), University of Leipzig, Leipzig, Germany, [2] Max Planck Institute for Mathematics in the Science, Leipzig, Germany, [3] Fraunhofer Institut für Zelltherapie und Immunologie (IZI), Leipzig, Germany, [4] Department of Theoretical Chemistry, University of Vienna, Wien, Austria, [5] Santa Fe Institute, Santa Fe, USA

Genomes are far from being the most compact representations of biological information memorized in the form of genes. Instead, the extensive amount of ***non-protein-coding DNA*** (**ncDNA**) found in Eukarya has been invoked as the key genetic component in the evolution of ***complex multicellularity*** (**CM**) (Taft *et al*., 2007; Lozada-Chávez *et al*., 2011). Although functional relevance has been demonstrated for some instances of eukaryotic ncDNA, the distribution of different types of ncDNA within a genome, its relationship with genome size and its influence on the emergence of CM are still unclear and controversial (Petrov, 2001; Graur *et al*., 2013). In this work, we analyzed the distribution and contribution to the genome size of five types of ncDNA: *spliceosomal introns*, *repeats* (transposable elements and simple repeats), *pseudogenes*, *non-repetitive intergenic regions* and *non-protein-coding RNAs* (ncRNAs) in 350 completely sequenced genomes distributed across all major Eukaryote linages. Both repetitive and unique intergenic ncDNA sequences turn out to be fast builders of genomes, since their genomic content positively correlates with an increase in genome size. In contrast, introns, pseudogenes, ncRNAs and specific families of repeats evolve less strictly linked to genome size and are dominated instead by evolutionary history and life style. Our analysis underscores the pivotal role of introns and ncRNAs as promoters of CM both in basal forerunners to multicellularity and in complex multicellular organisms. Our findings suggest that the differential evolution of some ncDNA forms could be driving the emergence of independent pathways to multicellularity, despite their mainly nonadaptative origin in Eukarya.

190D

Alternative splicing as a source of innovation at the origin of animals

Xavier Grau-Bové [1], Manuel Irimia[2], Iñaki Ruiz-Trillo[1]

[1] Institut de Biologia Evolutiva, Barcelona, Catalonia, Spain, [2] Centre de Regulació Genòmica, Barcelona, Catalonia, Spain

Alternative splicing (AS) is a major mechanism of transcriptome regulation that has been described in a wide range of eukaryotes. The differential selection of splice sites constitutes a powerful force in evolution, e.g. enabling a diversification of the proteome or adding an additional layer of gene expression control. We explore the landscape of AS events in animals, their closest unicellular relatives and a wide range of other eukaryotes, aiming to understand how this relates to the emergence of multicellularity. Using high-coverage RNA-seq data, we quantify the frequency of two types of AS, exon skipping (ES) and intron retention (IR), and compare events across taxa. We reconstruct the evolution of regulated ES, a landmark feature of animal genomes that enables the establishment of sets of co-regulated splicing events marking different cell types. This fine-tuned regulation of the cell lineages' transcriptomic profile is crucial to the formation of multicellular organisms. We show that the ability to regulate ES across cell stages predates the origin of animal multicellularity, being present in the unicellular holozoans Capsaspora owczarzaki (Sebé-Pedrós et al. 2013, eLife), Creolimax fragrantissima and Sphaeroforma arctica, among other eukaryotes. In most of the analysed eukaryotes, however, IR is the most frequent form of AS. Notably, we find that there is an extremely low degree of conservation of AS events across eukaryotes, with new AS events being evolved in every lineage. Altogether, these results reinforce the view that the foundations of animal genomic complexity were already laid long before the emergence of multicellularity.

191A

A broad genomic survey reveals the evolutionary hierarchy, species specific expansions and N-terminal domain diversifications of “Adhesion” class G protein-coupled receptor families in holozoans

Arunkumar Krishnan, Helgi Schiöth
Uppsala University,Department of Neuroscience, Functional Pharmacology Unit, Uppsala, Sweden

Cell-cell adhesion is one of the major factors involved in transition to metazoan multicellularity from unicellular holozoans. Here, we survey a wide dataset of metazoan and the closest unicellular non-metazoan genomes to reconstruct the evolutionary history of one of the crucial family of receptors that participate in cell-cell adhesion, the “Adhesion” class G protein-coupled receptors (aGPCRs). In humans and most vertebrates, aGPCRs are classified in to nine distinct groups/families and possess a highly diverse N-terminal domain architectures containing large number of classical cell adhesion domains. Our phylogenetic analysis suggests that the emergence of most vertebrate-like aGPCRs families are largely confined within metazoans, with homologues of family III, IV, V and VIII found across most bilaterians, while family I, II and VII members are largely restricted to deuterostomes.Furthermore, several metazoans encode many species specific aGPCRs that do not belong to any of these known families. Intriguingly, in comparisons to unicellular relatives, aGPCR repertoires have expanded largely in metazoans, and indeed the largest expansions are mainly found in the most basal non-bilaterian metazoans, suggesting that transition to metazoan multicellularity perhaps required a more expanded aGPCR system. Moreover, the diverse arrays of vertebrate aGPCR N-terminal domains are also present in almost all analyzed metazoans, while most non-metazoan aGPCRs contains only the core seven-transmembrane segment and the GPCR proteolysis site (GPS). Overall, our analysis postulates the dynamic evolution of aGPCRs and provides a basis to further compare the diversifications of aGPCR system and its cell-adhesion functions in various metazoan lineages.

192B

Genome annotation of *Acrasis kona*

Sanea Sheikh, Chengjie Fu, Sandra Baldauf
Uppsala University, Uppsala, Sweden

Acrasids are single-celled amoebae that can undergo aggregative multicellularity in response to adverse environmental conditions, similar to the well-studied dictyostelid social amoebas. However, acrasids are unrelated to dictyostelids (supergroup Amoebozoa), being instead the only multicellular lineage in the eukaryotic supergroup Excavata. This makes *Acrasis* an interesting model system to study parallel evolution of social behavior in microbes as well as to explore the diversity of eukaryotes in general. We have sequenced the genome and transcriptome of *Acrasis kona* and are currently preparing transcriptomes from the four main stages of its life cycle. In initial work, we assembled the complete *A. kona* mitochondrial genome (mtDNA) and find that it is missing 14 protein genes present in the mtDNA of its closest sequenced relative, *Naegleria gruberii*. We further identified 11 of these protein genes in *A. kona* nuclear DNA and find that they carry mitochondrial important signals (transit peptides, Fu et al. 2014). We are now using RNAseq data and the *N. gruberii* genome in an annotation pipeline to create a fully annotated *A. kona* nuclear assembly. The results will be used to investigate parallel evolution of simple multicellularity, early steps in the evolution of eukaryotes and to aid in resolution of the eukaryote tree of life by breaking up some of the longer deep branches.

193C

Acrasis kona genomics: parallel evolution of aggregative multicellularity and rampant genetic exchange among eukaryotic soil microbes

Cheng-Jie Fu, Sanea Sheikh, Sandra Baldauf
Department of Organismal Biology, Uppsala University, Uppsala, Sweden

Acrasids (Heterolobosea, Excavata) are unicellular soil-dwelling amoebae. However when starved, they can cooperate to form simple tree-like fruiting bodies consisting of spores from formerly free-living cells. This type of quasi-multicellular behavior is well studied in dictyostelids (e.g. Dictyostelium discoideum). However, molecular phylogeny shows that "aggregative multicellularity" has evolved independently in acrasids and dictyostelids. Thus it makes Acrasids the only mullticellular taxon in the vast, ancient and poorly studied eukaryotic supergroup Excavata, for which there is very limited molecular data and a high potential for the discovery of genetic novelty. This is a challenging undertaking because the closest free-living relative of Acrasis with complete genome sequence data is Naegleria gruberi, the only completely sequenced heterolobosean. There is also considerable genome data from the more distantly parasites Trypanosoma and Leishmania, although these genomes are highly derived and reduced, as well as several hundred million years distantly related. Our initial analyses of genomic and transcriptomic data suggest the presence of illegitimate genetic exchange between Acrasis and other eukaryotic soil microbes, particularly dictyostelids, currently the main source of data on eukaryotic soils microbes. It will be particularly interesting to investigate whether these transfers involve specific metabolic pathways, possibly even some involved social behavior.

28 The golden age of Archaea: unveiling the diversity and evolution of the third Domain of Life

28.1

From enigmatic curios to global players: current understanding of the diversity and unique contribution of Archaea to global biogeochemical cycles

Graeme Nicol [1]

[1] *Université de Lyon, Lyon, France,* [2] *University of Aberdeen, Aberdeen, UK*

Since the mid-1980's, molecular microbial ecology has given us the ability to detect and characterise microbial communities without prior cultivation, and allowed us to begin to comprehend the vast diversity that exists in the environment. Perhaps the most startling discovery of this 'molecular revolution' was the realisation that organisms placed within the domain Archaea did not simply represent a rather strange collection of enigmatic organisms living at the environmental extremes of the natural world. Today, we recognise that they represent a major proportion of microbial biomass on the planet, and make both essential and unique contributions to the planet's carbon and nitrogen biogeochemical cycles. While there have been advances in culturing representatives of the many lineages that were first detected through 16S rRNA-based surveys, there remains a large number of archaeal phyla that have yet to be grown and characterised in the laboratory. However, the use of genomic and metagenomic-approaches have allowed us to gain insights into the potential ecological contributions of these abundant lineages.

This presentation will provide an overview of the most recent advances in our understanding of the diversity of Archaea and their contribution to biogeochemical cycles through both cultivation and genomic-based approaches, and will describe in detail their singularly unique contributions to the nitrogen cycle in both aquatic and terrestrial environments.

28.2

New approaches to rooting the archaeal radiation

Tom A. Williams, Sarah E. Heaps, Svetlana Cherlin, Tom M. W. Nye, Richard J. Boys, T. Martin Embley
Newcastle University, Newcastle upon Tyne, UK

The Archaea play a key role in some of the most important debates about the evolution of cellular life, but their origins and early diversification remain unclear. A rooted archaeal phylogeny would enable us to address some of the most fundamental questions, including the nature of the ancestral archaeon and the branching order of the major groups. It would also help to constrain scenarios for the original divergence of the Archaea and Bacteria, which some recent analyses suggest may represent the oldest extant cellular lineages. However, traditional outgroup rooting approaches are problematic for Archaea because the long branch leading to the Bacteria may induce phylogenetic artifacts, and published inter-domain trees disagree as to the placement of the archaeal root. To overcome these limitations, we have developed new non-reversible models and branch-heterogeneous substitution models in which the root of the tree can be inferred as an integral part of the analysis, without the need for outgroups. We apply these methods to sequence alignments encompassing the full breadth of archaeal diversity and compare our results to complementary analyses using protein concatenations and supertrees. Our results place the archaeal root between a clade comprising the Euryarchaeota and the recently-described “DPANN” lineages on the one hand, and the TACK superphylum on the other, but do not robustly exclude a root position within a clade of early-diverging euryarchaeotal methanogens. These analyses provide new insight into the root of the archaeal tree and demonstrate the utility of non-reversible and branch-heterogeneous models for rooting major cellular radiations.

28.3

Evolutionary history of adaptation to environmental temperature in Archaea

Manolo Gouy
CNRS, Lyon, France

Archaea vary remarkably in their optimal growth temperatures, containing many very diverse lineages adapted to the full spectrum of temperatures compatible with life as we know it. Several molecular features of the archaeal genomes and proteomes are shaped by adaptation to environmental temperature. Conversely, extant archaeal genomes and proteomes contain traces of temperatures at which their ancestors probably lived. I will describe the methods we employ to recover information about these ancestral states, and the resulting history of adaptation to environmental temperatures within the archaeal domain. This history states that the last common ancestor of the archaeal domain and that of the bacterial domain were both thermophilic organisms. In contrast, LUCA, their last common ancestor appeared to have been mesophilic. I will describe the nature of the signal within extant genomes and proteomes that informs us about this non parsimonious feature of the history of optimal growth temperatures.

28.4

Phylogenomic rooting of the domain Archaea: taxonomic and evolutionary implications

David Moreira, Celine Petitjean, Philippe Deschamps, Purificacion Lopez-Garcia
CNRS - Universite Paris-Sud, Orsay, France

The first 16S rRNA-based phylogenies of the Archaea showed a deep division between two groups, the kingdoms Euryarchaeota and Crenarchaeota. However, this classification has been challenged by the recent discovery of many deep-branching lineages (Thaumarchaeota, Aigarchaeota, Nanoarchaeota, Korarchaeota, Parvarchaeota, Aenigmarchaeota, Diapherotrites, and Nanohaloarchaeota) which have also been given the same taxonomic status of kingdoms. Nevertheless, the phylogenetic position of some of these lineages remains controversial. Moreover, as phylogenetic analyses of the Archaea have mostly been carried out without outgroup sequences, it is difficult to determine if these taxa actually define lineages at the same level as the Euryarchaeota and Crenarchaeota. We have addressed the question of the position of the root of the Archaea by reconstructing archaeal trees rooted by using bacterial sequences as outgroup. These trees were based on classical protein markers (32 ribosomal proteins) as well as on 38 new conserved markers identified through phylogenomic analysis. We thus gathered a total of 70 conserved proteins that we analyzed individually and in concatenated datasets. In contrast with previous analyses, our trees consistently placed the root of the Archaea between the Euryarchaeota (including the Nanoarchaeota and other fast-evolving lineages) and the rest of archaeal species. This has important consequences regarding the taxonomic status of many of the newly discovered archaeal groups as well as on the inference of the characteristics of the last archaeal common ancestor, which most likely was a hyperthermophilic complex organism with a gene-rich genome.

28.5

"Lokiarchaeum, a novel deep-sea archaeon illuminates the prokaryote to eukaryote transition"

Anja Spang [1], Jimmy Saw[1], Steffen Jørgensen[2], Katarzyna Zaremba-Niedzwiedzka[1], Joran Martijn[1], Anders Lind[1], Roel van Eijk[1,4], Christa Schleper[2,3], Lionel Guy[1,5], Thijs Ettema[1]
[1] Department of Cell- and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden, [2] Department of Biology, Centre for Geobiology, University of Bergen, Bergen, Norway, [3] Division of Archaea Biology and Ecogenomics, Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria, [4] Groningen Institute for Evolutionary Lifes Sciences, University of Groningen, Groningen, The Netherlands, [5] Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

The origin of the eukaryotic cell represents an enigmatic evolutionary puzzle. Yet, recent studies have started to provide growing support for the emergence of the eukaryotic host cell from within the archaeal domain of life. Thus far however, the identity and nature of the alleged archaeal ancestor of eukaryotes remains obscure, as are the events that triggered the increase of cellular and genomic complexity that is characteristic for eukaryotes.

Applying a metagenomics sequencing approach to deep-sea sediment samples derived from Loki`s Castle, allowed the obtainment of genomes from uncultivated deep-branching members of the archaeal TACK superphylum such as the 'Deep Sea Archaeal Group'. Comparative analyses and phylogenomics of these novel archaeal genomes suggest that they comprise a novel candidate phylum, the ‘Lokiarchaeota’ and constitute a monophyletic group with eukaryotes. Most surprisingly, however, was the finding that Lokiarchaeota encode a distinct repertoire of eukaryotic signature proteins including for instance actins, Ras/Rab-type GTPases and components for all subunits of the eukaryotic ESCRT-modifier system.

These findings are suggestive of sophisticated membrane remodeling capabilities in this novel archaeal phylum, provide strong support for a scenario in which the eukaryotic host evolved from a *bona fide* archaeon, and suggest that the genetic basis for several components that govern eukaryotic cellular complexity has evolved in the archaeal ancestor of eukaryotes.

178A

Laboratory resurrected ancestral extremozymes: An efficient approach to describe protein evolution

Samuel Blanquart[1], Mathieu Groussin[2], Eric Girard[3], Bruno Franzetti[3], Manolo Gouy[2], Dominique Madern [3]

[1] INRIA Lille Nord Europe, Lille, France, [2]) Laboratoire de Biométrie et Biologie Évolutive, Lyon, France, [3] Institut de Biologie Structurale, Grenoble, France

At any time in the cytoplasm, a given protein exists as a population of conformers, each *(i)* differing slightly from others by its structure, *(ii)* being characterized by a given free energy, *(iii*) and being statistically represented in a specific proportion. Environmental parameters that modify the thermodynamic equilibrium of this population will therefore impact protein properties. Shape, energy and proportion of conformers are determined at the level of the amino-acid sequence, upon which acts the mutation process. Each mutation accumulated during the adaptive pathway of a protein facing extreme physico chemical conditions will change the protein energy landscape, the rates of inter-conversion between conformers, and the population structure. Experimental data that probe evolvability potential of consecutive amino acid replacement, on the protein conformational landscape, are scare. To precisely understand how substitutions have driven changes in properties, the evolutionary pathway of malate dehydrogenase (MalDH) in the Archaea was reconstructed. Both the evolutionary trees and the ancestral sequences were inferred by using probabilistic models of molecular evolution. The careful inferences account for frequent reconstruction artefacts, the potential lateral gene transfers and paralogous duplications, and the occurence of insertions and deletions during the evolution of sequences. We focused our study on the main clades that reflect the evolutionary divergence within the extreme halophilic *Halobacteriaceae*. Nine ancestral and four contemporary MalDH gene were then synthesised, overexpressed and purified. Properties of these laboratory resurrected enzymes highlight how the trades-offs between biophysical, biochemical, structural and dynamical parameters drive evolution of protein facing extreme physico chemical conditions.

179B

Ancient and rapid diversification of Thaumarchaeota correlates with phenotypic adaptation

Cecile Gubry-Rangin [1], Christina Kratsch[3], Tom Williams[2], Alice McHardy[3], Martin Embley[2], James Prosser[1], Daniel Macqueen[1]

[1] University of Aberdeen, Aberdeen, UK, [2] University of Newcastle, Newcastle, UK, [3] University of Dusseldorf, Dusseldorf, Germany

Thaumarchaeota form an ammonia-oxidizing archaeal phylum that is abundant in many ecosystems, including the soil, performing a key function in the global nitrogen cycle. Previous high throughput sequencing analysis of the ammonia monooxygenase gene *amoA* has demonstrated that pH is the major driver of thaumarchaeal niche specialization and community structure of these organisms.

While many studies have examined the adaptive distribution and ecophysiology of extant Thaumarchaeota, the evolutionary rise of these prokaryotes to a position of ecological dominance in many habitats has never been considered. Therefore, we characterized thaumarchaeal diversification with respect to ancestral reconstructions of soil pH adaptation, employing state-of-the-art comparative phylogenetic methods based on extensive *amoA* and 16S rRNA sequence data. Our analysis shows a striking increase in lineage diversification rates during early thaumarchaeal evolution that was coupled to major pH adaptation events. While radiation of eukaryotes usually involves an explosion in diversity followed by a decrease in diversification rate, the high initial rate of diversification of Thaumarchaeota remained globally stable during the last 400-700 Ma, resulting in a high level of thaumarchaeal diversity nowadays.

Overall, this study highlights the suprising pattern of thaumarchaeal diversification together with the important role played by pH specialization in thaumarchaeal evolution and is the first of its kind to link diversification with phenotypic adaptation in a prokaryotic phylum. This study provides a framework for comparing dynamics of evolutionary processes across the tree of life and to better understand the diversification processes and past molecular innovations in archaea.

180C

Evolutionary impact of massive gene acquisitions by lateral gene transfer in Methanosarcinales

Nabor Lozada-Chávez, Filipa L. Sousa, Shijulal Nelson-Sathi, William F. Martin
Institute of Molecular Evolution, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

Prokaryotes are well recognized as the most metabolic diverse and old organisms on Earth. Their evolutionary success and metabolic plasticity to adapt to new environments has been mainly attributed to *lateral gene transfer* (LGT) (Deppenmeier et al., 2002, Nelson-Sathi et al., 2012). In order to understand with more detail the contribution of LGT during the evolution of archaeal organisms, we focused this work in one of the most metabolic versatile methanogenic groups of this domain, the Methanosarcinales.

By using standard bioinformatics and phylogenomics methods (Nelson-Sathi et al., 2015), we detected the presence of a massive lateral acquisition of 1,045 genes in 10 species of Methanosarcinales from Bacteria. The distribution of these genes in Methanosarcinales species shows that some genes are present in specific genus and genera, which can reflect either their secondary loss or differential acquisition after the divergence of the last common ancestor of Methanosarcinales. The reconstruction of ancestral metabolic pathways for these genes shows that some play specific roles in particular metabolic pathways, as for instance, enzymes involved in aceticlastic and methylotrophic methanogenesis as well as nitrogen fixation, while others are involved in more general processes (F-ATPase, Rnf complex, ABC transporter complexes). Though further analysis of the reconstructed metabolic pathways is necessary, horizontal gene transfer has clearly affected the evolution of alternative metabolic pathways in Methanosarcinales.

181D

Novel single-cell genomes expand the archaeal tree of life

Jimmy H. Saw [1], Anja Spang[1], Anders E. Lind[1], Guillaume Reboul[2], Dan R. Colman[3], Cristina Takacs-Vesbach[3], Thijs J. G. Ettema[1]

[1] Uppsala University, Uppsala, Sweden, [2] University of Rouen, Rouen, France, [3] University of New Mexico, Albuquerque, NM, USA

Although prokaryotes are known to be dominant members in various biospheres, less than one percent of them present on Earth are cultivated. The rest remain poorly characterized. However, with advancements in sequencing technologies, previously uncultivated microbes can be identified from environments and through sequencing their genomes, their metabolic potential can be predicted, and their evolutionary connections can be pinpointed. Among the uncultivated prokaryotes, Archaea are the least represented in terms of number of cultured species and sequenced genomes. Moreover, the taxonomic classification for the Archaea needs major revisions as new phylum-level lineages are discovered in recent years.

Here, we present the three archaeal lineages isolated from a hot spring sediment sample in Yellowstone National Park, USA. A Single-Cell Genomics (SCG) approach was applied to recover draft genomes of these three lineages and metagenomic sequencing helped to improve assemblies. The draft genome sizes for the Single-Cell Amplified Genomes (SAGs) A3, F10, and N21 are 0.8, 0.7, and 1.2 Mbp, respectively, with genome completeness estimates of 38%, 50%, and 63%. Based on a maximum likelihood phylogenomics tree constructed using a set of 36 conserved single-copy marker genes, SAGs A3 and F10 were placed together with an archaeal SAG belonging to Miscellaneous Crenarchaeotal Group (MCG) from a study by Lloyd et al (2013), and N21 (belonging to SAGMEG group) was placed at the base of Euryarchaeota. Analysis of the gene content of the MCG SAGs revealed several eukaryotic signature proteins (ESPs) thus expanding the already increasing number of ESPs within the domain Archaea.

