Please click the embedded links to view this content on the SMBEv2021 YouTube Channel

2021 Graduate Student Symposium:

- Introduction
- Influence of the Last Glacial Period on the genetic diversity of current Southeast Asians - Catarina Branco
- Inferring selective sweep parameters through supervised learning - Ian Vasconcellos Caldas
- Sourcing Elephant Ivory from a Sixteenth-Century Portuguese Shipwreck - Alida de Flamingh
- Recombination and population demographics impact allele-specific expression of deleterious variants in human populations - Michelle P. Harwood
- Pan-mammalian analysis of molecular constraints underlying extended lifespan in mammals - Amanda Kowalczyk
- Historical memory, epistasis, and contingency in long-term protein evolution - Yeonwoo Park
- Sex-specific selection drives the evolution of alternative splicing in birds - Thea F. Rogers
- Preferred synonymous codons are more accurately translated: proteomic evidence, among-species variation, and mechanistic basis - Mengyi Sun

2021 SMBE Presidential Address - Harmit Malik – Constraints and adaptive potential of antiviral proteins

2021 SMBE Business Meeting – Harmit Malik

2021 SMBE Past-President Address - Marta Wayne - Neutral theories

2021 Faculty Awards Symposium

- Early Career Faculty Award Winner Introduction - Harmit Malik
- Early Career Faculty Award Winner - Kelley Harris - Investigating common mutator phenotypes in humans, mice, and yeast
- Mid-Career Faculty Award Winner Introduction - Harmit Malik
- Mid-Career Faculty Award Winner - Tanja Stadler - What do SARS-CoV-2 genomes tell us about pandemic spread?
- Lifetime Achievement Award Winner Introduction - Harmit Malik
- Lifetime Achievement Award Winner - Michael Lynch - Forty years of effective neutrality

OS1-1

The evolutionary consequences of hybridisation for grey wolves and free-ranging domestic dogs

Malgorzata Pilot[1], Andre E. Moura[1], Innokentiy M. Okhlopkov[2], Nikolay V. Mamaev[2], Ninna H. Manaseryan[3], Vahram Hayrapetyan[4], Natia Kopaliani[5], Elena Tsingarska[6], Abdulaziz N. Alagaili[7], Osama B. Mohammed[7], Elaine A. Ostrander[8], Wiesław Bogdanowicz[9]

[1]Museum and Institute of Zoology, Polish Academy of Sciences, Gdansk, Poland. [2]Institute of Biological Problems of Cryolithozone, Siberian Branch of Russian Academy of Sciences, Yakutsk, Russian Federation. [3]Scientific Center of Zoology and Hydroecology, National Academy of Sciences, Yerevan, Armenia. [4]Green Artsakh Biosphere Complex SNCO, Artsakh, Armenia. [5]Institute of Ecology, Ilia State University, Tbilisi, Georgia. [6]BALKANI Wildlife Society, Sofia, Bulgaria. [7]KSU Mammals Research Chair, Department of Zoology, King Saud University, Riyadh, Saudi Arabia. [8]Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [9]Museum and Institute of Zoology, Polish Academy of Sciences, Warszawa, Poland

Abstract

Introgressive hybridisation between domestic animals and their wild relatives is an indirect form of human-induced evolution, altering gene pools and phenotypic traits of wild and domestic populations. Although this process is well documented in many taxa, its evolutionary consequences are poorly understood. In this study, we assess introgression patterns in admixed populations of Eurasian wolves and free-ranging domestic dogs (FRDs), identifying chromosomal regions with significantly overrepresented hybrid ancestry and assessing whether functional genes within these regions show signatures of selection. Although the dog admixture proportion in West Eurasian wolves (2.7%) was greater than the wolf admixture proportion in FRDs (0.75%), the number and average length of chromosomal blocks showing significant overrepresention of hybrid ancestry was smaller in wolves than FRDs. In wolves, 6% of genes located within these blocks showed signatures of positive selection compared to 23% in FRDs. We found that introgression from wolves may provide a considerable adaptive advantage to FRDs, counterbalancing some of the negative effects of domestication, which can include reduced genetic diversity and excessive tameness. In wolves, introgression from FRDs is mostly driven by drift, with a small number of positively selected genes associated with brain function and behaviour. The predominance of drift may be the consequence of small effective size of wolf populations, which reduces efficiency of selection for weakly advantageous or against weakly disadvantageous introgressed variants. Small wolf population sizes result largely from human-induced habitat loss and hunting, thus linking introgression rates to anthropogenic processes.