182A

Deciphering the evolutionary origin of Halobacteriales

Monique Aouad, Najwa Taib, Anne Oudart, Céline Brochier-Armanet
Université de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France, France

The recent advances in sequencing technologies have opened the way to large scale genome sequencing projects and provided a valuable material to study the evolutionary history of Life. In fact, this rainfall of genomic data offers a unique opportunity to improve phylogenetic inferences, in particular by allowing the simultaneous analysis of hundreds of markers and thus to combine the weak phylogenetic signal carried by each individual marker toward a stronger signal. In the case of Archaea, the analysis of proteins involved in transcription, translation and replication, and more recently of 200 additional proteins widely distributed and strongly conserved among the archaeal orders have allowed disentangling the evolutionary history of this Domain and deciphering a number of important relationships such as those among methanogens Class I as well as the proposal of two novel super-classes (the ‘Diaforarchaea' and the ‘Methamonada').
While significant advances emerged from these analyses, some important relationships remain unresolved, such as those among methanogens Class II and Halobacteriales. Answering this question is however crucial in order to understand the evolutionary paths that have led to the emergence of the most extreme halophilic present-day organisms from methanogenic ancestors. Here we present an in-depth phylogenomic analysis aiming at tackling this issue. Using comparative genomics approaches, we identified more than 300 protein markers carrying a reliable phylogenetic signal to study the relationships among these lineages. Accurate analyses of these markers combined with noise-removing approaches shed new light on the origin of Halobacteriales.

183B

Lokiarchaeaota from deep marine sediments shed new light on eukaryogenesis

Katarzyna Zaremba-Niedzwiedzka, Emmelien Vancaester, Thijs J. G. Ettema
Department of Cell- and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Eukaryogenesis represents a major event in evolution and one that is still rather enigmatic. According to the leading hypotheses, eukaryotes have evolved via an endosymbiosis between a bacterium and a host cell. Whereas the identity of the alphaproteobacterial endosymbiont is widely accepted, the identity and nature of the host cell remains subject of heated debate. During the recent years, support is accumulating for scenarios in which the host emerged from within the Archaea, more specifically from a lineage that is affiliated with the archaeal TACK superphylum. Culture-independent approaches such as metagenomics and single cell genomics have the potential to generate new genomic data of uncultured micro-organisms, including archaeal lineages that are ancestrally related to eukaryotes. Here, we focus on the genomic exploration of low-abundant, deeply-rooting members of the recently discovered candidate phylum Lokiarchaeota through ultra-deep sequencing of a deep marine sediment metagenome and by using a variety of binning strategies.

184C

Genome assembly of hot ammonia-oxidizing archaea: a glimpse into the evolutionary success of Thaumarchaeota.

Sophie Abby, Romana Bittner, Christa Schleper
Archaea Biology and Ecogenomics Division, Dep. of Ecogenomics and Systems Biology, University of Vienna, Wien, Austria

Over the past 10 years, Thaumarchaeota have been shown to range among the dominant microbial lineages in various ecosystems. They are found abundantly in the ocean's plankton, but also in deep-sea marine sediments, estuarines and terrestrial environments. Their abundance and their capability to oxidize ammonia indicate that they are major players in the global nitrogen cycle. Deeply branching clades of Thaumarchaeota seem to be represented by species from hot environments, suggesting a thermophilic ancestor for the phylum.

We have recently succeeded to enrich ammonia-oxidizing archaea (AOA) from a hot spring in southern Italy (Ischia) that grow optimally at around 74°C and produce nitrite from ammonia. They are closely related to *Ca. N. yellowstonii*, another thermophilic AOA, and together form a clade with a basal phylogenetic position in the Thaumarchaeota. We are currently in the process of assembling the genome of one of these thermophilic organisms from metagenomic data of the enrichment culture, which will allow comparative analyses with other mesophilic AOA and thermophilic archaea. This could provide key information on the process of adaptation to mesophily and niche diversification in Thaumarchaeota, and help to decipher the origins of ammonia-oxidizing metabolism in archaea. First genomic and comparative genomic studies of thermophilic AOA will be presented.

185D

Origin and evolution of the haem-copper oxidases superfamily in Archaea

Anne Oudart [1], Simonetta Gribaldo[2], Céline Brochier-Armanet[1]

[1] Université de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France, [2] Département de Microbiologie, Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Paris, France

Understanding the origin and evolution of dioxygen reductases, the terminal electron acceptors of aerobic respiratory chains, can provide precious information on the emergence of this very important bioenergetic conversion process. The haem-copper oxidases superfamily (HCO) includes three families of dioxygen reductases (A-HCO, B-HCO, and C-HCO) and a family of NO reductases. Phylogenetic analyses of the HCO superfamily showed that these enzymes have very different evolutionary histories. In particular, A-HCO was proposed to be the most ancient and to have emerged likely prior to the diversification of Archaea and Bacteria, thus before the emergence of Cyanobacteria and oxygenic photosynthesis, while B-HCO are of archaeal origin. The debate on the origin of these enzymes was recently revivified by a structural-based analysis that suggests that A-HCO are more recent members of the HCO superfamily.
The recent outburst of genome sequencing projects, in particular for archaea, offers a unique opportunity to tackle this issue. Here we present an in-depth phylogenomic analysis of the HCO superfamily with a special focus on Archaea. Our results confirm that A-HCO are very ancient, and were likely already present in the Last Universal Common Ancestor. Furthermore, the analysis of archaeal A-HCO and B-HCO regulatory and accessory proteins revealed a complex evolutionary history involving several specific duplications, as well as horizontal gene transfers from archaea to bacteria but also among archaea. Implications for the origin and the evolution of aerobic respiration are discussed.

186A

Intein Invasion in Haloarchaeal Communities

Shannon Soucy, J Peter Gogarten
University of Connecticut, Storrs, Connecticut, USA

We survey metagenomes from Deep Lake, a hyper saline lake in Antarctica, for the presence of inteins, a type of mobile genetic element. In this cold environment replication is limited, and thus evolution occurs mainly through genetic exchange[1]. Inteins have a unique mechanism of propagating themselves in that they associate with a homing endonuclease (HEN), which allows them to home into homologous but un-invaded alleles. Interestingly inteins and HENs have no known mechanism for penetrating a cell wall, thus **intein distribution is dependent on established mechanisms of gene flow within a given population**. Previously we have used the distribution and phylogeny of inteins to reconstruct pathways of gene flow between both closely related and distantly related organisms[2–4]. In Deep Lake where rates of gene exchange are high, and substitution rates are low most inteins are maintained at rates close to fixation, this is in contrast to other natural populations such as the Mediterranean coastline (unpublished data), where inteins exist at much lower frequencies.

POSTER INDEX

Symposia
Posters 1 – 28: Evolution of molecular pathways and networks: Molecular evolution meets systems biology
Posters 29 – 56: Exploring the consequences of ancient and contemporary gene flow
Posters 57 – 68: Inferring fitness landscapes from experimental evolution
Posters 69 – 87: Micro-evo-devo: using natural variation to explain the how and why of phenotypic evolution
Posters 88 – 116: PopGen in space! Theory and inference in spatial population genetics
Posters 117 – 159: Population genomics of rapid adaptation
Posters 160 – 177: Reframing the demography vs. selection debate using 21st century models and data
Posters 178 – 186: The golden age of Archaea: unveiling the diversity and evolution of the third domain of life
Posters 187 – 193: The origins of multicellularity under the light of functional genomics
Posters 194 – 226: Untangling information, noise, and phylogenetic reconstruction in genome scale data
Posters 227 – 236: Novel functional approaches to evolutionary genomics
Posters 237 – 293: Origins and evolution of molecular innovation
Posters 294 – 297: Inference of demography and selection under violations of the Kingman coalescent assumptions
Posters 298 – 327: Ancient genomes: A time machine for investigating natural selection
Posters 328 – 348: Beyond the equilibrium paradigm: The role of temporal processes in population genetics and evolution:
Posters 349 – 361: Genomic and epigenomic evolution of sex chromosomes: Broad patterns and intriguing cases
Posters 362 – 374: Genomics of sex bias: Addressing questions with or without genomes
Posters 375 – 414: Speciation genomics
Posters 415 – 440: Within- and between-host viral evolution
Posters 441 – 475: Evolution and ecology of microbial communities
Posters 476 – 494: The biological impact of transposable elements
Posters 495 – 504: Cancer as a Darwinian process
Posters 505 – 517: Fungal evolutionary genomics: Unravelling mysteries from the forgotten kingdom
Posters 518 – 636: Open symposium
Posters 637 – 650: The evolution of alternative splicing
Posters 651 – 718: Adaptive and non-adaptive evolution of gene expression and regulation
Posters 719 – 725: Short tandem repeats in the post-genomics age: Accurate typing, variability, evolution, and function
Posters 726 – 747: The horizontal component of microbial evolution

SPEAKER AND AUTHOR INDEX

The number following the contributor's name indicates the talk or poster abstract/board number. A, B, C, or D following the abstract number refers to the respective poster session.

Abadie, Catherine: 152D
Abascal, Federico: 15.2
Abby, Sophie: 184C
Abe, Hideaki: 24.5
Abramyan, John: 669C
Abril, Josep F.: 267C
Acevedo-Nieves, Jonathan: 356D
Achaz, Guillaume: 297D
Acuna-Alonzo, Victor: 11.13
Acutis, Pier Luigi: 436B
Adam, Hélène: 568C
Adams, Alexandra: 136D
Adams, Alexandra F.: 307B
Adhikari, Kaustubh: 11.13
Admon, Arie: 549D
Adolfsson, Sofia: 349A
Adzhubey, Ivan: 2.13
Aeschbacher, Simon: 23.2, 25.4
Afanador, Yashira M.: 724B
Agorreta, Ainhoa: 393C
Agosto-Perez, Francisco: 92A
Ågren, J. Arvid: 10.13
Ågren, Jon: 86B
Aguadé, Montserrat: 578A
Aguiar, Bruno: 271C, 272D
Aguilar-Rodríguez, José: 22.2
Aguileta, Gabriela: 505A, 506B
Ahn, Do Hwan: 233C
Aira, Manuel: 463C
Ait-hamlat, Adel: 15.4
Ajay, Jerry: 304C
Akanni, Wasiu: 210A
Akbari, Ali: 6.2
Akey, Joshua: 16.1, 16.3, 22.11, 22.23, 39C
Akgül, Gülfirde: 324C
Akhunov, Eduard: 122B
Akiyama, Tokiho: 341B
Akoz, Gokce: 406D
Aktipis, Athena: 14.4
Alain, Froment: F.4
Alawad, Mohammed: 237A
Albert, Jan: 18.4
Albrechtsen, Anders: 5.6, 305D, 377C, 533D
Albritton, Sarah: 17.2
Albà, M. Mar: 7.1, 246B, 269A, 268D
Alcala, Nicolas: 89B
Aldehoff, Florian: 469A
Alexander, Helen: 598A
Alexander, William G.: 21.9
Alexis, Maria: 15.1
Alföldi, Jessica: 394D
Alhaija, Rana Abu: 630A
Allaby, Robin: 299B
Allen, Richard: 30B
Alleyne, Angela: 552C
Allienne, Jean Francois: 26.3
Allon, Samuel: 643C
Alm, Eric: 448D
Almeida, Fábio: 589D
Almeida, Vasco: 82B
Almudi, Isabel: 22.10
Alon, Uri: 20.14
Alonso, Santos: 682D
Alonso-Blanco, Carlos: 11.19
Alston, Meggan: 344A
Alt, Kurt: 303B
Altenberg, Lee: 7.10
Altizer, Sonia: 90C
Altmüller, Janine: 378D
Amand, Paul St: 122B
Amish, Stephen: 16.4
Amorim, Carlos Eduardo G.: 5.8
Amoutzias, Grigorios: 19C
Amselem, Joelle: 512D
Ananda, Guruprasad: 24.4
Andam, Cheryl: 8.4
Anderson, Dave: F.8
Andersson, Siv G E: 451C, 457A
Andolfatto, Peter: 395A
Andolfi, Gennaro: 707A
Andreatta, Matt: 7.4
Andrews, Simon: 525D
Andrews, Sophie: 418D
Andrés, Aida M.: 16.2, 23.8, 25.3, 177B, 362A
Ane, Cecile: 375A
Anfora, Gianfranco: 551B
Anjos, António dos: 17.10
Anselm, Cherysa: 475C
Anthony, David: 303B
Antia, Rustom: 424B
Antonini, Dario: 707A
Aoki, Kenichi: 556C
Aouad, Monique: 182A
Appleton, Janine: 559B
Aquadro, Charles F.: 146B, 531B
Arachchilage, Madara Hetti: 426D
Arakawa, Nami: 700B
Araki, Hitoshi: 580C
Aravind, L: 518A
Arbeithuber, Barbara: 24.3
Arbiza, Leonardo: 22.1
Arce, Agustin: 666D
Arguello, J. Roman: 6.7, 22.17
Ari, Eszter: 121A
Arias–Rodriguez, Lenin: 10.11
Arif, Saad: 22.10, 22.14
Aristotelous, Andreas: 502D
Armstrong, Philip: 422D
Arnaiz, Olivier: 22.16
Arnaud, Jean-François: 401C
Arndt, Peter F.: 2.16, 214A
Arneodo, Alain: 553D
Arroyo-Pardo, Eduardo: 101B
Arsuaga, Juan Luis: 327B
Artieri, Carlo: 22.21
Arunkumar, Ramesh: 10.13
Arunyawat, Uraiwan: 442B, 577D
Ascencao, Kelly: 26.4
Ashkenazy, Haim: 209D, 576C
Assaf, Zoe June: 5.15
Athanasiadis, Georgios: 43C
Atkinson, Claire: 18.9, 434D
Auboeuf, Didier: 650B
Audit, Benjamin: 553D
Aury, Jean-Marc: 515C
Austerlitz, Frédéric: 5.16, 388B
Austin-Dennehy, Michele: 310A
Ávila-Arcos, Maria: 539B
Avila, Victoria: 26.2
Avram, Oren: 576C, 740C
Axtell, Michael: 687A
Ayala-Agosto, Nikole: 628C
Ayoub, Albert: 230D
Ayub, Qasim: 607B, 707A
Azeredo-Espin, Ana: 697C
Azevedo, Ricardo: 601D
Babarinde, Isaac Adeyemi: 340A, 658D
Babbitt, Courtney: 228B
Babbitt, Gregory: 237A
Babic, Ana: 65A
Babik, Wieslaw: 6.5, 32D, 162C, 171D, 197D
Baca, Mateusz: 320C
Bachmann, Herwig: 4.19
Bachmann, Jörg: 156D
Bachtrog, Doris: Pl.4, 17.3, 358B, 360D, 414D, 648D
Badouin, Helene: 505A
Badouin, Hélène: 506B
Bae, Hye-Ri: 147C
Baek, Jiyeon: F.6
Baer, Sara: 122B
Bafna, Vineet: 6.2
Bagnato, Allessandro: 9.5
Bai, Xue: 134B
Bai, Yalai: 14.2
Bailey, Susan: 13.7, 13.9
Bailly-Bechet, Marc: 200C
Baines, John: 4.2

Bajec, Simona Sušnik: 624C
Bajic, Djordje: 715A
Baker, Antoine: 553D
Baker, Richard: 201D
Baldauf, Sandra: 192B, 193C
Balding, David: 11.16
Baldwin, Ian: 646B
Baldwin-Brown, James G.: L 3C
Balkenhol, Niko: 107D
Ballerini, Evangeline: 410D
Balloux, Francois: 9.1, 18.9, 143C, 151C
Band, Gavin: 621D
Bank, Claudia: 13.3
Banovich, Nick: 5.10
Bapteste, Eric: 465A, 466B, 741D
Baptista, Rodrigo: 667A
Baranov, Pavel: 620C
Barbadilla, Antonio: 74B, 127C
Barbash, Daniel: 725C
Barbujani, Guido: 46B, 106C, 113B
Barghini, Elena: 489B
Barkoulas, Michalis: 12.1
Barmintseva, Anna: 52D
Barnaud, Adeline: 698D
Barnes, Ian: 317D
Barnett, Ross: 30B
Barradas, Ana: 16D
Barreiro, Luis: F.4
Barreto, Mauricio: 37A
Barribeau, Seth: 5.21
Barrie, Assiatu: 291C
Barron, Maite: 127C
Barros, Alberto: 82B
Barros, Nuno: 82B
Barroso-Batista, João: 1.1
Barrón, Maite G.: 482C, 484A
Barshir, Ruth: 501C
Bartha, Denes: 5.1
Bartholomeu, Daniella: 667A
Barton, Nick: 1.11, 9A, 11.18, 11.9, 22.3, 331D
Barzin, Giti: 522A
Basnyat, Buddha: 23.1
Basu, Analabha: 528C
Bataillon, Thomas: 6.1, 8.5, 13.7, 23.3, 175D
Batini, Chiara: 101B, 352D
Battista, Vincent: 6.12
Battistuzzi, Fabia U.: 208C, 719A
Batty, Elizabeth: F.5
Batziakas, Efstratios: 630A
Bauer, Ulrich: 62B
Baumlé, Veronique: 387A
Baums, Iliana: 4.8
Bayes, Mónica: 149A
Bayliss, Sion: 335D, 663A
Bayzid, Md. Shamsuzzoha: 2.1
Bazykin, Georgii A: 157A, 404B
Beall, Bernard: 8.4
Beall, Cynthia: 23.1
Beaulieu, Jeremy: 205D
Becker, Claude: 11.17
Bedddows, Ian: 105B
Bedoya, Gabriel: 11.13
Beekman, Madeleine: 417C, 433C
Beheregaray, Luciano B: 16.4
Behrman, Emily: 1.5
Beier, Claus: 6.1
Beja-Pereira, Albano: 516D
Bekada, Asmahan: 47C
Bekaert, Michael: 5.18
Bekele, Endashaw: 11.16
Bekpen, Cemalettin: 238B, 256D
Bell, Connor: 559B
Bello, Nora: 122B
Belov, Katherine: 536C
Benazzo, Andrea: 23.8, 101B, 106C, 345B
Bendalak, Keren: 549D
Benhamamouch, Soraya: 47C
Benito, Blas: 43C
Benke, Harald: 11.22
Bennett, E. Andrew: 54B, 314A
Benny, Gerald: 21.1
Benoit-Pilven, Clara: 650B
Bento, Pascal: 515C
Benton, Richard: 6.7
Benzerara, Karim: 464D
Berbee, Mary: 21.1
Berendonk, Thomas U.: 563B
Berenguer, Maite Wilke: 9.3
Berenyi, Maria: 720B
Berg, Douglas: 591B
Berg, Ingrid: 328A
Berg, Jeremy: 130B
Berg, Leonard H van den: 317D
Berg, Paul R.: 306A
Berger, Bonnie: 3.6
Bergero, Roberta: 359C, 409C
Berggren, Per: 11.22
Bergland, Alan: 1.5
Bergland, Alan O.: 336A
Bergman, Casey M.: 708B
Bergman, Juraj: 172A
Bergmann, Holger: 744C
Bergstrom, Anders: 6.15
Bergstrom, Carl: 424B
Bergström, Anders: 607B
Bergström, Tomas: 434D
Bergès, Hélène: 85A
Berlin, Sofia: 354B
Bernard, Guillaume: 202A
Bernard, Aurelien: 6.1
Bernardi, Daniel: 655A
Bernardi, Giorgio: 604C
Bernatchez, Louis: 385C
Berry, Sarah: 693C
Bershtein, Shimon: 738A
Berthier, Karine: 116A
Berthouly-Salazar, Cecile: 698D
Berto, Stefano: 22.20
Bertolino, Paola: 4.9, 464D
Bertolucci, Cristiano: 113B, 618A
Bertorelle, Giorgio: 345B
Bertranpetit, Jaume: 6B, 528C
Besnard, Fabrice: 12.1
Besser, Daniel: 7.2
Betancourt, Andrea: 476A, 487D
Betsem, Edouard: F.4
Bezerra, Ana Rita: 149A
Bhadra-Lobo, Siddharth: 585D
Bhandari, Saurabh: 140D
Bhandari, Sushil: 125A
Bhaskar, Anand: 1.8, 139C
Bhattacharyya, Sanchari: 738A
Bhinija, Kisana: 565D
Bideau, Virgil: 552C
Bieysse, Daniel: 152D
Bilge, Arman: 211B
Billaud, Yves: 301D
Birgy, André: 13.6
Bitard-Feildel, Tristan: 7.3, 532C
Bitarello, Bárbara: 177B
Bittner, Noelle K. J.: 293A
Bittner, Romana: 184C
Bjerregaard, Peter: 5.6
Blake, Damer: 456D
Blanc, Guillaume: 18.2
Blancher, Antoine: 595B
Blanquart, Françios: 13.7
Blanquart, Samuel: 178A
Blath, Jochen: 9.3
Blaxter, Mark: 383A
Bleibtreu, Alexandre: 470B
Blevins, William: 269A
Blier, Pierre: 4D
Bligh, Frances: 693C
Blin, Aurelie: 446B
Blin, Camille: 746A
Blondin, Laurence: 723A
Blundell, Jamie: 6.6, 605D
Bocherens, Hervé: 325D
Boddy, Amy: 14.4
Bodova, Katarina: 1.11
Bodyl, Andrzej: 289A
Boehm, Thomas: 518A
Boerger, Christine: 405C
Boessenkool, Sanne: 597D
Bogos, Balazs: 4.16
Bogos, Balázs: 8D, 270B, 573D
Bogumil, David: 23C
Bogusz, Marcin: 619B
Boissier, Jérome: 26.3
Boitard, Simon: 5.16, 9.5
Bollback, Jonathan: 471C
Bolnick, Deborah: 55C, 579B
Bolon, Daniel N.A.: 13.3
Bolívar, Paulina: 221D
Bomhoff, Matthew: 293A
Bonazzi, Marion: 319B