OS1-2

Experimental assessment of predicted coevolving amino acids in *Escherichia coli*

Muhammad B. Haider[1], Jenna Gallie[1], Julien Dutheil[2]
[1]Max Planck Institute for Evolutionary Biology, Plön, Germany. [2]Max Planck Institute for Evolutionary Biology, Plön, Plön, Germany

Abstract

Amino acids within a protein sequence interact to maintain their structure and function. Consequently, mutations at a given functionally significant position can be potentially compensated by a mutation at an interacting position. For instance, a substitution from an amino acid with a small side chain to an amino acid with a big side chain can be compensated by a reciprocal substitution at an interacting position. Such a coevolution scenario implies that the first mutation leads to a fitness reduction, while the compensating mutation restores it. Several methods have been developed to detect coevolving positions from sequence alignments. However, the validation of the resulting predictions relies so far only on indirect evidence such as residue contact maps in proteins for which an experimental structure is available. In this study, we mapped substitutions for a protein sequence alignment on each branch of the phylogeny to detect coevolving amino acids in bacterial homologous protein families. Accounting for the biochemical properties of amino acids, we identified thousands of coevolving groups. We then selected candidate groups displaying a pattern of co-substitutions in the *Escherichia coli* branch, and experimentally reconstructed the local fitness landscape, resurrecting the ancestral genotype in *E. coli* and putting it in competition with single and double mutants using nutrient-enriched and single carbon source media. In this study, our results provide experimental evidence of the compensating nature of the predicted mutations in elongation factor 4 protein, highlighting the potential of coevolution detection methods as tools to understand molecular evolution.

OS1-3

Population structure of Brittany provides new insights on the introduction of steppe ancestry in Western Europe

Isabel Alves[1], Joanna Giemza[1], Michael G. Blum[2], Carolina Bernhardsson[3], Véronique Gallien[4], Elodie Cabot[5], Martial Monteil[6], Ashot Margaryan[7], Fernando Racimo[8], Eske Willerslev[8,9], Yves Coativy[10], Yvan Pailler[11], Clément Nicolas[12], Daniel Le Bris[10], Mael Jézéquel[10], Robert Olaso[13], Hélène Blanché-Koch[14], Anne Boland[13], Pierre Darlu[15], Mattias Jakobsson[3], Emmanuelle Genin[16], Jean-François Deleuze[13,14], Richard Redon[1], Christian Dina[1], FranceGenRef Consortium[17]
[1]L'institut du thorax, Inserm UMR 1087, CNRS UMR 6291, University of Nantes, Nantes, France. [2]TIMC-IMAG UMR 5525 CNRS, Univ. Grenoble Alpes, Grenoble, France. [3]Department of Organismal Biology and SciLife Lab, Uppsala University, Uppsala, Sweden. [4]INRAP / UMR 7264 CEPAM, CNRS-UNS Nice-Sophia Antipolis, Nice, France. [5]INRAP / UMR 7268 (Adès) Anthropologie Bio-Culturelle, Droit, Ethique et Santé, Faculté de Médecine Site Nord, Marseille, France. [6]Université de Nantes, UMR 6566 CReAAH, LARA, Nantes, France. [7]Center for Evolutionary Hologenomics, University of Copenhagen, Copenhagen, Denmark. [8]Lundbeck GeoGenetics Centre, Globe Institute, University of Copenhagen, Copenhagen, Denmark. [9]Department of Zoology University of Cambridge, Cambridge, United Kingdom. [10]Centre de Recherche Bretonne et Celtique, EA 4451, Université de Bretagne Occidentale, Brest, France. [11]INRAP; UMR 6554 LETG – Brest Institut Universitaire Européen de la Mer, Plouzané, France. [12]UMR 8215 « Trajectoires » Maison de l’Archéologie et de l’Ethnologie, Plouzané, France. [13]Commissariat à l'Energie Atomique, Université Paris-Saclay, Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, Labex GenMed, Evry, France. [14]Fondation Jean Dausset, CEPH, Labex GenMed, Paris, France. [15]CNRS/ MNHN/ Université Paris Diderot, UMR 7206 Eco-anthropologie et ethnobiologie, Paris, France. [16]UMR 1078 Génétique, Génomique fonctionnelle et Biotechnologies, Inserm, Univ Brest, EFS, CHU Brest, Brest, France. [17]France, France, France

Abstract

Present-day France lies at the confluence of the three migration waves that mostly contributed to the genetic ancestry of modern Europeans. However, little is known about the interaction between the edges of those migrations and how it gave rise to modern population structure.

To fill in this gap, we generated genome-wide data for >3,000 present-day individuals and ~850 high-coverage full genomes from the northern half of France and merged them with 100s of publicly available modern and ancient Europe-wide samples. Due to the complete absence of French samples from the two last millennia, we also present, for the first time, ancient DNA from six Medieval individuals (300-1100 CE) from France.

Haplotype and rare allele sharing revealed extensive fine-scale structure in Northwestern France. We also found relatively large differentiation of Western Brittany relative to the rest of the country, and an overall increased differentiation between the northern and southern sides of the river Loire accompanied by different proportions of northwestern European- versus Mediterranean-related ancestry, with Western Brittany individuals carrying unique levels (~75%) of Irish-related ancestry. Within France, individuals in Western Brittany show the largest levels of steppe ancestry and share levels of ancestry with Bell Beakers-associated individuals only comparable with those found in other populations lying on the Northwestern edges of Europe. Together, we provide evidence that the Bronze Age arrival of peoples from the Pontic steppe may have reached as far as present-day Brittany and massively reshaped the genetic makeup of Europeans living on the shores of the North Sea.

OS1-4

Detecting natural selection in trait-trait coevolution

Daohan Jiang, Jianzhi Zhang
University of Michigan, Ann Arbor, MI, USA

Abstract

No phenotypic trait evolves independently of all other traits, but the cause of trait-trait coevolution is poorly understood. While the coevolution could arise simply from pleiotropic mutations that simultaneously affect the traits concerned, it could also result from multivariate natural selection favoring certain trait relationships. To gain a general mechanistic understanding of trait-trait coevolution, we examine the evolution of 220 cell morphology traits across 16 natural strains of the yeast *Saccharomyces cerevisiae* and the evolution of 24 wing morphology traits across 110 fly species of the family Drosophilidae, along with the variations of these traits among gene deletion or mutation accumulation lines (a.k.a. mutants). For numerous trait pairs, the phenotypic correlation among evolutionary lineages differs significantly from that among mutants. Specifically, we find hundreds of cases where the evolutionary correlation between traits is strengthened or reversed relative to the mutational correlation, which, according to our population genetic simulation, is likely caused by multivariate selection. Furthermore, we detect selection for enhanced modularity of the yeast traits analyzed. Together, these results demonstrate that trait-trait coevolution is shaped by natural selection and suggest that the pleiotropic structure of mutation is not optimal. Because the morphological traits analyzed here are chosen largely because of their measurability, our conclusion is likely general.

OS1-5

The genetic structure of western Remote Oceanians

Lara R Arauna[1], Jeremy Choin[1,2], Javier Mendoza-Revilla[1], Jacob Bergstedt[1], Christine Harmant[1], Laure Lemee[3], Olivier Cassar[4], Antoine Gessain[4], Lluis Quintana-Murci[1,5], Etienne Patin[1]

[1]Human Evolutionary Genetics Unit, Institut Pasteur, Paris, France. [2]Université Paris Diderot, Sorbonne Paris Cité, Paris, France. [3]Biomics Platform, Institut Pasteur, Paris, France. [4]Oncogenic Virus Epidemiology and Pathophysiology, Institut Pasteur, Paris, France. [5]Chair Human Genomics and Evolution, Collège de France, Paris, France

Abstract

The Vanuatu archipelago, the gateway to Remote Oceania, was peopled during the last human large-scale migration to uninhabited lands, ~3,000 years ago (ya). Ancient DNA studies support an initial settlement by East Asian-related peoples, associated to the Lapita culture, which was quickly followed by the arrival of Papuan-related populations from the Bismarck Archipelago, leading to a major replacement. Despite the high linguistic and cultural diversity of present-day ni-Vanuatu, little is known about the patterns of genetic diversity and admixture processes in contemporary ni-Vanuatu. We generated SNP array data from >1,000 ni-Vanuatu, originating from 30 islands. We observe limited genetic differentiation among ni-Vanuatu, suggesting a common peopling history for the archipelago. Accordingly, admixture dates between the East Asian-related and Papuan-related ancestors of ni-Vanuatu suggest a unique admixture event occurring ~2,000 ya. Yet, we observe fine-scale genetic structure that mirrors later admixture events, geographical barriers and sociocultural factors. We observe differences in ancestry proportions among islands, in line with a geographically uneven replacement. Our analyses detect Polynesian ancestry arriving ~600-1,000 ya in southern Vanuatu, supporting back migrations from Polynesia. Finally, using ~300 male-female couples, we find evidence for a mating preference for partners of similar genetic ancestry, while accounting for relatedness avoidance. However, the signal was not driven by genetic variation at specific loci, suggesting that ancestry is used as a marker for social assortative mating. Together, this study sheds new light onto the peopling of Vanuatu and its interactions with the rich sociocultural diversity of the region.