Bondioli, Luca: 318A
Bonew-Rivera, Yásen: 356D
Bonhoeffer, Sebastian: 13.5, 416B
Bonilla-Perez, Jenelle: 356D
Booker, Tom: 586A
Boone, Celia: 705C
Boore, Jeffery L.: 334C
Borenstein, Elhanan: 8.7
Bork, Peer: 4.1, 5.20
Bornberg-Bauer, Erich: 7.3, 64D, 188B, 266B, 532C, 564C, 22.1
Borner, Janus: 264D
Boross, Gábor: 689C
Bortolini, Maria Catira: 11.13
Bortolini, Maria Cátira: 174C
Bos, Kirsten: 25.1, 450B
Bos, Kirsten I.: 318A
Boscari, Elisa: 53A
Bossolini, Eligio: 10.1
Bossu, Christen: 399A
Botha, Anna-Maria: 521D
Botta, Giordano: 10.10
Boubli, Jean: 606A
Boulesteix, Matthieu: 3.15
Bourdier, Frederic: 109B
Bousios, Alexandros: 3.16
Boussau, Bastien: 2.1, 554A
Bouwman, Abigail S.: 324C
Bowden, Rory: F.5
Boyano, María Dolores: 682D
Boyko, Adam: 3.4
Boys, Richard J.: 28.2
Bozdag, Gönensin Ozan: 403A
Bracewell, Ryan: 517A
Bradburd, Gideon: F.1
Bradburn, Sarah: 344A
Bradbury, Ian R.: 306A
Bradley, Daniel G: 317D
Bradman, Neil: 11.16
Branca, Antoine: 21.5, 45A, 506B, 508D
Brand, Cara: 5.11
Brandao, Karina: 655A
Brandao, Marcelo: 655A
Brandstädter, Katharina: 83C
Brandvain, Yaniv: 34B
Branger, Catherine: 8.8
Brasó-Vives, Marina: 245A, 249A
Bratulic, Sinisa: 4.14
Braun, Dylan: 12.5
Bravo, Ignacio G.: 437C
Brayne, Adam: 428B
Brekke, Thomas D: 10.7
Brelsford, Alan: 374A
Brem, Gottfried: 361A
Brenner, Steven: 15.5
Bresadola, Luisa: 391A
Bresolin, Nereo: 315B, 599B, 600C
Breuer, Judith: 18.9, 427A, 434D, 438D
Brisse, Sylvain: 746A
Brochier-Armanet, Céline: 5A, 182A, 185D, 199B, 534A
Brockhurst, Michael: 8.2, 8.9
Brockmüller, Thomas: 646B
Bronstein, Omri: 625D
Brousseau, Louise: 118B
Brown, Anais: 208C
Brown, Amanda C: 18.9
Brown, Julianne: 18.9, 427A, 434D, 438D
Brown, Katherine D.: 693C
Brown, Susan: 122B
Bruchhaus, Iris: 264D
Brun, Christine: 5A
Brunke, Sascha: 1A
Bryant, David: 8.6, 224C
Bubnell, Jaclyn: 531B
Buchfink, Benjamin: 450B
Buckee, Caroline: 9.4
Buffry, Alexandra: 22.19
Buikstra, Jane: 25.1
Burbano, Hernán A.: 11.17, 321D
Burge, Chris: 15.1
Burger, Joachim: 25.6
Burger, Pamela: 300C
Burghgrave, Nicolas: L 6B
Burke, Jason R.: 263C
Burke, Terrence: 572C
Burmester, Thorsten: 264D, 683A
Burnell, Ann: 250B
Burri, Reto: 542A
Burridge, Christopher: 519B
Busa-Fekete, Róbert: 8D
Buschbeck, Elke: 72D
Bustamante, Carlos: 136D, 539B, 541D, 607B, 716B
Bustamante, Carlos D.: 17.9, 307B
Bybee, Seth: 463C
Byrne, Kevin P.: 21.11
Bódi, Zoltán: 149A
Böhm, Alexander: 287C
Chang, Yao-Ming: 686D
Cabec, Adeline Le: 319B
Cabrera, Carla Bello: 283C
Cabrera, Lilia: 591B
Caccone, Adalgisa: 587B
Caffell, Anwen: 317D
Cagliani, Rachele: 315B, 599B, 600C
Cai, Hong: 15C
Cai, Huiqiang: 7.2
Caicedo, Ana: 141A
Cain, John Wesley: 7.5
Calabria, Gemma: 146B
Calafell, Francesc: 110C
Calderoni, Luca: 618A
Callahan, Colin: 10.7
Callebaut, Isabelle: 7.3
Callendrello, Alanna: 8.4
Camargo, Sâmia: 51C, 280D
Camilo, Vânia: 665C
Campagna, Leonardo: 384B
Campos, Jose: 26.2, 127C
Campos, Jose L: 359C
Canelli, Elena: 436B
Canizales-Quinteros, Samuel: 11.13
Cannella, Vincenza: 436B
Capella-Gutíerrez, Salvador: 622A
Capilla, Laia: 394D
Capra, John: 491D
Capra, John A.: 16.3
Capy, Pierre: 478C
Cara, Angeles de: 388B
Carareto, Claudia: 490C
Carberry, Ciara: 452D
Carbonell, Eudald: 327B
Carbonetto, Peter: 69A
Card, Daren: 129A
Cardona-Feliciano, Christian: 356D
Cardoso, Gisele: 697C, 699A
Cardoso-Moreira, Margarida: 6.7, 22.17
Carelli, Francesco Nicola: 7.11
Carey, Lucas: 269A
Carillo, Enrique: 15.2
Carlier, Jean: 152D
Carlini, David: 77A
Carmel, Liran: 3.11
Carmi, Shai: 557D
Carmo, Camila: 51C
Carneiro, Jeferson: 606A
Caro, Angelia: 372C
Carpenter, Emma: 664B
Carpenter, Meredith: 307B
Carr, Robert: 539B
Carretero, José Miguel: 327B
Carreté, Laia: 513A, 622A
Carroll, John: 5.18
Carroll, Maureen: 317D
Carroll, Sean: 265A
Cartwright, Reed: 2.11, 11.5, 36D, 504B, 636C
Cartwright, Reed A.: 6.10
Caruso, Chiara: 52D
Caruso, Claudio: 436B
Carvalho, Ana Sofia: 82B
Carvalho, Filipa: 82B
Carvalho, Tiago: 722D
Casale, Francesco Paolo: 22.15
Casals, Ferran: 528C
Casanova, Adrian Gonzalez: 9.3
Caseys, Céline: 10.1, 391A
Casillas, Sònia: 558A
Casola, Claudio: 486C
Castellano, David: 5.19, 74B, 127C, 176A, 535B
Castellano, Sergi: 362A
Castiglione, Stefano: 5.1

Castilho, Ricardo Arraiano: 516D
Castillo-Morales, Atahualpa: 616C, 706D
Castoe, Todd: 129A
Castric, Vincent: 85A, L 7C
Castro, José María Bermúdez de: 327B
Castro, L. Filipe C.: 277A
Catania, Francesco: 639C, 691A
Cattani, M. Victoria: 5.11
Cattani, Victoria: 560C
Cavalleri, Gianpiero L.: 101B
Cavallini, Andrea: 489B
Ceolin, Andrea: 106C
Cerutti, Francesco: 436B
Chacon-Duque, Juan Camilo: 11.13
Chai, Juan Juan: 205D
Chain, Frédéric: 22.1, L 5A
Chaix, Raphaelle: 109B
Chaliotis, Anargyros: 19C
Challis, Richard J.: 10.5
Chalyvopoulou, Pavlina: 19C
Chan, Cheong Xin: 202A
Chan, Hue Sun: 564C
Chancerel, Emilie: 301D
Chang, Andrew Ying-Fei: 637A
Chang, Hsiao-Han: 9.4
Chang, Jessica: 13.2
Chang, Tean-Hsu: 8.10
Chapman, Tracey: 57A, 371B, 443C
Chapuis, Marie-Pierre: 116A, 723A
Chareonviriyaphap, Theeraphap: 442B, 577D
Charlesworth, Brian: 26.2
Charlesworth, Deborah: 17.1, 359C, 409C, 574A
Charlton-Perkins, Mark: 72D
Charrau, Pauline: 300C
Charrier, Jean-Philippe: 199B
Charrier, Pierre: 23.14
Chateigner, Aurélien: 3.3
Chatzou, Maria: 203B
Chaudhry, Sophia: 719A
Chautard, Emilie: 650B
Chavarria, Claudia: 5.10
Chavez, Andreas: 153A
Chaw, Shu-Miaw: 567B
Chebbi, Amine: 26.9
Chen, Chi-Kuan: 80D
Chen, Chih-Feng: 80D
Chen, Chun-Long: 553D
Chen, Chunyan: 7.9, 276D
Chen, Feng-Chi: 681C
Chen, Feng-chi: 637A
Chen, Hua: 633D
Chen, I-Chen Kimberly: 27.5
Chen, Jiun-Jie: 80D
Chen, Kun: 633D
Chen, Lei: 19.4
Chen, Ling-Ling: 8.10
Chen, Michael: 366A
Chen, Ming-Hui: 2.12
Chen, Minhui: 145A
Chen, Ping: 15.1
Chen, Wei: 7.2
Chen, Wei-Hua: 5.20
Chen, Xiao: 4.13
Chen, Xiaoshu: 702D
Chen, Yi-Ching: 681C
Chen, Zehua: 506B
Cheng, Jade: 43C
Cheng, Jen-Hao: 681C
Cheng, Jie: 378D
Chenoweth, Stephen F.: 368C
Cherlin, Svetlana: 28.2
Cherry, J. Michael: 507C
Cheverud, Jim: 70B
Chiang, Charleson: 25.7
Chiang, Tai-Wei: 681C
Chiapello, Hélène: 512D
Chiaria, Matteo De: 6.15
Chikhi, Lounes: 377C
Chikhi, Rayan: 26.8
Childs, Geoff: 23.1
Childs, Lauren: 9.4
Chirn, Gung-wei: 3.6
Chiva, Cristina: 268D
Cho, Wan-Sup: 453A
Cho, Won Kyong: 461A, 462B
Choi, Chi-Hwan: 453A
Choi, Hoseong: 461A, 462B
Choi, Jae Young: 531B
Chris, Tyler-Smith: 23.9
Christe, Camille: 10.1
Christensen, David: 8.1
Christiansen, Mette: 18.9, 427A
Christine, Harmant: F.4
Christou, Epaminondas: 630A
Chu, Chong: 477B
Chuang, Trees-Juen: 637A, 681C
Chuong, Cheng-Ming: 80D
Churches, Nathan: L 8D
Cibrian, Angelica: 298A
Ciccozzi, Massimo: 436B
Ciomborowska, Joanna: 282B, 488A
Cissé, Mame Codou Guèye: 698D
Clark, Andrew: 5.7, 17.5, 20.16, 137A, 725C
Clark, Andrew G.: 6.7, 22.17
Clark, Duncan A: 18.9
Clark, Nathan: 20.9
Clark, Richard: 22.15
Clarke, Andrew: 299B
Clarke, Rachel: 11.12
Clavereau, Isabelle: 478C
Clement, Yves: 709C
Clerici, Mario: 315B, 599B, 600C
Cleves, Phillip: F.6
Clucas, Gemma: 173B
Coelho, Rui: 51C, 280D
Coen, Enrico: 11.18
Coimbra, José: 17.10
Colbourne, John K: 231A
Coleman, Rhys: 16.4
Collins, Matthew J: 317D
Colman, Dan R.: 181D
Colonna, Vincenza: 707A
Colston, Sophie M.: 16.5
Colyn, Marc: 88A
Comai, Luca: 10.12
Comas, David: 47C, 49A, 110C, 174C, 722D
Comas, Iñaki: 745D
Conard, Nicholas J.: 325D
Condori, Edgar: 90C
Congiu, Leonardo: 52D, 53A
Consortium, Ag1000g: 10.10
Consortium, CANDELA: 11.13
Consortium, iGénolevures: 21.3
Consortium, NESCent: 472D
Consortium, The Brazilian EPIGEN: 37A
Consortium, The Vervet: 161B
Consortium, The 1000 bull genomes project: 9.5
Cook, Tiffany: 72D
Cooley, Arielle: 661C
Coop, Graham: F.1, 6.8, 17.4, 23.2, 25.4, 34B, 130B
Cooper, Alan: 11.12, 303B
Cooper, Elizabeth: 410D
Cooper, Max: 518A
Copp, Gordon: 380B
Corbett-Detig, Russ: 275C
Corcoran, Pádraic: 166C
Cordaux, Richard: 3.3, 26.9
Cordellier, Mathilde: 673C
Cordes, Matt: 7.4
Cordes, Matthew: 4.11
Corel, Eduardo: 741D
Cornejo, Omar: 716B
Cornetti, Luca: 345B
Cornille, Amandine: 45A
Coronado, Marta: 74B, 127C, 535B
Corradi, Nicolas: 21.1
Cortez, Diego: 26.4
Cosseau, Céline: 26.3
Coster, Stefanie De: 588C
Couce, Alejandro: 13.6
Couderc, Marie: 698D
Couloux, Arnaud: 515C
Coupland, George: 86B
Courtiol, Alexandre: 41A
Craig, Sienna: 23.1
Crawford, Gregory: 228B
Crawford, Jacob: 627B
Crawford, Jacob E.: 10.9
Crawford, Michael H.: 55C
Creanza, Nicole: 170C

Crease, Teresa: 485B
Creevey, Christopher: 210A, 261A, 452D
Cristescu, Melania E.: L 5A
Croll, Daniel: 16.10, 21.6
Crollius, Hugues Roest: 229C, 709C
Cruaud, Astrid: 116A
Cryer, Nicolai: 6.1
Crystal, Ronald G.: 92A
Cseke, Botond: 66B
Csörgő, Bálint: 4.16, 8D, 20.2, 573D
Cui, Chaoying: 125A
Cunha, Ana: 272D
Cuomo, Christina A.: 505A, 506B
Currat, Mathias: 11.10, 11.11
Cury, Jean: 8.11
Cvijovic, Ivana: 1.14
Czarkwiani, Anna: 12D
Cáceres, Mario: 5.19, 176A, 558A
d'Aubenton-Carafa, Yves: 553D
Dabad, Marc: 581D
Dagan, Tal: 23C, 83C, 155C, 224C, 467C, 727B, 742A
Dai, Fangyin: 124D
Dai, Kael: 6.10
Daish, Tasman: F.5
Dal-Kilinc, Gulsah Merve: 692B
Daligault, Julien: 323B
Dall'Ara, Irene: 113B
Dall'Olio, Giovanni: 6B
Dalmay, Tamas: 57A
Dalquen, Daniel: 31C
Daly, Toni: 2.14
Dalén, Love: 313D
Damião, Isabel: 82B
Danchin, Etienne: 250B
Danecek, Petr: 23.9
Dang, Guoquan: 125A
Daniels, Rod: F.7
Danko, Charles: 19.3
Dannemann, Michael: 16.2, 25.8
Dannenberg, Richard: 4.8
Daraba, Andreea: 149A
Darzentas, Nikos: 3.16
Das, Sabyasachi: 518A
Dash, Phil: 14.5
Daskalaki, Evangelia: 327B
Daub, Josephine: 3C
Daubin, Vincent: 8.5, 737D
Davalos, Liliana: 379A
Davies, Richard: 91D
Davis, Simon: 326A
Davydov, Iakov: 198A
De-Biasi, Juliana: 51C
De-Franco, Bruno: 51C
DeGiorgio, Michael: 25.5, 42B, 128D
DeGregori, James: 501C
DePamphilis, Claude W.: 5.17
Dearlove, Bethany: 428B
Debarre, Florence: 11.8
Debuchy, Robert: 21.5
Decker, Jared: 9.5
Deereper, Alexis: 454B
Defoort, Jonas: 13A, 274B
Defossez, Pierre-Antoine: 314A
Degnan, Bernard: 27.4
Degnan, James: 2.3
Degnan, Sandie: 7.6
Dekker, Teun: 551B
Delaporte, Cécile: 297D
Delser, Pierpaolo Maisano: 101B, 352D
Demengeot, Jocelyne: 1.1
Dennenmoser, Stefan: 3.13
Denny, Joshua: 16.3
Depledge, Daniel: 427A, 434D, 438D
Depledge, Daniel P: 18.9
Derevianko, Anatoly: 309D
Derieg, Nathan: 410D
Derényi, Imre: 503A
Desai, Michael: F.2, 1.14, 13.1, 13.10, 68D, 142B
Desai, Michael M.: 1.16, 449A
Desalle, Rob: 201D
Deschamps, Matthieu: 126B
Deschamps, Philippe: 4.9, 28.4
Deszo, Marina: 699A
Detry, Cleia: 326A
Deun, Katrijn Van: 588C
Devos, Jane: 321D
Dharmayanthi, Anik Budhi: 78B
Dhinaut, Julien: 316C
Diack, Oumar: 698D
Diamant, Arik: 561D
Diaz, Maria Rivera: 27C
Dickins, Benjamin: 342C, 507C
Diez, Concepcion M.: 3.16
Diffloth, Gerard: 109B
Dikow, Rebecca: 2.9
Dillmann, Christine: 672B
Din, Israr Ud: 568C
Ding, Yuantong: 502D
Dion-Côté, Anne-Marie: 385C
Diouf, Diaga: 698D
Djurdjevič, Ida: 624C
Doak, Tom: 257A
Dobrynin, Pavel: 412B, 413C
Doerder, Paul: 601D
Dohmen, Elias: 64D
Dolezal, Marlies: 9.5
Dominguez, Jorge: 463C
Domman, Daryl: 732C
Dong, Jianrong: 2.3
Dong, Tao: 418D
Donnelly, Peter: F.5
Doorn, Sander van: 20.3
Dorp, Lucy van: 11.16
Dosch, Julian: 744C
Douady, Christophe: 23.14
Doucet, Aurelien: 3.4
Douglas, Gavin M.: 10.13
Dowell, Noah: 265A
Drew, Phillip: 22.15
Driller, Christine: 114C, 402D
Drillon, Guénola: 21.2
Drucker, Dorothée G.: 325D
Drummond, Alexei: 5.12, 211B, 423A
Du, Xiaochen: 735B
Duan, J.: 175D
Dubin, Manu: 22.15
Duchemin, Wandrille: 302A
Duchen, Pablo: 537D
Duchene, Sebastian: 2.7
Dudas, Gytis: F.7, 18.3, 421C
Dudek, Katarzyna: 162C, 171D, 197D
Dudycha, Jeffry L.: 260D
Dufkova, Petra: 613D
Dufresne, France: 4D
Dujon, Bernard 21.10
Dukić, Marinela: 94C
Dumas, Emile: 21.5
Dumas, Emilie: 508D
Dunham, Maitreya: 3.2
Dunn, Barbara: 13.2, 507C
Dupont, Joëlle: 508D
Duprat, Simon: 515C
Dupuy, Berit Myhre: 101B
Durbin, Richard: 10.5, 11.12, 607B
Duret, Laurent: 8.5, 22.16, 23.14
Durrett, Rick: 502D
Durvasula, Arun: 585D
Dutheil, Julien Y.: 10.6, 21.8
Duval, David: 26.3
Duveau, Fabien: 22.22
Dyken, J David Van: 449A
Dykstra, Emily: 4.11
Dyson, Gregory: 137A
Dávalos, Liliana: 548C
Eagle, Shannon: 485B
Early, Angela: 20.16
Ebersberger, Ingo: 514B, 744C
Ebert, Dieter: 21.7, 94C
Ebranati, Erika: 436B
Eckert, Kristin: 24.4
Eckhart, Leopold: 251C
Edmands, Suzanne: 353A
Ehlers, Bodil: 6.1
Ehrman, Lee: 4.7
Eijk, Roel van: 28.5
Eimes, John: 569D
Einer-Jensen, Katja: 18.9
Eizaguirre, Christophe: 22.1
Eldholm, Vegard: 151C
Eldon, Bjarki: 9.3
Elias, Marianne: 388B
Elleder, Daniel: 431A
Ellegaard, Kirsten M: 451C
Ellegren, Hans: 221D

Ellis, Peter: 693C
Elmer, Kathryn: 582A
Elodie, Laine: 15.4
Embley, Martin: 179B
Embley, T. Martin: 28.2, 732C
Eme, David: 23.14
Eme, Laura: 729D
Emera, Deena: 4.20, 230D
Emery, Sarah: 11.14
Emond, Stephane: 292D
Enard, Wolfgang: 688B
Engler, Jan O.: 107D
Epert, Aline: 18.2
Ercan, Sevinc: 17.2
Erdem, Babur: 71C
Erickson, Priscilla: F.6
Eriksen, Heidi A.: 101B
Eriksson, Anders: 169B
Erkenswick, Gideon: 543B
Erlich, Yaniv: 24.1
Ermini, Luca: 23.14
Ernenwein, Lise: 26.9
Erpf, Anna-Christina: 22.20
Esbjornsson, Joakim: 415A, 418D
Escalante, Lucio: 490C
Escudero, Jose Antonio: 65A
Esson, Heather: 613D
Esteve, Carla: 146B
Etienne, Patin: F.4
Etingova, Anna: 157A
Ettema, Thijs J.G.: 28.5, 181D, 183B, 473A, 457A
Evans, Thomas: 4.3
Evrard, Olivier: 109B
Ewens, Warren: 614A
Ewing, Gregory: 168A
Excoffier, Laurent: 3C, 11.1, 97B, 167D, 376B
Exposito-Alonso, Moises: 11.17
Eyre-Walker, Adam: 11.20, 61A, 127C
Ezkurdia, Iakes: 15.2
Fabrizius, Andrej: 683A
Fabryova, Helena: 431A
Fadhlaoui-Zid, Karima: 47C, 49A
Fairhead, Cécile: 21.3, 513A
Faivre, Nicolas: 6.1
Fakhro, Khalid A.: 92A
Falkowski, Lisa: 22.20
Fallahshahroudi, Amir: 680B
Faltyskova, Zuzana: 312C
Falush, Daniel: 4.5, 11.3
Fan, Dingding: 19.4
Fan, Weishu: 583B
Fan, Wen-Lang: 80D
Fang, Xiaodong: 19.4
Fang, Yongjun: 474B
Farci, Patrizia: 18.8
Fares, Mario A.: 58B
Farias, Izeni: 606A
Faridi, Nurul: 486C
Farkas, Zoltán: 149A, 703A
Farré, Marta: 394D
Fawcett, Jeffrey: 494C
Faye, Laurélène: 93B
Feder, Alison: 18.7
Fehren-Schmitz, Lars: 636C
Fehér, Zoltán: 115D
Feil, Ed: 335D
Feil, Edward: 663A
Feiner, Nathalie: 481B
Fekete, Gergely: 8D, 20.2
Feldman, Marc: 614A
Feldman, Marcus: 20.12, 136D, 170C, 556C
Feldman, Marcus W.: 336A
Feldstein, Tamar: 561D
Felgontov, Pavel: 596C
Fellner, Lea: 653C
Felsenstein, Joseph: Pl.1
Feng, Guoping: 631B
Feng, Jie: 735B
Feng, Ying: 704B
Ferguson-Smith, Malcolm: 26.8
Ferlin, Minca: 739B
Fernandez-Valverde, Selene: 27.4
Ferrand, Nuno: 544C
Ferrari, Giada: 324C
Ferreira, Zélia: 20.9, 665C
Ferrette, Bruno: 51C
Ferris, Kathleen: 153A
Fescemyer, Howard W.: 26.8
Feschotte, Cedric: 291C
Feuk, Lars: 408B
Feulner, Philine: 10.4, 22.1
Feurtey, Alice: 45A
Feyertag, Felix: 430D
Fickel, Joerns: 41A
Field, David: 11.18
Field, Yair: 22.8
Fields, Peter: 94C
Figueira, Antonio: 655A
Figueroa-Vázquez, Zahudy: 356D
Figuet, Emeric: 308C
Fijarczyk, Anna: 162C, 171D
Filiault, Daniele: 410D
Filippi, Giulia: 599B
Filippo, Cesare de: 23.8, 177B, 362A
Finci, Iris: 15.3
Fischer, Daniel: 9.5
Fischer, Gilles: 21.2
Fisher, Charles: 4.8
Fisher, Daniel: 6.6, 605D
Fiston-Lavier, Anna Sophie: 482C
Flamand, Marie-Christine: 88A
Flandrois, Jean-Pierre: 199B
Flannick, Jason: 631B
Flatscher, Ruth: 132D
Flatt, Thomas: 11.21
Fleming, James: 222A
Florez-Rueda, Ana Marcela: 5.4, 397C
Fluch, Silvia: 720B
Flutre, Timothée: F.4
Flynn, Jullien M.: L 5A
Fobis-Loisy, Isabelle: 85A
Foerster, Daniel: 41A
Fofana, Amadou: 698D
Fogarty, Laurel: 556C
Foll, Matthieu: 6.9, 165B, 167D
Fonseca, Filipa: 82B, 665C
Font, Neus: 110C
Fontanillas, Eric: 505A
Fontes, Fernanda: 602A
Foresti, Fausto: 51C, 280D
Forin, Niccolò: 53A
Forni, Diego: 315B, 599B, 600C
Foster, Charles: 196C
Foster, Charles SP: 2.7
Foulkes, Nicholas S.: 618A
Fournier, Elisabeth: 512D
Fournier, Greg: 726A
Foy, Scott: 7.4, 285A
Fraaije, Bart: 333B
Frajman, Bozo: 132D
Franchini, Paolo: 279C, 657C
Franco, Marco A.P.: 23.3
Francois, Clementine: 23.14
Frangoulis, Constantin: 630A
Franssen, Susanne: 6.14, L 1A
Frantz, Laurent: 6.12, 30B
Franzetti, Bruno: 178A
Fraser, Hunter: 4.10, 22.21
Freel, Kelle: 6.15, 21.10
Freimer, Nelson: 161B
Freitas, Lucas: 627B
Frenkel, Evgeni M: 449A
Fridman, Eyal: 549D
Friedlander, Tamar: 20.14
Friedman, Lilach: 499A
Friedrich, Anne: 6.15, 21.10
Frigato, Elena: 618A
Fritzemeier, Claus Jonathan: 2B, 20.2
Frost, Simon: 428B
Fruchard, Cécile: F.3
Fruciano, Carmelo: 657C
Fu, Chengjie: 192B, 193C
Fu, Qiaomei: 25.3
Fuchs, Nina: 7.2
Fucikova, Karolina: 2.12
Fuenzalida, Gonzalo: 656B
Fujii, Kiyotaka: 341B
Fujimoto, Kazuma: 341B
Fujito, Naoko: 670D
Fulgione, Andrea: 11.19
Fuller, Zachary: 365D
Fullmer, Matthew S.: 16.5
Fumagalli, Luca: 542A
Fumagalli, Matteo: 5.6, 143C
Fungtammasan, Arkarachai: 24.4
Furmanek, Tomasz: 624C

Furstenau, Tara: 11.5
Fuselli, Silvia: 618A
Fusté, Elena Bosch: 343D
Futschik, Andreas: 566A
Félix, Marie-Anne: 12.1
Green, David: 330C
Gabaldón, Toni: 513A, 622A
Gaffney, Stephen: 14.2
Gagat, Przemyslaw: 289A
Gagneux, Sebastien: 25.1
Gagnevin, Lionel: 454B
Gakuhari, Takashi: 594A
Galdeano, Alicia García de: 682D
Galdos, Maria B. Lorente: 581D
Galili, Sheli Ostrow -: 501C
Galliart, Matthew: 122B
Gallina, Sophie: 85A
Gallo, Carla: 11.13
Gallot, Lucie: 18.2
Galtier, Nicolas: 23.3, 308C
Gansauge, Marie-Theres: 309D
Gao, Ziyue: 5.8
Garcia, Sonia: 526A
García, Oscar: 682D
Gardeazábal, Jesús: 682D
Gardner, Kyle: 584C
Garg, Sriram: 255C
Garnier-Gere, Pauline: 160A
Garrett, Karen: 122B
Garud, Nandita: 23.10
Gascuel, Olivier: 2.6
Gasparich, Gail E.: 728C
Gatfield, David: 644D
Gattepaille, Lucie: 337B
Gauffre, Bertrand: 116A
Gaughan, Sarah: 390D
Gaut, Brandon S.: 3.16, 684B
Gauthier, Laurent: 400B
Gavilán, Brenda: 267C
Gavryushkina, Alexandra: 5.12
Gay, Laurène: 316C, 339D
Gayà-Vidal, Magdalena: 5.19, 176A
Gebert, Daniel: 694D
Geigl, Eva-Maria: 54B, 314A, 323B
Geiler-Samerotte, Kerry: 20.7
Gemmell, Neil J.: 24.5, 571B
Geneva, Anthony: 3.1
Genner, Martin J.: 10.5
George, Oliver: 525D
Georgiev, Stoyan: 24.1
Gerbault, Pascale: 5.6, 143C
Gerber, Florian: 4.14
Gerlach, Daniel: 3.6
Gerrard, Dave T.: 708B
Gesemann, Matthias: 248D
Gessain, Antoine: F.4
Gessner, Cornelia: 571B
Geyrhofer, Lukas: 1.15
Ghalichi, Ayshin: 692B
Ghanbarian, Avazeh: 7.2, 22.13
Ghirotto, Silvia: 23.8, 46B, 101B, 106C, 113B
Ghiselli, Fabrizio: 244D
Giammarioli, Monica: 436B
Gibilisco, Lauren: 648D
Gignoux, Christopher: 136D
Gilad, Yoav: 5.10, 22.7
Gilbert, Clément: 3.3, 26.9
Gilbert, M. Thomas P.: 310A
Gilbert, Thomas: 539B
Gilbert, Thomas P.: 35C
Gilchrist, Michael: 205D, 510B
Gillard, Thomas: 417C
Gilligan, Dean: 16.4
Gilman, Robert: 591B
Giner-Delgado, Carla: 5.19, 176A
Ginja, Catarina: 326A
Gioia, Luca De: 599B
Giordani, Tommaso: 489B
Giordano, Christina: 548C
Giorgianni, Matt: 265A
Girard, Eric: 178A
Giraud, Isabelle: 26.9
Giraud, Tatiana: 21.5, 45A, 505A, 506B, 508D
Gittelman, Rachel: 22.11
Gladieux, Pierre: 45A
Glaser-Schmitt, Amanda: 22.9
Glassberg, Emily: 614A
Glemin, S.: 175D
Glemin, Sylvain: 23.3
Glocke, Isabelle: 322A
Gloria-Soria, Andrea: 422D
Gloss, Andrew: 138B
Glémin, Sylvain: 672B
Gneme, Awa: 10.9
Godinho, Raquel: 544C
Goebel, Julien: 542A
Gogarten, Johann Peter: 16.5, 186A
Goillot, Evelyne: 5A
Gojobori, Jun: 154B
Gokcumen, Omer: 304C
Goldberg, Amy: 1.7
Goldman, Nick: 632C
Goldstein, Richard: 10.8, 18.5, 427A
Gomez, Karen: 495A
Goncalves, Paula: 21.9
Gong, Qiang: 498D
Gongora, Jaime: F.5
Gontijo, Alisson: 14B
Gonzalez, Alberto Nakauma: 20.3
Gonzalez, Josefa: 22.12, 482C
Gonzalez-Candelas, Fernando: 745D
Gonzalez-Jose, Rolando: 11.13
Gonzalez-Martinez, Santiago: 5.1
Gonzalez-Quevedo, Catalina: 91D
González, Favio: 583B
González, Josefa: 3.5, 479D, 484A
Good, Benjamin: F.2, 1.16, 1.14
Good, Jeffrey: 10.7, 213D
Good, Robert: 284D
Goodall-Copestake, Will: 344A
Goodman, Aaron: 20.12
Gopal, Shakuntala: 143C
Gopalakrishnan, Shyam: 69A
Gophna, Uri: 8.1
Gorbunova, Vera: 3.1
Gordo, Isabel: 1.1
Gos, Gesseca: 10.13
Gossmann, Toni: 523B
Gottipati, Srikanth: 6.7, 22.17
Goubert, Clément: 3.15
Goubet, Pauline: 85A
Goudet, Jérome: 11.21
Gould, Billie A.: 321D
Gould, Sven: 255C
Gounot, Jean-Sebastien: 21.10
Gout, Jean-François: 235A, 257A
Gouveia, Mateus: 37A
Gouy, Manolo: 28.3, 178A, 199B, 534A
Gouzy, Jérôme: 21.5, 505A, 506B, 508D
Gower, David: 261A
Graf, Joerg: 16.5
Graham, Catherine: 379A
Gramzow, Lydia: 273A
Grandaubert, Jonathan: 21.8
Grande, Francesco Dal: 514B
Grange, Thierry: 54B, 314A, 323B
Granka, Julie: 136D
Grant, Tom: F.5
Granvik, Mikaela: 588C
Grarup, Niels: 5.6
Grassi, Elena: 695A
Grassl, Guntram: 4.2
Grau-Bové, Xavier: 190D
Graves, Jennifer A. Marshall: 17.6
Gray, Douglas: 239C, 649A
Gray, Miranda: 122B
Green, Anna G.: 24D
Green, Eric D.: 23.8
Greenbaum, Gili: 96A
Greenwood, Jenny: 532C
Gregis, Veronica: 568C
Gregoire, Laura: 3.9
Greig, Duncan: 403A
Greminger, Maja: 722D
Grenier, Jennifer K.: 6.7, 22.17
Greshake, Bastian: 514B
Gribaldo, Simonetta: 185D
Griegel, Pierre: 454B
Griffiths, Chris: 23.9
Griffiths, Paul D: 18.9
Griffiths-Jones, Sam: 22.19

Grigoriev, Igor: 21.1
Groenen, Martien: 6.12
Groenendaal, Laura: 20.12
Gronau, Ilan: 22.1, 384B
Gross, Joshua: 668B
Grossen, Christine: 16.10
Grosser, Stefanie: 519B
Grossniklaus, Ueli: 5.4
Groussin, Mathieu: 178A, 448D, 534A
Gruhl, Franziska: 644D
Grunau, Christoph: 26.3
Grutzner, Frank: F.5
Gryganskyi, Andrii: 21.1
Grève, Pierre: 26.9
Gschwend, Andrea R.: 640D
Gu, Jianying: 15C
Gu, Langyu: 241A
Gu, Xun: 500B
Guardiano, Cristina: 106C
Guardiola, Ombretta: 707A
Gubanova, Aleksandra: 630A
Gubry-Rangin, Cecile: 179B
Gueguen, Laurent: 23.14
Guelbeogo, Wamdaogo M.: 10.9
Guenther, Catherine: 136D
Guercio, Annalisa: 436B
Guglielmini, Julien: 13.6
Guillaume, Frederic: 11.8
Guillot, Gilles: 103D
Guimaraes, Silvia: 323B
Guio, Lain: 22.12
Guirao-Rico, Sara: 359C
Guizar, Maria: 298A
Guldbrandtsen, Bernt: 9.5, 611B
Gulisija, Davorka: 104A
Gulko, Brad: 22.1
Gunnlaugsson, Þorvaldur: 11.22
Guo, Qianyun: 13.7
Gupta, Sunetra: 418D
Gut, Ivo: 149A
Gutaker, Rafal: 299B
Gutenkunst, Ryan: 20.6, 23.4, 530A
Gutierrez, Humberto: 616C, 706D
Guy, Lionel: 28.5, 457A, 473A
Guzman, Lynette: 7.4
Guéguen, Laurent: 200C
Guérin, Cyprien: 512D
Gymrek, Melissa: 24.1
Györkei, Ádám: 4.16, 8D, 270B, 679A
Gärtner, Kerstin: 566A
Gómez-Sánchez, Daniel: 159C
Götherström, Anders: 326A, 327B
Günther, Torsten: 327B
Haag, Christoph: 94C
Haag, Karen: 21.7
Haak, David: 10.2
Haak, Wolfgang: 11.12, 303B
Habas, Rémy: 152D
Habel, Jan Christian: 107D
Habib, Michel: 465A
Haeseler, Arndt von: 378D
Hahn, Beatrice: 4.4
Hahn, Christoph: 386D
Hahn, Jang-Ho: 17A
Hahn, Matthew: 10.2
Hajduskova, Eva: 596C
Halbert, Jean: 22.4, 26.4
Hall, David: 20.7
Hall, Erica: 291C
Hall, James: 8.9
Hallast, Pille: 101B, 352D
Hallatschek, Oskar: 1.15, 11.7
Halloran, M. Elizabeth: 424B
Hallsson, Jon: 148D
Hamm, Christopher: 367B
Hammond, Thomas M.: 21.4
Hampl, Vladimir: 729D
Hanage, William: 8.4
Hancock, Angela: 11.19
Hane, Masaya: 670D
Hanihara, Tsunehiko: 79C, 341B
Hankeln, Thomas: 683A
Hansen, Greta: 310A
Hansen, Marianne HS: 306A
Hansen, Matthew: 87C, 366A
Hansen, Thomas F.: 76D
Hansen, Torben: 5.6
Hanson, Sara J.: 21.11
Hao, Huaining: 498D
Hao, Weilong: 509A
Haque, Tanzina: 434D
Hara, Yuichiro: 218A
Hardy, Matt: 364C
Haring, Elisabeth: 115D, 603B, 610A, 625D
Harkins, Kelly M: 636C
Harpak, Arbel: 139C
Harpe, Marylaure de La: 98C
Harris, Kelley: 1.6
Harris, Robert: 24.4
Harrison, Ellie: 8.2, 8.9
Harrison, Peter: 26.5
Harrison, Richard: 84D
Harrop, Thomas: 568C
Harrow, Jen: 364C
Harshman, Lawrence G.: 22.17
Hart, Kristen: 129A
Hart, Tom: 173B
Hartasánchez, Diego: 245A, 249A
Hartl, Daniel: 275C
Hartley, Molly: 559B
Hartmann, Fanny: 21.6
Harvati, Katerina: 325D
Haudry, Annabelle: 3.9
Hautaniemi, Sampsa: 15.1
Havill, Nathan: 587B
Havukainen, Heli: 5.21
Hawkins, Nichola: 333B
Hayakawa, Toshiyuki: 670D
Hayes, M. Geoffrey: 55C
Hazkani-Covo, Einat: 499A
Hazzouri, Khaled M.: 10.13
He, Fei: 666D
He, Yibo: 125A
Heaps, Sarah E.: 28.2
Heath, Tracy: 5.12
Heberlein, Magdalena: 7.3, 64D, 266B
Hedgecock, Dennis: L 8D
Hedges, S. Blair: 447C
Heel, David van: 23.9
Heerschop, Sacha: 402D
Hefter, Avigayel: 576C
Heil, Caiti Smukowski: 3.2
Heinze, Berthold: 5.1
Heißl, Angelika: 24.3
Helbing, Sophie: 3.12
Hellenthal, Garret: 47C
Hellenthal, Garrett: 11.13, 11.16, 11.3
Heller, Rasmus: 377C
Hellmann, Ines: 128D, 163D
Henderson, Donna: 173B
Henn, Brenna: 23.6, 136D
Henry, Isabelle: 10.12
Herbert, Matthew: 17.7
Herbig, Alexander: 25.1, 319B, 450B, 469A
Herissant, Lucas: 13.2
Herlyn, Holger: 20D
Hermisson, Joachim: 6.13, 11.19, 16.6, 50B, 117A, 163D, 168A
Hernandez, Erick: 490C
Hernandez, Ryan: 1.10, 23.12, 555B
Herniou, Elisabeth: 3.3
Hernández, Jéssica: 268D
Hernández-Muñiz, Jenniffer: 356D
Herrel, Anthony: 465A
Herrero, Javier: 364C
Hershberg, Ruth: 460D, 501C
Herzyk, Pawel: 582A
Hettiarachchi, Nilmini: 5.14
Hey, Jody: 99D
Heyer, Evelyne: F.4
Hibner, Urszula: 14.4
Hietpas, Ryan T.: 13.3
Higgie, Simone: 7.6
Higham, Tom: 539B
Highnam, Gareth: 722D
Hilbrant, Maarten: 22.19
Hile, Suzanne: 24.4
Hill, Tom: 476A, 487D
Hilmarsson, Hrannar: 148D
Hilu, Khidir: 2.10
Hirano, Masayuki: 518A
Hirbo, Jibril: 87C
Hirsch, Penny: 525D

Hittinger, Chris Todd: 21.9
Ho, Simon: 2.7, 30B, 195B, 196C
Ho, Wei-Chin: 20.11
Hoal, Eileen: 136D
Hoban, Sean: 345B
Hobolth, A.: 175D
Hobolth, Asger: 557D
Hobza, Roman: 350B
Hochberg, Michael: 14.4
Hochreiter, Sepp: 29A
Hodges, Scott: 410D
Hoekstra, Hopi E.: 6.9, 167D
Hoerger, Anja C.: 123C
Hofman, Corinne: 441A, 539B
Hofreiter, Michael: 41A, 300C
Hogeweg, Paulien: 609D
Hohmann, Gottfried: 362A
Hohmann, Nora: 133A
Holdstock, Jolyon: 18.9
Holl, Anne-Catherine: 85A
Holland, Stephen: 518A
Hollfelder, Florian: 64D, 266B, 292D
Holm, Karl: 10.13
Holmstrup, Martin: 6.1
Holst, Malin: 317D
Holstege, Frank C. P.: 703A
Hombert, Jean-Marie: F.4
Honap, Tanvi Prasad: 543B
Hong, Soon Gyu: 453A
Honkela, Antti: 332A
Hood, Michael: 505A, 506B
Hoogland, Menno: 441A
Hopf, Thomas: 24D
Hopfen, Corinna: 17.7
Hopkins, Rachel: 539B
Horak, Ales: 613D
Horianopoulos, Linda: 705C
Horikoshi, Renato: 655A
Horimoto, Andrea: 37A
Horn, Matthias: 457A, 732C
Horner, David: 551B
Hornett, Emily: 135C
Horváth, Balázs: 8D
Horváth, Péter: 149A
Hossain, Mukarram: 428B
Hou, Yali: 134B
Howe, Christopher: 664B
Howell, Megan: 36D
Hron, Tomas: 431A
Hrtyan, Mónika: 8D
Hsieh, Li-Ching: 26B
Hsieh, PingHsun: 366A
Hu, Haichao: 704B
Hu, Haiyang: 71C
Hu, Jinyong: 12.4
Hu, Songnian: 642B
Hu, Zheng: 496B
Hua-Van, Aurélie: 478C
Huang, Yu-Ching: 370A
Huang, Yun: 22.1
Huang, Zhongyun: 141A
Huber, Christian: 163D
Huber, Christian D.: 128D
Hubisz, Melissa J.: 22.1
Hublin, Jean-Jacques: 309D, 319B
Huchon, Dorothee: 561D
Huehne, Pattana Srifah: 565D
Huerta-Sanchez, Emilia: 5.6, 16.7
Hughes, Austin: 426D
Hulsey, C. Darrin: 6.16
Hunt, Toby: 364C
Hunter, Margaret: 129A
Hunter-Mann, Kurt: 317D
Hurst, Greg: 135C
Hurst, Laurence: 7.2, 22.13, 645A, 663A, 679A
Huson, Daniel: 450B, 469A
Hussaini, Yusuf: 434D
Hussin, Julie: F.5
Huylmans, Ann-Kathrin: 673C
Hvilsom, C.: 175D
Hyrien, Olivier: 553D
Hytönen, Timo: 148D
Hänfling, Bernd: 380B
Hélène, Quach: F.4
Hünemeier, Tábita: 174C
Iamarino, Atila: 14.2
Ijaz, Umer: 4.3
Ilhan, Judith: 23C, 83C, 155C
Illingworth, Chris: 1.2, 346C
Immel, Alexander: 319B
Indjeian, Vahan: 12.3
Ingram, Brett A: 16.4
Innan, Hideki: 494C
Intakham, Chidchanok: 592C
Invergo, Brandon: 6B
Ion-Scotta, Michela: 446B
Ip, Melissa: 504B
Irene, Hernando-Herraez: 5.10
Irez, Selen: 22.15
Iriarte, Eneko: 327B
Irimia, Manuel: 187A, 190D
Irisarri, Iker: 204C
Irwin, Kristen: 432B
Ishida, Hajime: 341B
Ishida, Takafumi: 492A
Ishizaki, Naoya: 341B
Ito, Kimihito: 435A
Ivivs, Zoltan: 7.2
Iwamoto, Eisuke: 216C
Iwasaki, Risa: 154B
Iyer, Lakshminarayan: 518A
Izaabel, Hassan: 47C
Izagirre, Neskuts: 682D
Izquierdo, David: 5.19, 176A
Izsvak, Zsuzsanna: 7.2
Jackson, Ben: 572C
Jackson, Robert: 525D
Jacob, Katja: 22.20
Jacobs, Arne: 582A
Jacquier, Hervé: 13.6
Jacquin-Joly, Emmanuelle: 387A
Jagoda, Joanna: 22.15
Jain, Kanika: 583B
Jain, Kavita: 329B
Jakobsen, Kjetill S.: 306A
Jakobsson, Mattias: 327B, 337B
Jaksch, Katharina: 115D
Jakšić, Ana Marija: 638B
Jallow, Muminatou: 621D
James, Jennifer: 11.20
James, Katherine: 415A
James, Timothy: 21.1, 21.7
Janich, Peggy: 644D
Janouskovec, Jan: 596C
Janoušek, Václav: 381C
Jansen, Gunther: 14.4
Janssen, Arnold: 8.6, 224C
Jaramillo, Laura Salazar: 383A
Jardillier, Ludwig: 4.9
Jasinska, Anna: 161B
Jaszczyszyn, Yan: 553D
Jauffrit, Frédéric: 199B
Jay, Flora: 5.16
Jean, Greraldine: 22.15
Jeffreys, Alec: 328A
Jeffries, Daniel: 380B
Jenkins, Paul: 296C
Jenner, Ronald: 222A
Jensen, Jeffrey: 16.6, 50B, 168A, 240D, 425C, 432B, 652B
Jensen, Jeffrey D.: 6.9, 13.3, 167D
Jensen, Per: 680B
Jentoft, Sissel: 306A
Jeong, Choongwon: 23.1
Jerison, Elizabeth: 13.10
Jesús, Adriana Gordon-De: 356D
Ji, Hyeon-So: 17A
Jia, Mingming: 498D
Jiang, Li: 145A
Jiang, T.: 175D
Jiang, Xiaowei: 430D
Jin, X.: 175D
Jinam, Timothy: 340A
Jirku, Milan: 596C
Jo, Yeonhwa: 461A, 462B
Joachim, Anja: 456D
Jobling, Mark: 101B, 111D, 352D
Joh, Keiichiro: 341B
Johannesson, Hanna: 5.25, 21.4
Johnsen, Jill: 4.2
Johnson, Gary: 23.7
Johnson, Jeremy: 631B
Johnson, Loretta: 122B
Johnson, Louise: 14.5
Johnson, Philip L. F.: 305D, 424B
Johnsson, Martin: 680B
Johri, Parul: 257A

Jonas, Agnes: 332A
Jones, Felicity: 22.14
Jones, Meaghan: F.4
Jordan, Daniel: 2.13
Joron, Mathieu: 388B
Josephs, Emily: 10.13, 22.26
Jouannic, Stefan: 568C
Jové, Thomas: 8.11
Joy, Abigail: 468D
Joyce, Domino: 386D
Jr., José de Sousa e Silva: 606A
Jue, Nathaniel: 3.10
Julia, MacIsaac: F.4
Julian-Sanchez, Adriana: 717C
Julien, Philippe: 26.4
Jung, Matthieu: 2.6
Juric, Ivan: 25.4
Jónsson, Hákon: 103D
Jørgensen, Frank Grønlund: 43C
Jørgensen, Marit E: 5.6
Jørgensen, Steffen: 28.5
Kablik, Lea: 714D
Kabza, Michal: 282B, 488A
Kacharia, Fenil: 743B
Kaeppler, Adrienne: 310A
Kaessmann, Henrik: 7.11, 15.3, 22.4, 26.4, 644D
Kahles, André: 22.15
Kahli, Malik: 553D
Kahn, Daniel: 7C
Kaiser, Tobias S.: 5.3
Kaja, Elzbieta: 687A
Kalaghatgi, Prabhav: 440B
Kalajdzic, Predrag: 238B
Kalapis, Dorottya: 149A, 703A
Kalbe, Martin: 455C, 22.1
Kaltenbach, Miriam: 263C
Kaltenegger, Elisabeth: 259C
Kamm, John: 164A
Kana, Vlastislav: 320C
Kane, Ndjido Ardo: 698D
Kaneko, Kunihiko: 609D
Kang, Jonathan: 217D
Kang, Seunghyun: 233C
Kang, Yoonjee: 690D
Kankare, Maaria: 647C
Kapopoulou, Adamandia: 652B
Kapralov, Maxim: 20.10
Kapun, Martin: 11.21
Karakoc, Emre: 220C, 304C, 403A
Karczewski, Konrad: 23.9
Karen, Hunt: 23.9
Karger, Amir: 7.5
Karin, Eli Levy: 576C, 740C
Karlsson, Elinor: 631B
Karnkowska, Anna: 729D
Karsten, Jordan: 594A
Kasala, Naga Apoorva: 455C
Kasapidis, Panagiotis: 630A
Kashyap, Mehr: 3.1
Kassen, Rees: 13.9
Kataoka, Hiroshi: 79C
Kater, Martin: 568C
Katsumura, Takafumi: 79C
Katz, Sophia: 460D
Kauffman, Kathryn: 468D
Kautt, Andreas: 10.3
Kawamura, Shoji: 79C, 341B
Ke, Huei-Mien: 615B
Kearns, Joanne: 16.4
Keeble, Sara: L 4D
Keeling, Patrick J.: 596C
Keena, Melody: 587B
Kehdy, Fernanda: 37A
Keightley, Peter: 586A
Keim, Daniel: 653C
Kelemen, Reka K.: 158B
Kellam, Paul: F.7
Keller, Laurent: 370A
Keller, Lukas: 16.10
Kelly, Brian: 5.18
Kelso, Janet: 16.2, 39C, 309D
Kemena, Carsten: 532C
Kemi, Ulla: 86B
Kemmeren, Patrick: 703A
Kempenaers, Bart: 373D
Kenny, David: 452D
Kent, Tyler: 585D
Kern, Andrew: 99D, 252D
Kerver, Halie N.: 26.4
Kessler, Michael: 10.1
Ketting, René: 694D
Key, Felix M.: 23.8, 25.3
Khaitovich, Philipp: 16.2, 71C
Khalil, Charbel A.: 92A
Khatri, Bhavin: 10.8
Khrameeva, Ekaterina: 16.2
Kidd, Jeffrey: 3.4, 11.14
Kile, Benjamin: 3.6
Kim, Eun-Young: 562A
Kim, Hie Lim: 95D
Kim, Kang-Wook: 572C
Kim, Kangchon: 120D
Kim, Kyung Mo: 453A
Kim, Lindsay: 8.4
Kim, Michelle: 42B
Kim, Nam-Kuk: 147C
Kim, Sanghee: 233C
Kim, Seungsoo: 22.23
Kim, Tae-Ho: 17A
Kim, Yeon-Ki: 17A
Kim, Yuseob: 6.3, 104A, 120D
Kimura, Ryosuke: 341B
Kin, Koryu: 20.15
King, Charlotte: 30B
King, Daniel B.: 10.15
King, Nicole: 27.1
King, Turi E.: 101B
Kingan, Sarah: 5.11
Kingma, Guilherme: 112A
Kingsley, David: 12.3, 22.14, 136D
Kintses, Bálint: 4.16, 270B
Kircher, Martin: 688B
Kirchner, Sandra: 603B, 610A, 618A
Kirsanow, Karola: 25.6
Kirschner, Marc W.: 7.5
Kissinger, Jessica: 291C
Kitajima, Ken: 670D
Kitano, Takashi: 595B
Kittelmann, Sebastian: 22.10
Kittke, Volker: 688B
Kivinen, Katja: 621D
Kivisild, Toomas: 26.6, 312C
Kjesbu, Olav S.: 306A
Kjærgaard, Peter: 43C
Klasberg, Steffen: 7.3
Klassmann, Alexander: 538A
Kleinert, Philip: 23.8
Klenerman, Paul: 18.8
Klepikova, Anna: 157A
Klink, Galya: 404B
Kloesges, Thorsten: 105B, 676B
Knijff, Peter de: 101B
Knipper, Corina: 608C
Knowles, James: 631B
Knowles, L. Lacey: 375A
Ko, Amy: 38B
Koay, Evelyn: 434D
Koblmüller, Stephan: 581D
Kobor, Michael: F.4
Koch, Evan: 348A
Koch, Marcus: 133A
Koch, Robin: 83C
Kochko, Alexandre de: 102C
Koebnik, Ralf: 454B
Koelle, Katia: 18.6
Koenig, Daniel: 22.25
Kofler, Robert: 3.14, 6.14
Koganebuchi, Kae: 341B
Kolaczkowski, Bryan: 253A
Kolaczkowski, Oralia: 253A
Kolchanova, Sofia: 412B, 413C
Koletou, Malamati: 24.2
Koltookian, Michele: 631B
Komarova, Natalia: 14.1
Komljenovic, Andrea: 718D
Kompella, Purnima S.: 25A
Konczal, Mateusz: 6.5
Kondrashov, Alexey: 157A
Kondrashov, Fyodor: 1.13, 67C, 283C
Konecny-Dupre, Lara: 23.14
Kongkachana, Wasitthee: 577D
Konik, Peter: 613D
Kono, Mieko: 511C
Konrad, Anke: 20.4
Koolkarnkhai, Pavitchaya: 592C
Koonin, Eugene: 8.1
Kopp, Artyom: 12.2
Kopp, Michael: 117A
Korb, Judith: 532C
Kordis, Dusan: 739B
Korithoski, Bryan: 253A
Korneliussen, Thorfinn S: 5.6
Korneliussen-Sand, Thorfinn:

305D
Koroteev, Maxim: 620C
Kosheleva, Katya: 142B, 449A
Kosiol, Carolin: 2.2, 332A
Koskela, Jere: 296C
Koteja, Pawel: 6.5
Kousathanas, Athanasios: 165B
Kovarik, Aleš: 526A
Kovács, Károly: 703A
Kowalik, Timothy: 425C
Kozaczek, Melisa: 583B
Kozlov, Alexey: 219B
Kramer, Maxwell: 17.2
Kranz, Anna-Lena: 17.2
Krasheninnikova, Ksenia: L 2B
Krasovec, Marc: 61A
Krasovec, Rok: L 9A
Kratochwil, Claudius: 19.5, 713C
Kratsch, Christina: 179B
Kraus, F Bernhard: 3.12
Kraus, Lee: 19.3
Krause, Johannes: 25.1, 303B, 318A, 319B, 321D, 325D, 450B, 469A, 608C
Krause-Kyora, Ben: 742A
Kreiswirth, Barry: 201D
Kremer, Antoine: 301D
Krenek, Sascha: 563B
Krishnan, Arunkumar: 191A
Kroh, Andreas: 625D
Kronforst, Marcus: 19.4
Kruckenhauser, Luise: 115D, 603B, 610A, 618A
Kruse, Lars H.: 11C
Krutzen, Michael: 722D
Kryazhimskiy, Sergey: 13.10
Kryuchkova, Nadja: 685C
Kryukov, Alexey: 399A
Ku, Chuan: 8.10, 730A
Kubera, Anchanee: 442B
Kubiak, Magdalena: 488A
Kuderna, Lukas: 49A, 408B
Kudla, Grzegorz: 66B
Kuhl, Heiner: 373D
Kujala, Sonja: 108A
Kuleung, Chatuporn: 634A
Kumabe, Toshihiro: 341B
Kumagai, Seiji: 9.2
Kumar, Sudhir: 447C, 495A
Kumirov, Vlad: 4.11
Kundu, Samit: 434D
Kunstman, Kevin: 562A
Kuo, Alan: 21.1
Kuo, Chih-Horng: 8.10, 728C
Kuo, Lynn: 2.12
Kupczok, Anne: 83C, 727B
Kuraku, Shigehiro: 218A, 278B
Kurt, Noemi: 9.3
Kuruoglu, Ercan E.: 214A
Kvikstad, Erika: 480A
Kwiatkowski, Dominic: 621D
Kwiatkowski, Marcel: 264D
Kyriakidou, Pelagia: 19C
Kégl, Balázs: 8D
Kéri, Zsófia Nóra: 737D
Künzel, Sven: 238B
Launay, Adrien: 470B
Laayouni, Hafid: 6B
Lachance, Joseph: 87C, 366A
Lachmann, Michael: 25.3, 25.8
Lack, Justin: 12.5
Lacoste, Sandrine: 21.5
Lacroix, Vincent: 650B
Laetsch, Dominik: 383A
Laffoon, Jason: 539B
Lafontaine, Ingrid: 21.2, 262B
Lafosse, Sophie: 109B
Lagane, Frédéric: 301D
Lah, Ljerka: 11.22
Lai, Cheng-Kuo: 26B
Lai, De-Hua: 596C
Lai, Mei-Chin: 615B
Laiba, Efrat: 549D
Lallier, François: 656B
Lalueza-Fox, Carles: 539B
Lamatsch, Dunja: 349A
Lamaze, Fabien: 385C
Lamb, Annika M: 16.4
Lambert, Amaury: 297D
Lan, Xun: 22.18
Landan, Giddy: 8.6, 224C, 727B
Landerer, Cedric: 510B
Landstorfer, Richard: 653C
Landweber, Laura: 4.13
Lanfear, Robert: 11.20
Lang, Gregory I: 449A
Langat, Pinky: F.7
Langdon, Quinn: 21.9
Lange, Jeremy: 236B
Lapierre, Marguerite: 297D
Lapoint, Richard: 138B
Lareau, Liana: 15.5
Larsen, Klaus: 6.1
Larson, Greger: 6.12, 30B
Larsson, Ronny: 21.7
Lartillot, Nicolas: 4D, 206A, 710D
Lascoux, Martin: 10.13
Lassalle, Florent: 8.5, 18.9, 438D
Lata, Rodrigo: 81A
Lau, Nelson: 3.6
Lau, Quintin: 660B
Laurent, Romain: 109B, 362A
Laurent, Stefan: 6.9, 167D, 240D
Laurin-Lemay, Simon: 710D
Lauzi, Stefania: 436B
Laval, Guillaume: 126B
Lavi, Eitan: 3.11
Lavrov, Dennis: 258B
Lawson, Daniel: 11.3
Lawson-Handley, Lori: 380B
Lazar, Viktoria: 4.16
Lazaridis, Iosif: 303B
Lazzaro, Brian: 10.9
Le, Si Quang: 621D
LeRiche, Kaitlynn: 485B
Leandre, Bastien Saint: 478C
Lebrasseur, Ophelie: 30B
Leclercq, Sébastien: 26.9, 735B
Lecocq, Michel: 534A
Ledda, Alice: 8.8
Lee, Changhee: 147C
Lee, Chih-Chi: 370A
Lee, Gang-Seob: 17A
Lee, Grace Yuh Chwen: 3.7
Lee, Hyoungseok: 690D
Lee, Jin-Ho: 147C
Lee, Jong-Yeol: 17A
Lee, Jungeun: 690D
Lee, Kristin: 6.8
Lee, Seong-Hun: 147C
Lee, Tae-Ho: 17A
Lee, Young Wha: 22.26
Lefebure, Tristan: 23.14
Lefeuvre, Pierre: 454B
Leffler, Ellen: 621D
Leftwich, Philip: 443C
Legendre, Xavier: 88A
Lehnherr, André: 248D
Leitch, Andrew: 56D
Leite, Daniel J.: 22.19
Lemaire, Katleen: 588C
Lemey, Philippe: 18.8
Lencer, Ezra: 84D
Lengauer, Thomas: 440B
Lenski, Richard: F.2
Lenz, Tobias: 1.3, 22.1
Leonard, Jennifer A.: 581D
Lepennetier, Gildas: 639C, 691A
Lerat, Emmanuelle: 3.9
Lercher, Martin: 5.20, 20.13, 20.2
Lercher, Martin J: 2B
Lessa-Horta, Bernardo: 37A
Leuenberger, Christoph: 165B, 537D
Leventhal, Gabriel: 13.5, 416B
Levine, Josh: 7.4
Levivier, Aurore: 4D
Levkivskyi, Ivan: 546A
Levsen, Nicholas: 574A
Levy, Sasha: 6.6, 605D
Lewis, Louise: 2.12
Lewis, Paul: 2.12
Lexer, Christian: 5.1, 10.1, 391A, 98C
Li, Chuan: 59C
Li, Cuiping: 642B
Li, Jianxu: 518A
Li, Junya: 145A
Li, Lin-Feng: 141A
Li, Tao: 497C
Li, Weiyi: 235A
Li, Wen-Hisung: 80D

Li, Wen-Hsiung: 615B, 686D
Li, Xin: 124D
Li, Xueyan: 19.4
Li, Y.: 175D
Li, Yawei: 496B, 498D
Li, Yuli: 642B
Li, Yuping: 13.2
Li, Zhen: 223B, 274B
Liang, Cong: 4.18, 20.15, 696B
Liang, Yipeng: 713C
Liao, Ben-Yang: 637A, 678D
Liberles, David: 20.4, 243C
Libkind, Diego: 21.9
Lichtenstein, Leonie: 22B
Liechti, Angélica: 22.4, 26.4
Lifton, Richard: 14.2
Lilburn, Timothy: 15C
Lilley, Andrew: 8.9
Lillie, Mette: 536C
Lim, Marisa: 379A
Lima, Thiago: 671A
Lima-Costa, Maria Fernanda: 37A
Lin, Hsuan-Yu: 637A
Lin, Meng: 136D
Lin, Qiang: 363B
Lin, Xuan: 486C
Lin, Yen-Lung: 304C
Lind, Anders E.: 28.5, 181D
Lindblad-Toh, Kerstin: 394D, 631B
Linderholm, Anna: 30B
Linderoth, Tyler: 483D
Lindo, John: 25.5
Lindström, Jan: 4.3
Lindtke, Dorothea: 10.1, 391A
Ling, Zhihao: 646B
Linheiro, Raquel S.: 708B
Linnen, Catherine R: 167D
Linnenbrink, Miriam: 654D
Lintermans, Mark: 16.4
Lipecki, Grzegorz: 320C
Lipsitch, Marc: 8.4
Lis, John: 19.3
Lisa, McEwen: F.4
Lischer, Heidi: 16.10
Liti, Gianni: 6.15
Litovchenko, Maria: 240D
Liu, Baoning: 17.10
Liu, Gangbiao: 500B
Liu, Guichun: 19.4
Liu, Guojing: 498D
Liu, Jian-Feng: 145A
Liu, Jianlin: 497C
Liu, Kevin: 2.3
Liu, Liang: 20.4
Liu, Qingpo: 704B
Liu, Shiping: 143C
Liu, Xiaomin: 136D
Liu, Xingquan: 704B
Livesey, Rick: 19.1
Lizzo, Giulia: 314A
Llamas, Bastien: 11.12
Llopart, Ana: 16.8
Lo, Nathan: 2.7, 195B
Lo, Wen-Sui: 8.10, 728C
Lockyer, Christina: 11.22
Loe, Louise: 11.12
Loeffler, Taylor: 6.10
Loenen, Freek B van: 18.9
Logacheva, Maria: 157A
Logsdon, John M.: 334C
Lohmueller, Kirk E.: 1.9, 23.7
Lommel, Leentje Van: 588C
Long, Anthony D.: L 3C
Long, Manyuan: 7.9, 276D, 640D
Long, Quan: 22.15, 85A
Longini, Ira: 424B
Longobardi, Giuseppe: 106C
Loo, Bert van: 64D, 266B
Loog, Liisa: 30B
Loot, Céline: 65A
Lopes-Marques, Monica: 277A
Lopez, Alberto: 673C
Lopez, Beatriz: 151C
Lopez, Philippe: 465A, 466B, 741D
Lopez-Garcia, Purificacion: 4.9, 28.4
Lopez-Villavicencio, Manuela: 21.5
Lorente-Galdos, Belen: 49A
Lorenz, Camila: 75C
Lorenzen, Eline: 143C
Lorincz, Matthew: 7.2
Lourido, Santiago Lens: 11.22
Loveland, Jane: 364C
Lovette, Irby: 384B
Lozada-Chávez, Irma: 189C
Lozada-Chávez, Nabor: 8.6, 180C
Lu, Guoqing: 390D
Lu, Mei-Yeh Jade: 80D
Lu, Xiangwu: 5.18
Lu, Xiaowen: 20.2
Lu, Xuemei: 134B, 496B, 497C
Lui, Prudence: 12D
Luikart, Gordon: 16.4
Luisi, Pierre: 6B
Lukes, Julius: 596C
Lund, Mogens: 9.5
Lunt, David: 383A
Lunter, Gerton: 173B, 480A
Luobu, Ouzhu: 125A
Luria, Victor: 7.5
Luse, Meagan: 334C
Luyten, Isabelle: 482C
Luzzago, Camilla: 436B
Lv, Xuemei: 498D
Lwin, Nyein: 719A
Ly, Goki: 109B
Lycett, Samantha: 2.6
Lynch, Kristen: 643C
Lynch, Michael: 235A, 257A, 529D
Lynch, Vincent: 3.8
Lyon, Jarod P: 16.4
Lyons, Alice: 11.12
Lázár, Viktória: 8D, 13.8, 20.2, 270B
López, Saioa: 682D
López-Cerdán, Adolfo: 58B
López-García, Purificación: 464D
López-Villavicencio, Manuela: 508D
Löytynoja, Ari: 629D
Mason, Chris: 330C
MacArthur, Daniel: 23.9
MacLean, Craig R.: 8.3, 4.15
MacLeod, Iona: 9.5
MacMahon, Eithne: 434D
Macas, Jiři: 10.14
Machado, Heather: 1.5
Machado, Moara: 37A
Machado, Pablo: 490C
Machado-Schiffiano, Gonzalo: 10.3
Mackiewicz, Dorota: 347D
Mackiewicz, Pawel: 289A, 320C
Macqueen, Daniel: 179B
Madeoy, Jennifer: 22.23
Madern, Dominique: 178A
Madrigal, Jazmín Ramos: 35C, 310A
Maere, Steven: 274B, 311B
Magalhaes, Carolina: 51C
Magalhaes, Wagner: 591B
Magalhaes, Wagner CS: 37A
Magis, Cedrik: 203B
Magne, Maxime: 454B
Mahadevaiah, Shantha: 26.4
Mahajan, Shivani: 358B, 648D
Maher, Eamonn: 23.9
Mahmoudisaber, Morteza: 392B
Mai, Dat: 360D, 414D
Mailund, T.: 175D
Mailund, Thomas: 10.6, 23.11, 43C
Maio, Nicola De: 2.2, 11.4
Maixner, Frank: 450B
Majander, Kerttu: 469A
Majumder, Partha: 528C
Makalowska, Izabela: 234D, 281A, 282B, 488A, 687A
Makohon-Moore, Sasha: 19.2, 712B
Makova, Kateryna: 24.4
Makova, Kateryna D.: 26.8
Malard, Florian: 23.14
Malaspinas, Anna-Sapfo: 305D, 546A
Malausa, Thibaut: 446B
Malek, Joel A.: 92A
Maley, Carlo: 14.4
Malhi, Ripan: 25.5
Malhi, Ripan S.: 55C

Malinsky, Milan: 10.5
Malinverni, Roberto: 394D
Mallard, François: 22.24, 159C
Mallet, Ludovic: 512D
Mallick, Swapan: 607B
Mallory, Michael: 643C
Malmström, Helena: 327B
Mandage, Rajendra: 540C
Manel, Stéphanie: 102C
Manhart, Michael: 738A
Manica, Andrea: 169B
Mank, Judith: 26.5
Mann, Allison: 441A
Mannakee, Brian: 530A
Mano, Shuhei: 341B
Manousaki, Tereza: 369D
Mantelatto, Fernando Luis: 590A
Mantica, Federica: 695A
Marais, Gabriel: F.3
Marande, William: 85A
Marcet-Houben, Marina: 622A
Marchet, Camille: 650B
Marco, Antonio: 651A
Marcus, Joseph: 23.1, 25.7
Maricle, Brian: 122B
Mariette, Stephanie: 160A
Marin, Julie: 447C
Marin, Ray: 26.4
Marinho, Marco: 697C
Marino, Ilaria Anna Maria: 53A
Markov, Gabriel: 247C
Marks, Debora: 24D
Markus, Barak: 24.1
Marnetto, Davide: 695A
Marquardt, Jeannine: 56D
Marques, David: 376B
Marques, Lucas: 215B
Marques, Patrícia Isabel: 82B, 665C
Marquès-Bonet, Tomàs: 3.4, 5.10, 49A, 268D, 408B, 581D, 722D
Marsden, Clare: 23.7
Martijn, Joran: 28.5, 457A
Martin, Alicia: 136D
Martin, Hilary: F.5
Martin, Hélène: 401C
Martin, Trevor: 4.10
Martin, Ulrich: 688B
Martin, William: 8.6, 180C, 224C, 255C, 730A
Martinez-Cadenas, Conrado: 682D
Martiniano, Rui: 317D
Martino, Lucia De: 707A
Martins, Andre: 19.3
Martí-Solano, Maria: 146B
Martínez, Pedro: 267C
Martínez-Cruzado, Juan C.: 356D, 626A, 628C, 677C, 724B
Martínez-Fundichely, Alexander: 558A
Maruki, Takahiro: 529D
Marzo, Mar: 708B
Marín-Maldonado, Frances M: 677C
Marón, Carina: 344A
Mascagni, Flavia: 489B
Masel, Joanna: 7.4, 20.4, 285A
Masly, John: 17.7
Masoero, Loretta: 436B
Mason, Dan: 23.9
Massip, Florian: 2.15, 2.16
Massy, Ken: 608C
Mata-Míguez, Jaime: 579B
Matchett, Ashley: 27C
Materna, Christopher: 714D
Mathema, Barun: 201D
Mather, Niklas: 417C, 433C
Mathew, Lisha: 652B
Mathews, Carol: 631B
Mathiesen, Rune: 82B
Matos, José: 326A
Matos, Luis: 18.10
Matrin, William F.: 733D
Matschiner, Michael: 306A
Matsumae, Hiromi: 156D, 594A
Matsumoto, Tomotaka: 580C
Mattila, Tiina: 593D
Matts, Jessica: 3.6
Matuszewski, Sebastian: 13.3, 117A
Mauchline, Tim: 525D
Maud, Fagny: F.4
Maumus, Florian: 18.2
Mauro, Diego San: 261A, 393C
Mavingui, Patrick: 3.15
Maxime, Rotival: F.4
May, Celia: 328A
Mayer, Christine: 76D
Mayo, Avraham E.: 20.14
Mayrose, Itay: 254B
Mazel, Didier: 65A
Mazel, Florent: 448D
Mazzocchi, Maria Grazia: 630A
McCabe, Matthew: 452D
McCann, Honour: 444D
McCann, Jamie: 10.14, 550A
McCarthy, Shane: 23.9, 607B
McCauley, John: F.7
McConnell, Ellen: 654D
McCreight, Jennifer: 22.5
McCune, Amy: 84D
McCutcheon, John: 517A
McDonald, Michael: F.2
McDonald, Michael J: 449A
McDonnell, Christine: 317D
McGee, Lesley: 8.4
McGrath, Casey: 257A
McGregor, Alistair P.: 17.7, 22.10, 22.19
McHardy, Alice: 179B
McInerney, James: 8.6
McKenzie, Joanna: 664B
McKeown, Alesia: F.8
McLaren, Michael: 286B
McLaughlin, Russell: 317D
McManus, Kimberly: 716B
McNellis, Timothy: 687A
McTavish, Emily Jane: 207B
McVean, Gilean: 415A
Mead, Simon: 11.3
Meaux, Juliette de: 666D
Medlar, Alan: 629D
Medvedev, Paul: 24.4, 26.8
Meier, Cecile: 5.1
Meisel, Richard: 17.5
Melderen, Laurence Van: 734A
Melzer, Rainer: 273A
Memari, Yasin: 607B
Mendonça, Fernando: 51C, 280D
Mendoza, Alex de: 187A
Mendoza-Revilla, Javier: 47C
Meneu, Juan R.: 23.8
Meneu, Juan Ramón: 362A
Meng, Dazhe: 18B, 22.15, 85A
Menozzi, Giorgia: 315B
Merenciano, Miriam: 479D
Merilä, Juha: 351C, 597D
Merker, Stefan: 114C, 402D
Merkin, Jason: 15.1
Merritt, Justin: 743B
Merry, Alexandra: 36D
Mestres, Francesc: 146B
Meter, Michael van: 3.1
Metspalu, Mait: 225D
Metzger, Brian: 22.22
Metzler, Dirk: 23.5
Meyer, Austin: 419A
Meyer, Axel: 10.3, 19.5, 204C, 279C, 657C, 713C
Meyer, Diogo: 177B
Meyer, Eric: 22.16
Meyer, Jan: 247C
Meyer, Matthias: 322A, 544C
Meyer, Wynn: 414D
Meyniel, Jean-Philippe: 21.2
Mezey, Jason G.: 92A
Mezzavilla, Massimo: 46B
Mhuantong, Wuttichai: 442B
Mia, Gian Mario De: 436B
Michalek, Jan: 596C
Michalovova, Monika: 26.8
Mikkelsen, Teis: 6.1
Milani, Liliana: 244D
Milasin, Jelena: 101B
Milinski, Manfred: 455C, 22.1
Millan, Alvaro San: 4.15, 8.3
Miller, Craig: F.6
Miller, Michael R.: 516D
Miller, Webb: 95D
Miller, Wolfgang: 4.7, 487D
Milne, Richard S: 18.9
Minchiotti, Gabriella: 707A
Mineta, Katsuhiko: 580C
Ming, Ray: 26.1

Mingkwan, Nipitwattanaphon: 370A
Mirarab, Siavash: 2.1
Miró-Herrans, Aida T.: 55C
Mishmar, Dan: 396B
Miska, Eric A.: 10.5
Missero, Caterina: 707A
Mitani, Hiroshi: 79C
Mitchell, Patrick: 8.4
Mitrano, Amy: 5.10
Mittelman, David: 722D
Mitteroecker, Philipp: 70B
Mittnik, Alissa: 325D, 608C
Miura, Sayaka: 495A
Miyagi, Ryutaro: 398D
Miyamoto, Michael: 18.10
Miyazawa, Seita: 48D
Miyazawa, Takayumi: 420B
Mizrahi, Itzhak: 155C
Mlitz, Veronika: 251C
Modolo, Laurent: 3.15
Moeller, Andrew: 4.4
Mohammed, Jaaved: 6.7
Mohandesan, Elmira: 300C
Mohorianu, Irina: 57A, 635B
Molak, Martyna: 338C
Molineris, Ivan: 695A
Mollion, Maéva: 23.3
Molnar, Ruxandra I.: 5.25
Moltke, Ida: 5.6
Molyneaux, Pamela: 434D
Mondal, Mayukh: 528C
Monette, Matthew: 12.5
Money, Daniel: 584C
Monfardini, Raquel: 14B, 697C, 699A
Monit, Christopher: 18.5
Montagné, Nicolas: 387A
Montagud-Martínez, Roser: 58B
Montanucci, Ludovica: 6B
Montenegro, Alvaro: 95D
Montero, Yanetsy: 152D
Monteserin, Johana: 151C
Montgomery, Janet: 317D
Montgomery, Stephen: 26.5
Montoya-Burgos, Juan I.: 2.8, 11.11
Monzón-Sandoval, Jimena: 616C, 706D
Mooers, Arne: 93B
Moorjani, Priya: 5.8
Morales, Millke: 51C
Moran, John: 3.4
Moran, Nancy: 4.6
Moreira, David: 4.9, 28.4, 464D
Moreira, Rennan: 591B
Morello, Timothy: 3.1
Moreno, Ana: 436B
Moreno-Hagelsieb, Gabriel: 736C
Moreno-Mayar, Víctor: 305D
Moretti, Sébastien: 3C
Morgan, Colleen: 317D
Morgan, Ted: 122B
Moriarty, Ryan V.: 21.9
Moritz, Robin FA: 3.12
Morris, Chandler: 6.10
Mortimer, Richard: 11.12
Mortimer, Tatum: 6.11
Morvan, Claire: 23.14
Moses, Alan M.: 20.5, 25A
Mouguiama-Daouda, Patrick: F.4
Moumen, Bouziane: 26.9
Moura, Mónica: 527B
Mourier, Tobias: 477B
Mower, Jeffrey: 583B
Moyers, Bryan: 7.8
Moyle, Leonie: 10.2
Mozzi, Alessandra: 315B, 599B, 600C
Mueller, Jakob C.: 373D
Mueller, Kai F.: 5.17
Mugal, Carina F.: 221D
Mugue, Nikolai: 52D
Mullen, Kristen: 366A
Munain, Adolfo López de: 101B
Munch, Kasper: 10.6, 23.11, 175D
Munclinger, Pavel: 381C
Munoz, François: 102C
Murillo, Oscar: 495A
Murray, Brent: 705C, 721C
Murray, Gemma: 747B
Musharoff, Shaila: 17.9
Musser, Jacob: 4.18, 696B
Mustonen, Ville: 1.12, 346C
Muyle, Aline: F.3
Myers, Simon: 11.16, 11.3
Myles, Sean: 584C
Mylonas, Constantinos: 369D
Méheust, Raphaël: 44D, 741D
Méhi, Orsolya: 8D, 573D
Möller, Marlo: 136D
Müldner, Gundula: 317D
Müller, Stefan: 349A
Nabholz, Benoit: 308C
Nachman, Michael: 153A, 293A
Nadachowska-Brzyska, Krystyna: 32D
Nadachowski, Adam: 320C
Nagard, Hervé Le: 13.6
Nagy, István: 8D, 13.8
Naidoo, Thijessen: 327B
Nakagawa, Shinichi: 349A, 571B
Nakagawa, So: 420B, 429C
Nakagome, Shigeki: 79C, 341B
Nakhleh, Luay: 2.3
Nam, Kiwoong: 10.6, 23.11
Namasivayam, Sivaranjani: 291C
Naor, Adit: 8.1
Narang, Pooja: 17.8
Naranjo, Miguel A.: 622A
Naranjo, Santiago: 22.21
Narasimhan, Vagheesh: 23.9
Narechania, Apurva: 201D
Narum, Shawn: 16.4
Natali, Lucia: 489B
Nater, Alexander: 221D
Natri, Heini: 351C
Naumenko, Sergey: 157A
Navarro, Arcadi: 245A, 249A, 540C
Necsulea, Anamaria: 15.3
Neher, Richard: 18.4
Neiman, Maurine: 334C
Nelson-Sathi, Shijulal: 8.6, 180C, 730A, 733D
Neme, Rafik: 7.5, 7.7, 238B
Nene, Nuno: 346C
Nesme, Xavier: 8.5
Neuditschko, Markus: 400B
Neuffer, Barbara: 10.13
Neuhaus, Klaus: 653C
Neuhauss, Stephan C.F.: 248D
Neumann, Peter: 400B
Neupane, Maniraj: 23.1
Neuvéglise, Cécile: 21.2, 21.3
Ng, Chen Siang: 80D
Ngernsiri, Lertluk: 641A
Nichols, Richard: 56D
Nicholson, William: 299B
Nicol, Graeme: 28.1
Nicolas, Galtier: 6.1
Nieberding, Caroline: 387A
Nielsen, Rasmus: 5.6, 10.9, 16.7, 38B, 103D, 128D, 275C, 305D, 477B, 483D, 533D
Nielsen, Vivi Hunnicke: 611B
Nielsen, William: 228B
Nieves-Colon, Maria A.: 307B
Nimon, Kim: 88A
Ninova, Maria: 22.19
Nisbet, Ellen: 664B
Nishihara, Hidenori: 194A
Nishiumi, Isao: 569D
Nivina, Aleksandra: 65A
Nnamani, Mauris: 20.15
Noel, Benjamin: 515C
Noel, Joseph P.: 263C
Noguera, Isaac: 176A, 558A
Nogué, Fabien: 18.2
Noh, Hyun Ji: 631B
Nolte, Arne: 3.13, 378D, 403A
Nolte, Viola: 3.14, 6.14, 22.24, 119C, 121A, 158B, 159C, 332A
Noonan, James: 4.20, 230D
Norberg, Peter: 434D
Nordborg, Magnus: 18B, 22.15, 85A, 133A, 161B, 163D, 406D, 410D
Norris, Stephen: 468D
Nosenko, Tetyana: 675A
Notebaart, Richard: 20.2, 270B, 703A
Notebart, Richard A: 4.16
Notredame, Cedric: 203B

Nourmohammad, Armita: 439A
Novak, Sebastian: 331D
Novelletto, Andrea: 101B
Novembre, John: 1.9, 11.15, 11.2, 23.1, 25.7, 348A
Novikova, Polina: 133A
Novák, Petr: 10.14
Nowick, Katja: 22.20
Nozawa, Masafumi: 22.6
Nunes, Maria Daniela Santos: 17.7, 22.10
Nunes, Vanessa: 271C
Nussberger, Béatrice: 11.11
Nuzhdin, Sergey: 28D, L 8D
Nye, Tom M. W.: 28.2
Nyerges, Ákos: 8D, 573D
Nygaard, Sanne: 5.21
O'Brien, Katherine: 1.5
O'Donnell, Allyson: 20.9
O'Donnell, Kerry: 21.1
O'Meara, Brian: 205D
O'Neill, Mary: 6.11
O'Neill, Michael: 3.10
O'Neill, Rachel: 3.10
O'Reilly, Paul F: 176A
Obada, Theodor: 320C
Obbard, Darren: 421C
Ober, Dietrich: 11C, 259C, 547B
Oberstaller, Jenna: 291C
Obornik, Miroslav: 596C, 613D
Ochman, Howard: 4.4
Oda, Shoji: 79C
Odenthal-Hesse, Linda: 328A
Ogawa, Motoyuki: 341B
Ogutcen, Ezgi: 10.15
Oh, Jeongsu: 453A
Ohkita, Masashi: 524C
Ohmura, Yoshihito: 511C
Ohta, Toshio: 524C
Oldroyd, Benjamin: 433C
Oleksyk, Taras: 356D, 412B, 628C, 724B
Oleksyk, Taras K.: 413C, 626A, 677C
Oliva, Meritxell: 5.19
Oliveira, Claudio: 51C
Oliveira, Cristina Mendes de: 437C
Oliver, Stephen: 19C
Oliveri, Paola: 12D
Olofsson, Tobias: 451C
Olsen, Kenneth: 141A
Olson, Bradley J.S.C.: 7.5, 27.2
Ometto, Lino: 551B
Omoto, Celso: 655A
Onara, Dalia: 52D
Onkamo, Päivi: 469A
Oota, Hiroki: 79C, 341B, 594A
Oppold, Ann-Marie: 150B
Orek, Yesim Ak: 630A
Oren, Yaara: 740C
Orengo, Dorcas J.: 578A
Orlando, Ludovic: 23.14, 25.2, 103D, 301D
Osada, Naoki: 580C
Osborne, Edward: 22.15
Otto, Thomas D.: 596C
Otwinowski, Jakub: 439A
Oudart, Anne: 182A, 185D
Ouyang, Yidan: 640D
Oxley, John: 317D
Ozkurt, Ezgi: 71C, 692B
Öztürk, Ayaka: 11.22
Öztürk, Bayram: 11.22
O'Dushlaine, Colm: 631B
Parejo, Melanie: 400B
Paajanen, Pirita: 11.12
Paape, Timothy, 389C
Pabijan, Maciej: 171D, 197D
Pabón-Mora, Natalia: 583B
Paces, Jan: 431A
Pagani, Luca: 11.16, 16.9
Paijmans, Johanna: 41A
Pain, Arnab: 596C
Paixão, Tiago: 9A, 22.3, 62B
Pal, Csaba: 4.16, 20.8
Pal, Ferenc: 4.16
Palacios, Julia: 545D
Palero, Ferran: 446B
Pallares, Luisa F.: 69A
Palmer, Abraham A.: 69A
Palmer, Michael: 22.21
Palmer, Sarah: 299B
Palmieri, Nicola: 361A, 456D
Pamjav, Horolma: 101B
Pampoulie, Christophe: 306A
Pan, Chia-Lin: 637A
Pan, Dunfei: 145A
Panchal, Mahesh: 455C, 22.1
Panda, Oishika: 247C
Pang, TY: 20.13
Pannell, John: 156D
Pantano, Lorena: 5.19
Panyon, Larry: 6.10
Panziera, Alex: 618A
Pao, Yueh-Yang: 459C
Papaceit, Montserrat: 578A
Papaevangelou, Vassiliki: 434D
Papandroulakis, Nikos: 369D
Papetti, Chiara: 53A
Papp, Balázs: 2B, 4.16, 8D, 13.8, 20.2, 270B, 573D, 679A, 689C, 703A, 715A
Parat, Florence: 295B
Paris, Margot: 5.1, 5.4, 10.1, 397C
Park, Hyun: 233C, 690D
Park, Min-Kyu: 453A
Park, Soomin: 147C
Parker, Clarissa C.: 69A
Parmley, Joanna: 22.16
Parra, Genís: 23.8, 362A
Parsch, John: 22.9
Partha, Raghav: 20.9
Pascual, Marta: 146B
Pass, Günther: 287C
Passamonti, Marco: 244D
Passet, Virginie: 746A
Patané, José S. L.: 75C
Patel, Simit: 10.11, 150B
Paterno, Marta: 53A
Paterson, Steve: 8.2, 8.9
Pathmanathan, Jananan: 466B
Patin, Etienne: 126B
Patterson, Murray: 7C
Patterson, Nick: 303B
Patz, Edward: 502D
Paul, Reeba: 426D
Paulouskaya, Volha: 487D
Paun, Ovidiu: 132D
Pavard, Samuel: 109B
Pavlicev, Mihaela: 70B, 491D
Pavlidis, Pavlos: 304C
Pavlova, Alexandra: 16.4
Pawliczka, Iwona: 11.22
Payne, Joshua L.: 22.2
Payne, Pavel: 471C
Pearson, Gareth: 17.10
Pearson, Neil: 299B
Pease, James: 10.2
Peccoud, Jean: 3.3
Pecnerová, Patrícia: 313D
Pedersen, Casper-Emil Tingskov: 377C
Pedersen, Oluf: 5.6
Peer, Yves Van de: 13A, 19C, 223B, 274B, 311B
Pegueroles, Cinta: 146B
Peisajovich, Sergio G.: 25A
Peischl, Stephan: 11.1, 97B
Peletto, Simone: 436B
Pelletier, Eric: 515C
Penel, Simon: 199B,
Peng, Min-Sheng: 33A
Peng, Yi: 125A
Penin, Alexey: 157A
Pennings, Pleuni: 18.7
Penny, David: 2.14, 5.24, 570A
Penugonda, Sudhir: 562A
Peona, Valentina: 493B
Pepperell, Caitlin: 6.11
Perea-Atienza, Elena: 267C
Pereira, Alexandre C: 37A
Peres, Amelie: 229C
Perez, Raquel Garcia: 5.10
Perez-Losada, Marcos: 463C
Perez-Vicente, Luis: 152D
Peris, David: 21.9
Peris-Bondia, Francesc: 734A
Perlin, Michael H.: 505A, 506B
Perrin, Nicolas: 374A
Perrière, Guy: 5A
Perry, George: F.4, 95D
Perwitasari-Farajallah, Dyah: 114C, 402D
Pesando, Igor: 695A
Pestle, William J.: 307B
Peter, Arndt: 2.15

Peter, Benjamin: 11.15
Peter, Jackson: 6.15
Peters, Joris: 300C
Peterson, Brant K: 167D
Petit, Elsa: 505A
Petit, Natalia: 3.5
Petitjean, Celine: 28.4
Petkova, Desislava: 11.2
Petrov, Dmitri A.: Pl.3, 1.5, 5.15, 6.6, 13.2, 18.7, 23.10, 336A, 472D, 482C
Petryk, Nataliya: 553D
Pett, Walker: 206A
Peucker, Amanda: 519B
Peyrégne, Stéphane: 25.8
Pfeifer, Susanne: 6.9, 167D
Pfenninger, Markus: 10.11, 150B
Pfister, Luz-Andrea: 543B
Pflieger, David: 6.15
Phadnis, Nitin: 365D
Phifer-Rixey, Megan: 153A, 293A
Philippe, Hervé: 710D
Philippon, Heloise: 5A
Phillips, Barret C.: 353A
Phillips, Karl: 91D
Phunngam, Prin: 577D
Picard, Marion: 26.3
Pichler, Arthur: 603B
Pick, Christian: 264D
Pico, Xavier: 11.19
Piganeau, Gwenael: 61A
Pimenoff, Ville: 437C
Pineda, Laura: 146B
Pineiro, Rosalia: 411A
Pinto, Paulo: 4.17
Pinto, Pedro Vaz: 544C
Piontkivska, Helen: 18.10, 426D
Piou, Cyril: 723A
Piper, Julia: 68D
Piquet, Jésaelle: 698D
Pires, Ana Elisabete: 326A
Piry, Sylvain: 116A
Pisani, Davide: 210A, 222A, 551B
Pissinatti, Alcides: 606A
Plachetzki, David: 559B
Planet, Paul: 201D
Plantamp, Christophe: 723A
Plath, Martin: 10.11
Platts, Adrian E.: 10.13
Plessis, Louis du: 13.5
Plomion, Christophe: 301D
Plotkin, Joshua: 5.22, 439A
Poelstra, Jelmer: 399A
Pointer, Marie: 26.5
Pokalyuk, Cornelia: 425C
Pokusaeva, Victoria: 67C
Poland, Jesse: 122B
Polechová, Jitka: 11.9
Poletti, Giovanni: 11.13
Polte, Christine: 653C
Polz, Martin: 468D
Pombert, Jean-François: 21.7
Poncet, Valérie: 102C
Pontremoli, Chiara: 315B, 599B, 600C
Pool, John: 12.5, 236B
Pootakham, Wirulda: 442B
Popa, Alexandra: 22.16
Popescu, Elizabeth: 11.12
Popinga, Alex: 423A
Popovic, Danijela: 320C
Porcel, Betina M.: 515C
Posnien, Nico: 81A
Posth, Cosimo: 325D
Poulin, Lucie: 454B
Pouyet, Fanny: 200C
Povolotskaya, Inna: 1.13
Povysil, Gundula: 29A
Powell, Adam: 25.6
Powell, Jeffrey: 422D
Poyatos, Juan F.: 715A
Pozo, Angela del: 15.2
Pozzoli, Uberto: 315B, 599B, 600C
Prachumwat, Anuphap: 615B
Prat, Elisa: 85A
Pratas, Înes: 271C
Prats-Escriche, Maria: 58B
Presgraves, Daven: 5.11
Press, Maximilian: 8.7
Preston, Gail M.: 123C
Priedigkeit, Nolan: 20.9
Prieur, Stéphanie Le: 506B, 508D
Pritchard, Jonathan K.: 5.10, 22.18, 22.8, 24.1, 139C
Pritham, Ellen: 291C
Privman, Eyal: 5.21
Procaccini, Gabriele: 53A
Program, NISC Comparative Sequence: 23.8
Prohaska, Sonja: 189C
Proigione, Alessandro: 7.2
Project, Gambie Genome Diversity: 621D
Project, Great Apes Genome Diversity: 23.11
Project, The Open Tree of Life: 207B
Promerová, Marta: 542A
Prosperi, Jean-Marie: 316C
Prosser, James: 179B
Provero, Paolo: 695A
Prufer, Kay: 39C
Prum, Richard: 4.18
Prunier, Jérôme G.: 88A
Pruvost, Mélanie: 323B
Pruvost, Olivier: 454B
Pryszcz, Leszek: 622A
Przeworski, Molly: 5.8
Prüfer, Kay: 16.2, 25.8, 309D
Pucholt, Pascal: 354B
Puchta, Olga: 66B
Puelles, Maria Luz Fernandez de: 630A
Puerma, Eva M.: 578A
Pugh, Charles: 253A
Puig, Marta: 5.19, 176A
Pupko, Tal: 209D, 576C, 740C
Pusey, Anne: 4.4
Putnam, Patrick: 617D
Pybus, Marc: 6B
Pybus, Oliver G: F.7, 18.1, 18.8
Pál, Csaba: 2B, 8D, 13.8, 20.2, 149A, 270B, 573D, 689C, 703A
Pál, Ferenc: 270B, 573D
Pääbo, Svante: 39C, 309D, 362A
Pérez-Espona, Sílvia: 344A
Périan, Séverine: 8.5
Qi, Furong: 134B
Qi, Qin: 8.3
Qi, Xuebin: 125A
Qian, Wenfeng: 59C, 701C
Qian, Y.: 175D
Qiu, Suo: 359C
Quaglietti, Bastien: 446B
Queitsch, Christine: 8.7
Quesada, Victor: 82B, 665C
Quesneville, Hadi: 482C
Quilodrán, Claudio S.: 11.11
Quince, Christopher: 4.3
Quinn, Liam: 11.3
Quintana-Murci, Lluis: F.4, 126B
R.Araúna, Lara: 47C
Robertson, David: 20.1
Rab, Petr: 385C
Rabbers, Iraes: 4.19
Racimo, Fernando: 5.5, 5.6, 16.7
Rada, Petr: 255C
Radwan, Jacek: 6.5
Raff, Jennifer A.: 55C
Ragan, Mark A.: 202A
Raghavan, Rahul: 743B
Raghwani, Jayna: F.7, 18.8
Rago, Alfredo: 231A
Ragsdale, Aaron: 23.4
Rahman, Reazur: 3.6
Raices, Julia: 659A
Rainey, Paul: 444D
Raj, Srilakshmi: 137A
Rakic, Pasko: 230D
Ralph, Peter: F.1, 28, D, 34B
Ram, Asha: 4.3
Ramachandran, Sohini: 17.9, 23.13, 170C, 545D
Ramasamy, Sukanya: 551B
Rambaut, Andrew: F.7, 18.3
Ramirez, Oscar: 581D
Ramstetter, Monica: 92A
Ramírez-Aponte, Edwin: 356D
Ranciaro, Alessia: 87C
Randall, Tori: 312C

Rangavittal, Samarth: 26.8
Ranwez, Vincent: 339D
Rasko, Tamas: 7.2
Rasmussen, David: 18.6
Rasteiro, Rita: 111D
Ratan, Aakrosh: 95D
Rausch, Philipp: 4.2
Rauscher, Martina: 140D
Ravindran, Suda: 673C
Rawlence, Nicolas: 519B
Raynes, Yevgeniy: 1.4
Reboul, Guillaume: 181D
Recknagel, Hans: 582A
Recuenco, Sergio: 90C
Reddy, Aparna: 23.5, 105B
Redelings, Benjamin: 9.2
Reeves, Matthew B: 18.9
Regoes, Roland: 424B
Reich, David: 49A, 303B, 607B
Reid, Jack: 6.4
Reilly, Steven: 4.20, 230D
Reimegård, Johan: 22.25
Reis, Mario dos: 520C
Reis, Micael: 81A
Reis-Cunha, João Luis: 667A
Reisser, Céline: 94C
Remigereau, Marie-Stanislas: 22.15
Remnant, Emily: 417C, 433C
Renaud, Gabriel: 5.5, 309D
Renzette, Nicholas: 425C
Repar, Jelena: 445A
Ressayre, Adrienne: 672B
Restoux, Gwendal: 4.9
Reusch, Thorsten: 22.1
Revadi, Santosh: 551B
Rey, Carine: 554A
Reyes-Oliveras, Alfredo: 356D
Reynolds, Austin: 579B
Reynolds, Louise: 135C
Rhee, Soo-Yon: 18.7
Rheenen, Wouter van: 317D
Rhodes, Dylan: 20.12
Rhule, Emma: 747B
Riccardi, David: 22.17
Ricci, Marco: 493B
Rice, Daniel P.: 1.16
Richard, Durbin: 23.9
Richard, Hugues: 15.4
Richard, Trembath: 23.9
Richardson, David: 91D
Richardson, Mark: 6.4
Ridush, Bogdan: 320C
Rieckh, Georg: 9A
Riehle, Michelle M.: 10.9
Rienzo, Anna Di: 23.1
Rieux, Adrien: 151C
Rimm, David: 14.2
Ringbauer, Harald: 11.6
Ris, Nicolas: 446B
Ritacco, Viviana: 151C
Ritchie, Andrew: 195B
Rivera-Pagán, Ingrid: 356D
Rivera-Rivera, Carlos J.: 2.8
Rivera-Álvarez, Gustavo: 356D
Riveros-Rosas, Hector: 717C
Robay, Amal: 92A
Roberson, Robert: 21.1
Roberts, Nicholas: 222A
Roberts, Rosemary: 415A
Robertson, David: 16D, 430D
Robillard, Emilie: 478C
Robin, Charles: 284D
Robinson, Mark: 722D
Robinson-Rechavi, Marc: 3C, 198A, 685C, 718D
Roblickova, Martina: 320C
Rocchi, Mariano: 352D
Rocha, Eduardo P: 746A
Rocha, Eduardo PC: 8.11
Rocha, Joana L.: 544C
Rockett, Kirk: 621D
Rockman, Matthew: 560C
Rode, Nicolas: 339D
Rodrigo, Allen: 458B, 502D
Rodrigue, Nicolas: 710D
Rodrigues, Maíra R: 37A, 112A
Rodrigues, Nicolas: 374A
Rodrigues-Luiz, Gabriela: 667A
Rodriguez, Jose Manuel: 15.2
Rodriguez, Juan: 15.2
Rodriguez-Flores, Juan L.: 92A
Rodriguez-Garcia, Priscila Marie: 626A
Rodriguez-Varela, Ricardo: 327B
Rodriguez-Verdugo, Alejandra: 684B
Roettger, Mayo: 730A
Roger, Andrew J.: 729D
Rogers, David: 403A
Rogers, Rebekah: 242B
Rohlfs, Rori: 71C
Rollins, Lee: 6.4
Romagné, Frédéric: 25.3
Romero, Maria Lorenzo: 132D
Romiguier, Jonathan: 308C
Rommel, Sophie: 676B
Ronen, Roy: 6.2
Ronfort, Joëlle: 316C
Ronshaugen, Matthew: 22.19
Roos, Anna: 11.22
Ropars, Jeanne: 21.5, 508D
Roquis, David: 26.3
Rosa, Ricardo Valentin-de la: 724B
Rose, Laura: 23.5, 105B, 676B
Rose, Rebecca: 18.8
Rosenberg, Noah: 6.2, 89B, 217D
Rosenberg, Noah A: 1.7
Rosenblum, Erica B.: 6.9
Rosengrave, Patrice: 571B
Rosenkranz, David: 694D
Rosenstiel, Philip: 467C
Rosenthal, Gil: 395A
Rosikiewicz, Wojciech: 234D
Roskams, Steve: 317D
Ross-Ibarra, Jeffrey: 585D
Rossum, Fabienne Van: 401C
Rostant, Wayne: 635B
Rota-Stabelli, Omar: 551B, 618A
Roth, Morgane: 397C
Rothhammer, Francisco: 11.13
Rouhana, Jessy: 371B
Roussel, Véronique: 152D
Roux, Julien: 5.10
Rovira, Quirze: 484A
Rowland-Jones, Sarah: 415A, 418D
Roy, Alexandra-Sophie: 467C
RoyChoudhury, Arindam: 2.4
Rubino, Francesco: 452D
Ruempler, Florian: 273A
Ruhl, Stefan: 304C
Ruis, Chris: 427A
Ruiz, Dannise: 4.8
Ruiz-Herrera, Aurora: 394D
Ruiz-Linares, Andres: 11.13
Ruiz-Orera, Jorge: 7.1, 268D
Ruiz-Trillo, Iñaki: 187A, 190D
Ruppin, Eytan: 270B
Russell, Colin: F.7
Russell, Shelbi: 275C
Russo, Claudia: 215B
Ryan, Hannah: 30B
Ryder, Oliver: 26.8
Rätsch, Gunner: 22.15
Rödder, Dennis: 107D
Rössler, Wolfgang: 21A
Röttger, Mayo: 8.6
Rúa, Concepción de la: 682D
Rühli, Frank J.: 324C
Sabater-Muñoz, Beatriz: 58B
Sabath, Niv: 254B
Sabidó, Eduard: 268D
Sabol, Martin: 320C
Sadowska, Edyta T.: 6.5
Saghaï, Aurélien: 464D
Saglam, Ismail K.: 516D
Saglican, Ekin: 71C
Sagnon, N'Fale: 10.9
Saito, Claire: 524C
Saito, Shigeru: 524C
Saitou, Marie: 492A
Saitou, Naruya: 5.14, 340A, 392B, 595B, 658D
Sajantila, Antti: 101B
Salamin, Nicolas: 10.1, 198A
Salazar, Isaac: 74B
Salcedo, Adriana: 10.13
Saleh, Dounia: 382D
Salimpour, Fahimeh: 522A
Salit, Jacqueline: 92A
Sallet, Erika: 21.5
Salman-Minkov, Ayelet: 254B
Salminen, Iiro: 5.21
Salo, Kati: 469A

Salvador, Irepan: 74B
Salzano, Francisco M: 174C
Salzburger, Walter: 241A
Samerotte, Kerry: 13.2
Samonte, Irene: 455C, 22.1
Sampaio, Iracilda: 606A
Sampaio, José Paulo: 21.9
Sampere, Gabriel: 49A
Sanchez, Rebeca Campos: 26.8
Sanchez-Buso, Leonor: 745D
Sanchez-Ferandin, Sophie: 61A
Sandhu, Manjinder: 607B
Sandoval, Alex Mas: 174C
Sandoval, Marcela: 539B
Sangsrakru, Duangjai: 442B
Sangthong, Pradit: 592C
Sanguinetti, Guido: 66B
Sankaranarayanan, Krithi: 441A
Santi, Lisa De: 46B
Santoni, Sylvain: 6.1
Santos, Manuel: 149A
Santos, Miguel: 51C, 277A, 280D
Santos, Paulo: 665C
Santpere, Gabriel: 540C
Sanz, David Macaya: 5.1
Sarallah, Rezazadeh: 3.1
Sargent, Daniel J: 148D
Sarikas, Srdjan: 1.11
Sarilar, Véronique: 21.2
Sarkissian, Clio Der: 23.14
Sarropoulou, Elena: 711A
Sarver, Brice: 213D
Sasaki, Eriko: 18B
Sasaki, Minami: 16.4
Satayavivad, Jutamaad: 565D
Sato, Chihiro: 670D
Sato, Kimitoshi: 341B
Satomura, Kazuhiro: 357A
Satta, Yoko: 78B, 154B, 232B, 511C, 569D, 660B, 670D, 700B
Satterfield, Dara: 90C
Sattmann, Helmut: 115D, 603B, 610A
Saudemont, Baptiste: 22.16
Savisaar, Rosina: 645A
Savolainen, Outi: 108A, 593D
Savolainen, Peter: 131C
Saw, Jimmy: 28.5
Saw, Jimmy H.: 181D, 473A
Sawaya, Sterling: 24.5
Sawyer, Susanna: 309D
Sayer, Duncan: 11.12
Sayres, Melissa Wilson: 17.8, 26.7
Schaack, Sarah: 227A
Schacherer, Joseph: 6.15, 21.10
Schaeffer, Steve: 365D
Schaerfe, Charlotta: 24D
Schaffer, Willam: 539B
Scharf, Jeremiah M: 631B
Schartl, Manfred: 10.11, 349A
Schaub, Florentine: 532C
Schbath, Sophie: 2.15
Scheib, Christiana: 312C
Schempp, Werner: 352D
Scherer, Siegfried: 653C
Scherling, Christian: 653C
Schick, Alana: 13.9
Schield, Drew: 129A
Schienfeldt, Laura: 87C
Schierup, Mikkel Heide: 10.6, 23.11, 43C, 175D
Schiffels, Stephan: 11.12
Schiffer, Philipp: 250B
Schiffman, Joshua: 28D
Schiller, Thomas: 563B
Schiöth, Helgi: 191A
Schleper, Christa: 28.5, 184C
Schlessinger, Joseph: 14.2
Schlupp, Ingo: 674D
Schlötterer, Christian: 3.14, 6.14, 17.7, 22.24, 119C, 121A, 158B, 159C, 332A, 361A, 476A, 638B, L 1A
Schlüter, Hartmut: 264D
Schlüter, Philipp: 10.16
Schmid, Karl J.: 382D
Schmidt, Anja: 5.4
Schmidt, Chloé: 11.21
Schmidt, Heiko: 2.5
Schmidt, Joshua: 284D
Schmidt, Paul: 1.5, 11.21
Schmidt, Ryan: 594A
Schmidt-Kopplin, Philippe: 653C
Schmitt, Imke: 514B
Schmitz, Jonathan: 188B
Schnabel, Robert: 23.7
Schneeweiss, Gerald M.: 5.17, 550A
Schneider, Daniela: 4.7
Schneider, Harald: 56D
Schneider, Horacio: 606A
Schorpp, Michael: 518A
Schrago, Carlos: 627B
Schraiber, Joshua: 6.12, 22.23
Schrempf, Dominik: 2.2
Schrider, Daniel: 252D
Schroeder, Frank: 247C
Schroeder, Hannes: 441A, 539B
Schuenemann, Verena J.: 319B, 321D, 469A
Schuit, Frans: 588C
Schultz, Thia: 521D
Schulz, Frederik: 457A
Schulze, Jana: 264D
Schumacher, Julia: 20D
Schumann, Gerald: 7.2
Schumer, Molly: 395A
Schuster, Stephan C.: 95D
Schutten, Martin: 18.9
Schwager, Evelyn E.: 22.19
Schwartz, Rachel: 2.11, 36D, 636C
Schönfeld, Joachim: 467C
Schönheit, Peter: 8.6
Schönswetter, Peter: 132D
Scliar, Marilia: 591B
Scofield, Douglas: 22.25
Scofield, Douglas G.: 10.13
Scofield, Paul: 519B
Scott, Jeffrey: 17.5
Scotti, Ivan: 118B
Sedeek, Khalid: 10.16
Sedghifar, Alisa: 34B
Sedlazeck, Fritz: 378D
Seehausen, Ole: 376B
Sefton, Maggie: 713C
Segarra, Carmen: 578A
Seger, Jon: 344A
Seguin-Orlando, Andaine: 23.14, 301D
Sehr, Eva Maria: 720B
Seixas, Susana: 82B, 290B, 665C
Sela, Itamar: 209D
Seluanov, Andrei: 3.1
Selvatti, Alexandre Pedro: 215B
Senior, Alistair: 349A
Sepil, Irem: 569D
Serohijos, Adrian: 738A
Serrana, Daniel Garcia de la: 711A
Serra-Vidal, Gerard: 49A
Serrão, Ester: 17.10
Servin, Bertrand: 9.5
Seto, Yosuke: 407A
Setter, Derek: 16.6, 50B, 168A
Severac, Dany: 6.1
Sevilla, Arrate: 682D
Shafer, Robert: 18.7
Shah, Premal: 5.22
Shakhnovich, Eugene: 738A
Shanklin, John: 10.16
Shao, Yi: 7.9, 276D
Sharbrough, Joel: 334C
Sharp, Andrew: 408B, 722D
Shaw, Liz: 525D
Shaw, Neethu: 583B
Sheikh, Sanea: 192B, 193C
Sheinman, Michael: 2.15, 2.16
Shen, Hao: 7.9, 276D
Shen, Quan-Kuan: 33A
Sheridan, Isabelle: 18.8
Sherlock, Gavin: 6.6, 13.2
Shi, Lei: 363B
Shih, Chun-Chieh: 686D
Shikano, Takahito: 351C
Shimizu, Kentaro K.: 156D, 389C
Shimode, Sayumi: 420B
Shin, Paul: 714D
Shine, Richard: 536C
Shine, Rick: 6.4
Shiott, Morten: 5.21
Shirsekar, Gautam: 321D

Shringarpure, Suyash: 17.9, 541D
Shunkov, Michael: 309D
Siebers, Bettina: 8.6
Siebert, Ursula: 11.22
Siegal, Mark: 20.7
Siegismund, Hans R.: 175D, 377C, 544C
Siepel, Adam: 19.3, 22.1, 384B
Sievert, Christian: 6.10, 11C
Signor, Sarah: 12.2
Siguenza, Sophie: 506B
Sigwalt, Anastasie: 6.15
Sikora, Martin: 716B
Sikosek, Tobias: 564C
Silva, Corinne Da: 515C
Silva, Nuno M: 11.10
Silva, Thushan De: 415A
Simon, Marianne: 4.9
Simon, Svenja: 653C
Simonti, Corinne: 16.3, 491D
Simões, Luciana: 326A
Sing, Charles: 137A
Singh, Gajinder Pal: 13.8
Singh, Manvendra: 7.2
Singh, Pooja: 405C
Siokou, Ioanna: 630A
Sirica, Roberto: 707A
Sironi, Manuela: 315B, 599B, 600C
Sisay-Joof, Fatou: 621D
Siska, Veronika: 169B
Siu-Ting, Karen: 210A
Skage, Morten: 306A
Skoglund, Pontus: 327B
Skov, L.: 175D
Skóra, Krzysztof: 11.22
Slate, Jon: 523B
Slatkin, Montgomery: 5.5
Sleator, Roy: 5.18
Sliwerska, Elzbieta: 11.14
Slodkowicz, Greg: 632C
Slotte, Tanja: 10.13, 22.25
Smet, Riet De: 274B
Smith, Bruce D.: 35C
Smith, Ian: 519B
Smith, J. Andrew C.: 123C
Smith, Justin: 22.21
Smith, Matthew: 21.1
Smith, Oliver: 299B
Smith, Rebecca: 284D
Smith-Zubiaga, Isabel: 682D
Sniegowski, Paul: 1.4, 288D
Snirc, Alodie: 506B, 508D
Soares-Souza, Giordano B: 112A
Sobkowiak, Benjamin: 151C
Sobotka, Roman: 613D
Sochorová, Jana: 526A
Soejima, Hidenobu: 341B
Soi, Sameer: 87C
Sojo, Victor: 731B
Sokhatsky, Mykhailo: 594A
Solferini, Vera: 602A
Solis-Lemus, Claudia: 375A
Solé-Morata, Neus: 110C
Somel, Mehmet: 71C, 612C, 692B
Sommer, Ralf: 247C
Sommerlandt, Frank: 21A
Sonay, Tugce: 24.2, 722D
Song, Giltae: 507C
Song, Qiang: 22.15
Song, Shiya: 11.14
Song, Shuhui: 642B
Song, Yun: 1.8, 13.4, 164A, 212C
Soodyall, Himla: 49A
Sorger, Thomas: 714D
Soria, Andrea Gloria: 587B
Soucy, Shannon: 186A
Sousa, Filipa L.: 8.6, 180C, 730A, 733D
Sousa, Tânia: 665C
Sousa, Vitor C: 167D
Souto, Laura Do: 515C
Souza, Alexandra de: 616C
Spaethe, Johannes: 21A, 22B
Spang, Anja: 28.5, 181D, 473A
Spano, Dario: 296C
Spatafora, Joseph: 21.1
Spataro, Nino: 343D
Spence, Jeffrey: 212C
Spencer, Chris: 621D
Sperduti, Alessandra: 318A
Sperling, Linda: 22.16
Spielman, Stephanie: 5.13
Spielmann, Tobias: 264D
Spiers, Andrew: 8.2, 8.9
Spohn, Réka: 4.16, 8D, 13.8, 270B
Springs, Lauren: 579B
Sprouffske, Kathleen: 288D
Spurgin, Lewis: 91D
Spyrou, Maria A.: 318A
Squire, Mattie: 395A
Sriboonlert, Ajaraporn: 575B
Srivastava, Dhruv: 549D
Staab, Paul: 23.5
Stadler, Peter: 189C
Stadler, Tanja: 5.12, 423A
Stahl, Aaron: 72D
Stahl, Bethany: 668B
Stairs, Courtney W.: 729D
Stajich, Jason: 21.1
Stalker, Jim: 621D
Stamatakis, Alexandros: 219B
Stankovic, Anna: 320C
Stanton, Kimmy: 661C
Star, Bastiaan: 306A, 597D
Staubach, Fabian: 472D
Staudt, Michelle R.: 92A
Stefaniak, Krzysztof: 320C
Stegele, Oliver: 22.15
Steige, Kim: 10.13, 22.25
Stenseth, Nils Chr.: 306A
Stepanauskas, Ramunas: 457A
Stephan, Wolfgang: 329B, 675A
Stephens, Matthew: 11.2, 562A
Sterck, Lieven: 223B
Stern, David: 17.7
Stertz, Florian: 2.5
Stewart, S Evelyn: 631B
Stinchcombe, John R.: 22.26, 321D
Stock, Jay: 312C
Stockhammer, Philipp: 608C
Stoelting, Kai N.: 391A
Stoll, Monika: 22.1
Stolle, Eckart: 3.12
Stone, Anne: 25.1, 307B, 543B, 636C
Stoneking, Mark: 25.8
Storz, Jay: 60D
Storå, Jan: 327B
Strasser, Bettina: 251C
Straub, Christina: 444D
Strauli, Nicolas: 555B
Streicker, Daniel: 90C
Streiff, Réjane: 723A
Streito, Jean-Claude: 446B
Struck, Travis: 530A
Stucken, Karina: 23C, 83C, 742A
Stuessy, Tod F.: 10.14, 550A
Stuglik, Michal: 171D, 197D
Stukenbrock, Eva H.: 21.8
Sturmbauer, Christian: 405C
Städler, Thomas: 5.4, 397C
Stölting, Jan: 255C
Stölting, Kai N.: 5.1, 10.1
Su, Bing: 125A, 363B
Su, Guosheng: 145A
Su, Marcia Shu-Wei: 24.4
Su, Zhixi: 500B
Subbotin, Vladimir: 73A
Suchard, Marc A: 18.8
Suciu, Radu: 52D
Suesdek, Lincoln: 75C, 589D
Suga, Hiroshi: 187A
Sugden, Lauren A.: 23.13
Sui, Zhihai: 735B
Sukumaran, Jeet: 458B
Sun, Chen: 24.4
Sun, Cheng: 291C
Sun, Genlou: 40D
Sun, Qiushi: 701C
Sun, Yuqiang: 704B
Sunnucks, Paul: 16.4
Sunyaev, Shamil: 2.13, 14.3
Supek, Fran: 445A
Surat, Wunrada: 442B, 592C
Surbeck, Martin: 362A
Sutherland-Smith, Andrew: 2.14
Sutrina, Sarah: 475C
Suwandi, Abdulhadi: 4.2
Suzuki, Taichi: 153A
Suzuki, Yutaka: 234D
Suárez, Alejandro Morales:

156D
Svardal, Hannes: 161B
Svedberg, Jesper: 21.4
Svensson, Emma: 326A, 327B
Sverrisdóttir, Óddny: 327B
Swaggart, Kayleigh: 491D
Swallow, Dallas: 693C
Swanson, Willie: 5.23, 22.5
Swatdipong, Akarapong: 565D
Swofford, Ross: 631B
Sy, Ousmane: 698D
Sylvester, Kayla: 21.9
Symonova, Radka: 385C
Syntnikova, Yuliya: 3.6
Szamecz, Béla: 149A
Szappanos, Balázs: 4.16, 2B, 8D, 20.2, 270B, 679A
Szczesniak, Michal: 281A, 687A
Szitenberg, Amir: 383A
Szkotak, Rafal: 171D
Szvetnik, Attila: 7.2
Szyszka, Taylor: 4.11
Szöllősi, Gergely J: 503A
Szöllősi, Gergely János: 737D
Sánchez-Guillén, Rosa Ana: 394D
Sánchez-Maldonado, Stephanie: 356D
Sémon, Marie: 554A
Sørensen, Jesper: 6.1
Tenaillon, Olivier: 470B
Tachezy, Jan: 255C
Tafforeau, Paul: 319B
Taib, Najwa: 182A
Takacs-Vesbach, Cristina: 181D
Takahashi-Kariyazono, Shiho: 232B
Takeuchi, Nobuto: 609D
Takezaki, Naoko: 194A
Takuno, Shohei: 3.16, 398D
Tall, Hamidou: 698D
Tamarit, Daniel: 451C
Tamura, Koichiro: 216C, 357A, 407A
Tan, Li Ying: 22.14
Tanaka, Kentaro: 17.7
Tandonnet, Sophie: 699A
Tang, Julian: 434D
Tang, Ruqi: 631B
Tanguy, Arnaud: 656B
Tannier, Eric: 302A
Tapanainen, Riikka: 647C
Tarazona, Eduardo: 112A
Tarazona-Santos, Eduardo: 37A, 591B
Tarekegn, Ayele: 11.16
Tasdighian, Setareh: 274B, 311B
Tassi, Francesca: 46B, 106C
Tataru, Paula: 23.3
Taus, Thomas: 158B
Tautz, Diethard: Pl.2, 7.7, 69A, 220C, 238B, 256D, 654D
Tavares, Hugo: 11.18
Tawfik, Dan S.: 263C
Taylor, Thomas: 21.1
Taylor, Tiffany: 14.5
Teasdale, Matthew: 317D
Technau, Ulrich: 27.3
Tedder, Andrew: 156D
Teeter, Katherine C.: 381C
Tegel, Willy: 301D
Teixeira, Vanessa: 272D
Tejedor, Ghildres: 356D
Tellier, Arélien: 294A, 295B
Tenaillon, Maud: 45A
Tenaillon, Olivier: 13.6, 684B
Terai, Yohei: 78B
Terai, Yohey: 232B, 398D, 700B
Terhorst, Jonathan: 13.4, 164A, 212C
Terossi, Mariana: 590A
Terrapon, Nicolas: 532C
Tesler, Glenn: 6.2
Tessmer, Heidi: 435A
Teufel, Ashley I.: 20.4, 243C
Teusink, Bas: 4.19
Theissen, Guenter: 273A
Theriault, Jolan: 10.15
Thermes, Claude: 553D
Theunert, Christoph: 25.8
Thi, Ha My Vy: 6.3
Thiergart, Thorsten: 8.6
Thomas, Mark: 11.16, 25.6, 317D, 693C
Thomas, W. Kelley: 235A
Thompson, Simon: 87C
Thongsaiklaing, Thanaset: 641A
Thorne, Michael: 250B
Thornton, Joseph: F.8
Thorpe, Harry: 335D, 663A
Thorrez, Lieven: 588C
Thuer, Ernst: 622A
Thézé, Julien: 26.9
Tiedemann, Ralph: 11.22, 674D
Tiemann-Boege, Irene: 24.3
Timmermann, Bernd: 373D
Tishkoff, Sarah: 87C, 366A
Tkacik, Gasper: 9A, 22.3
Tlusty, Tsvi: 20.14
To, Thu Hien: 2.6
Tobler, Ray: 119C, 121A
Tobler, Raymond: 22.24, 158B
Toda, Masanori: 407A
Toft, Christina: 58B
Toh, Su San: 506B
Toivainen, Tuomas: 148D
Tokuriki, Nobuhiko: 4.12, 292D
Toll-Riera, Macarena: 4.15, 8.3
Tollervey, David: 66B
Tolun, Aslıhan: 101B
Tomaszkiewicz, Marta: 26.8
Tomcala, Ales: 596C
Tominaga, Makoto: 524C
Tomlinson, Chad: 408B
Tong, K Jun: 2.7
Tonkin, Zeb: 16.4
Topa, Hande: 332A
Torres, Raul: 23.12
Torres, Tatiana: 14B, 697C, 699A
Torres-Hernández, Isamar: 356D
Torres-Sánchez, María: 261A
Toräng, Per: 86B
Touchon, Marie: 746A
Touhami, Meriem Khelifi: 630A
Toups, Melissa: 374A
Tournebize, Rémi: 102C
Touzet, Pascal: 401C
Townsend, Jeffrey: 14.2
Towsend, Andrea: 569D
Tranchevent, Léon-Charles: 650B
Trense, Daronja: 11.22
Tress, Michael: 15.2
Trono, Didier: 652B
Trontti, Kalevi: 647C
Trucchi, Emiliano: 132D
Tsai, Huai-Kuang: 681C
Tsai, Isheng J: 615B
Tsai, Zing Tsung-Yeh: 681C
Tsakogiannis, Alexandros: 369D
Tschachler, Erwin: 251C
Tschopp, Patrick: 26.4
Tsigenopoulos, Costas: 369D
Tsinganis, Markos: 397C
Tsuchimatsu, Takashi: 85A
Tucker, Priscilla K.: 381C
Tugrul, Murat: 22.3
Turan, Zeliha Gözde: 612C
Turchin, Michael: 562A
Turgeman, Israela: 8.1
Turner, George F.: 10.5
Turner, James: 26.4
Turner, Paul: 422D
Tutill, Helena J: 18.9
Tyers, Alexandra: 10.5
Tyler-Smith, Chris: 11.12, 11.16, 352D, 364C, 607B, 707A
Tyrmi, Jaakko: 593D
Uecker, Hildegard: 6.13, 16.6
Uerpmann, Hans-Peter: 300C
Ullastres, Anna: 3.5, 479D
Underhill, Gillian: 434D
Uren, Caitlen: 136D
Ureña, Irene: 327B
Urrutia, Araxi O.: 616C, 706D
Usmanova, Dinara: 1.13
Uyenoyama, Marcy: 9.2
Vigliotti, Chloe: 465A
Vakirlis, Nikolaos: 21.2, 262B
Valderrama, William: 90C
Valdiosera, Cristina: 327B
Valdivia, Hugo: 667A
Valencia, Alfonso: 15.2

Valero-David, Guillermo: 356D
Valiente-Moro, Claire: 3.15
Vallier, Marie: 4.2
Valloton, Delphine: 22.4, 26.4
Valverde, Selene: 7.6
Vamosi, Jana C.: 10.15
Vancaester, Emmelien: 183B
Vanderhaeghen, Sonja: 653C
Vanderpool, Dan: 517A
Vanneste, Kevin: 274B
vanVuuren, Bettine Jansen: 544C
Vartanyan, Sergey: 313D
Vasconcelos, Ana Tereza: 667A
Vaughan, Timothy: 211B, 423A
Vazquez, Jesus: 15.2
Vecchyo, Diego Ortega Del: 1.9, 23.7
Veeramah, Krishna: 366A
Vega, Ricardo C Rodríguez de la: 508D
Vega, Ricardo Rodriguez de la: 506B
Vega, Ricardo Rodríguez de la: 21.5
Vega-Medina, Michelle: 356D
Veidenberg, Andres: 629D
Vekemans, Xavier: 85A
Velasco-Villa, Andres: 90C
Veldink, Jan H: 317D
Velez-Valentin, Jafet: 724B
Velicer, Gregory: 27.5
Veltsos, Paris: 374A
Vendramin, Giovanni Giuseppe: 118B
Venkataram, Sandeep: 6.6, 13.2
Ventura, Jacint: 394D
Vera-Colón, Olga: 356D
Verdier, Valerie: 454B
Verjans, Georges MGM: 18.9
Vermeirssen, Vanessa: 13A
Vernesi, Cristiano: 345B
Vernick, Kenneth D.: 10.9
Vernière, Christian: 454B
Vernot, Benjamin: 16.3, 22.11, 39C
Veselis, Amelia: 328A
Vibranovski, Maria: 659A
Vicente-Salvador, David: 5.19
Vicoso, Beatriz: 17.3
Vieira, Cristina: 3.15, 23.14, 81A, 271C, 272D
Vieira, Filipe G: 533D
Vieira, Jorge: 81A, 271C, 272D
Vienne, Damien de: 203B
Vieth, Beate: 688B
Vigilant, Linda: 362A
Vigouroux, Yves: 102C, 698D
Vijay, Nagarjun: 399A
Viklund, Johan: 457A
Vila, Carles: 3.4
Vilaca, Sibelle Torres: 46B
Vilgalys, Rytas: 21.1
Vilhjálmsson, Bjarni: 22.15, 43C
Viljakainen, Lumi: 5.21
Vilkki, Johanna: 9.5
Villanueva-Cañas, José Luis: 7.1, 246B
Villaseñor, José: 10.14
Villatoro, Sergi: 5.19, 176A
Vilà, Carles: 581D
Viola, Bence: 309D, 319B
Vlasschaert, Caitlyn: 239C, 649A
Vlastaridis, Panayotis: 19C
Vogl, Claus: 172A, 361A
Vohsen, Samuel: 4.8
vonHoldt, Bridget: 724B
Vornin, Victor: 22.15
Vyskot, Boris: 350B
Vásquez, Alejandra: 451C
Vågene, Åshild J.: 318A
Vélez-Ruiz, Edgardo: 356D
Víkingsson, Gísli: 11.22
Vörös, Andrea: 8D
Vøllestad, Asbjørn: 597D
Wade, Juli: 26.4
Wagner, Allon: 270B
Wagner, Andreas: 4.14, 4.15, 22.2, 24.2, 288D, 722D
Wagner, Gunter: 4.18, 20.15
Wagner, Stefanie: 301D
Wakano, Joe: 556C
Wakeley, John: 545D
Walda, Claudia: 742A
Waldron, Fergal: 421C
Wales, Nathan: 35C, 310A
Wallner, Barbara: 361A
Walsh, Paul: 5.18
Walters, James: 367B
Wang, Aiguo: 145A
Wang, Chao: 735B
Wang, Chenghui: 390D
Wang, Daryi: 459C
Wang, Guo-Dong: 131C
Wang, J.: 175D
Wang, Jichang: 7.2
Wang, John: 370A
Wang, Jun: 19.4, 124D, 390D
Wang, Liuyang: 381C
Wang, Rachel: 1.8
Wang, Shengdian: 498D
Wang, Wei: 10.13
Wang, Wen: 19.4, 124D
Wang, Wen-Zhi: 33A
Wang, Xiliang: 642B
Wang, Xu: 5.7
Wang, Xuan: 6.10
Wang, Yufeng: 15C
Wang, Zongji: 5.2
Wanger, Gunter: 696B
Ware, Roselyn: 299B
Warinner, Christina: 441A
Warnefors, Maria: 7.11, 22.4
Warnow, Tandy: 2.1
Warren, Wesley C: 408B
Wass, Anastasia: 14.5
Waters, Jonathan: 519B
Waters, Sinead: 452D
Watsa, Mrinalini: 543B
Watson, Eric T.: 353A
Watson, Simon: F.7
Wayne, Marta: 18.10
Wayne, Robert: 3.4, 23.7
Wayne, Robert K.: 581D
Weber, Claudia C.: 243C
Webster, Claire: 421C
Weeks, Stephen C: L 3C
Weghorn, Donate: 14.3
Wegmann, Daniel: 165B, 295B, 537D
Wegner, Fanny: 438D
Wehnert, Daniel: 563B
Wei, Kevin: 725C
Wei, Xinzhu: 10B
Weigel, Detlef: 11.17
Weihmann, Antje: 23.8
Wein, Tanita: 467C
Weinert, Lucy: 747B
Weinreich, Daniel: 1.4
Weiss, Clemens L.: 321D
Weiss-Gayet, Michele: 23.14
Weiss-Schneeweiss, Hanna: 10.14, 550A
Weissenbach, Julia: 467C
Weissensteiner, Matthias: 399A
Weissman, Daniel: 11.7
Weißenbach, Julia: 23C
Welch, David: 5.12
Welch, John: 747B
Weng, Meng-Pin: 678D
Werely, Cedric: 136D
Werren, John: 5.7
Werren, John H: 231A
Wertheim, Bregje: 371B
Wesseling, Anne-Maria: 547B
West, Joe: 601D
Westall, Portia: F.5
Wetton, Jon: 111D
Whelan, Simon: 243C, 619B
Whibley, Annabel: 388B
White, Mark: 30B
White, Merlin: 21.1
Whiteman, Noah: 138B
Whittle, Edward: 10.16
Wicke, Susann: 5.17
Widmer, Alex: F.3, 5.4
Wiehe, Thomas: 140D, 250B
Wietrzynskib, Wojciech: 549D
Wikander, Johan: 451C
Wikramaratna, Paul: 418D
Wilburn, Damien: 5.23
Wilcox, Alexander: 342C
Wildschutte, Hans: 468D
Wildschutte, Julia: 3.4
Wilke, Claus: 5.13, 419A
Wilkinson, Gerald: 14.4
Wilkinson, Mark: 210A, 261A

Willems, Thomas: 24.1
Willerslev, Eske: 305D, 539B
Willett, Christopher: 671A
Williams, Rachel J: 18.9
Williams, Tom: 28.2, 179B, 732C
Williamson, Robert J.: 10.13
Wilson, Daniel: 11.4
Wilton, Peter: 557D
Wincker, Patrick: 515C
Winney, Bruce: 101B
Winter, David: 6.10
Winternitz, Jamie: 90C
Wirtz, Johannes: 140D
Wischnitzki, Elisabeth: 720B
Witonsky, David: 23.1
Witt, Christopher: 379A
Witte, Hanh: 247C
Witting, Michael: 653C
Wittkopp, Patrica: 22.22
Wittmann, Meike J.: 336A
Wodarz, Dominik: 14.1
Woehle, Christian: 467C
Wolf, Jochen: 399A
Wolf, Yuri: 8.1
Wolfe, Kenneth H.: 21.11
Wolff, Kirsten: 574A
Wolinsky, Steven: 562A
Wollman, Francis-André: 549D
Wong, Simon: 250B
Wongkhamprai, Bhawat: 565D
Wongyai, Wasana: 565D
Wonnapinij, Passorn: 575B, 592C
Woodward, Martin: 525D
Worliczek, Hanna: 456D
Wostrikoff, Katia: 549D
Wragg, David: 400B
Wray, Greg: 712B
Wray, Gregory: 19.2, 228B
Wright, Alison: 26.5
Wright, John: 23.9
Wright, Lori: 5.11
Wright, Stephen: 10.13, 22.26, 152D
Wu, Baojun: 509A
Wu, Chung-I: 131C, 496B, 497C, 498D
Wu, Chung-Shien: 567B
Wu, Dexiang: 40D
Wu, Gideon: 637A
Wu, Hao: 474B
Wu, Ping: 80D
Wu, Shi-Fang: 33A
Wu, Siao-Man: 80D
Wu, Steven: 458B
Wu, Tianyi: 125A
Wu, Xin: 498D
Wu, Yufeng: 477B
Wunder, Jörg: 86B
Wunderlich, Stephanie: 688B
Wurm, Yannick: 5.21
Wylie, Scott: 1.4
Wyman, Minyoung: 5.8
Wötzel, Stefan: 86B
Wüest, Raphael O.: 10.1
Xia, Xuhua: 239C, 649A
Xiang, Hui: 124D
Xiang, Kun: 125A
Xiao, Wenyuan: 291C
Xie, Chao: 450B
Xie, Chen: 238B
Xie, Gangcai: 7.2
Xiong, Peiwen: 279C, 657C
Xu, Chenling: 17.4
Xu, Shuqing: 646B
Xue, Yali: 23.9, 607B
Yahalomi, Dayana: 561D
Yampolsky, Lev: 157A
Yan, Guangmei: 5.2
Yan, Hua: 5.21
Yang, Deying: 125A
Yang, Hsiao-Pei: 19.3
Yang, JianRong: 63C
Yang, Yu-Fei: 701C
Yang, Yuyan: 704B
Yang, Ziheng: 31C, 520C
Yang, Zuyu: 498D
Yassin, Amir: 12.5
Yasui, Yoshie: 341B
Yasukochi, Yoshiki: 660B
Yeaman, Sam: 11.8, 100A
Yeger-Lotem, Esti: 501C
Yin, Jun: 4.20, 230D
Ying, Wei: 5.2
Yohe, Laurel: 548C
Yoon, Ung-Han: 17A
Young, Randee: 365D
Yu, Jun: 474B, 642B
Yu, Michael: 3.6
Yu, Yuen-Tsu Nicco: 27.5
Yu, Yun: 2.3
Yuan, Dave: 22.22
Zhuhu, Tianqi: 520C
Zadik, Daniel: 101B, 352D
Zahradka, Davor: 445A
Zahradka, Ksenija: 445A
Zakov, Shay: 6.2
Zala, Marcello: 21.6
Zalloua, Pierre: 47C, 49A
Zane, Lorenzo: 53A
Zang, Pei: 22.15
Zanini, Fabio: 18.4
Zapater, Marie-Françoise: 152D
Zaremba-Niedzwiedzka, Katarzyna: 28.5, 183B, 457A, 473A
Zarin, Taraneh: 20.5
Zehender, Gianguglielmo: 436B
Zehr, Simonida: 514B
Zekraoui, Leila: 698D
Zemp, Niklaus: F.3
Zeng, Kai: 166C, 523B, 572C
Zeng, Mei: 3.6
Zeng, Qinglong: 458B
Zhai, Weiwei: 131C
Zhang, Chengjun: 640D
Zhang, Ge: 617D
Zhang, Guojie: 5.2, 19.4
Zhang, Hengmu: 704B
Zhang, Jianzhi: 5.9, 10B, 20.11, 59C, 63C, 226A, 702D
Zhang, Liqing: 704B
Zhang, Lu: 19.4
Zhang, Mian: 22.21
Zhang, Pei: 18B
Zhang, Peng: 217D
Zhang, QianQian: 9.5, 611B
Zhang, Qin: 145A
Zhang, Wei: 19.4
Zhang, Xiaoming: 125A
Zhang, Xiaoqing: 701C
Zhang, Xiquan: 134B
Zhang, Ya-Ping: 33A, 131C
Zhang, Yong: 7.9
Zhang, Yong E.: 276D
Zhang, Yonghong: 418D
Zhang, Yuezheng: 497C
Zhang, Zhang: 474B
Zhao, Bixiao: 14.2
Zhao, Li: 19.4
Zhao, Shiming: 500B
Zhao, Zi-Ming: 14.2
Zhou, Jingqi: 500B
Zhou, Qi: 5.2, 360D, 648D
Zhou, Yiqi: 22.21
Zhou, Yongfeng: 108A
Zhou, Zhan: 500B
Zhu, Fangjun: 674D
Zhu, Joe: 173B
Zhu, Leyi: 704B
Zhu, Lifeng: 390D
Zhu, Tianqi: 31C
Zhu, Yuan: 20.7
Zielinski, Piotr: 32D, 171D, 197D
Ziesemer, Kirsten A.: 441A
Zimmer, Fabian: 26.5, 188B
Zimorski, Verena: 255C
Zink, Albert: 450B
Zischler, Hans: 20D, 114C, 402D, 694D
Ziv, Tamar: 549D
Zivanovic, Yvan: 464D
Znidaric, Nina: 405C
Zou, Yangyun: 500B
Zou, Zhengting: 226A
Zufall, Rebecca: 601D
Zuo, Hongwei: 40D
Zöllner, Sebastian: 217D