OS1-6

Evolution of sexual dimorphism facilitated by Y-linked genes in a seed beetle.

Philipp Kaufmann[1], Arild Husby[1], Matthew E. Wolak[2], Elina Immonen[1]
[1]Uppsala University, Uppsala, Sweden. [2]Auburn University, Auburn, USA

Abstract

Females and males commonly have different fitness optima for shared traits, such as body size, however their independent response to selection is constrained by their mostly shared genome. This sparks the question what are the genetic underpinnings that can facilitate the evolution of sexual dimorphism. The sex chromosomes have been hypothesised to play an important role due to their asymmetric inheritance. Molecular signatures of sexual antagonism on the sex chromosomes often align with this theory, but demonstrations of their role in phenotypic dimorphisms are scarce. This is particularly so for the Y chromosome, which is subject to progressive degeneration due to halted recombination that limits segregating polymorphism. Here we test empirically how body size dimorphism can evolve in the seed beetle Callosobruchus maculatus and what role the sex chromosomes play in the process. Using a combination of quantitative genetics and artificial selection, we show that Y-linked genetic variance has a major effect on male body size and that sexual dimorphism can evolve rapidly in this system, despite an otherwise high genetic correlation between the sexes for body size that is believed to constrain the evolution of dimorphism. We then isolated and quantified the effect of Y-linked genetic variance on male body size by introgressing different Y haplotypes into an isogenic background, demonstrating that sexual dimorphism can indeed evolve through Y-linked genetic variance. Further increase of dimorphism even after a Y haplotype is fixed, suggests that sexual dimorphism can evolve through multiple mechanisms simultaneously.

OS1-7

The role of gene expression plasticity during recurrent altitudinal divergence

Aglaia Szukala[1,2], Bozo Frajman[3], Peter Schönswetter[3], Daniel Schubert[4], Ovidiu Paun[1]
[1]University of Vienna, Vienna, Austria. [2]Vienna graduate school of population genetics, Vienna, Austria. [3]University of Innsbruck, Innsbruck, Austria. [4]Freie Universität Berlin, Berlin, Germany

Abstract

Independent instances of divergence with similar phenotypic outcome (*parallel evolution*) provide natural evolutionary replicates to investigate adaptation to different ecological niches. The alpine plant *Heliosperma pusillum* Waldst. & Kit. (Caryophyllaceae) comprises two elevational ecotypes (i.e. montane and alpine) that diverged five times independently in different geographic regions of the South Eastern Alps. The montane and alpine growing sites significantly differ in terms of temperature, and light and water availability (i.e. warmer/drier/shaded montane vs colder/humid/exposed alpine). To understand the genetic vs. environmental components of the phenotypic divergence, we performed transcriptomic analyses of plants grown in reciprocal transplantations at natural sites. Our results show that the adaptation to different altitudes involves a limited proportion of constitutive changes in gene expression. Interestingly, we observe that the montane ecotype bears significantly higher plasticity of gene expression then the alpine. Genes that change expression plastically are involved in response to high light, flavonoid biosynthesis, oxidation-reduction processes and methylation. We further tested if the two ecotypes are differentially affected by global inhibition of methylation using zebularine. We show that the montane ecotype recovers faster from the de-methylation shock, possibly due to higher plastic potential of methylation-related genes. We conclude that higher phenotypic plasticity, possibly mediated by dynamic methylation rewiring, likely evolved in response to drier and warmer environments in this plant system.

OS1-8

Mutational constraints on a transmembrane protein revealed by deep scanning mutagenesis

Steven K. Chen[1], Benjamin M. Scott[2], Alex Van Nynatten[1], Ryan K. Schott[3], Sergio Peisajovich[4], Belinda S.W. Chang[1]

[1]University of Toronto, Toronto, Ontario, Canada. [2]Concordia University, Montreal, Quebec, Canada. [3]York University, Toronto, Ontario, Canada. [4]Illumina, San Diego, California, USA

Abstract

G protein-coupled receptors (GPCRs) make up the largest family of transmembrane protein receptors in eukaryotes and are responsible for highly adaptive traits that range from light perception to immune system regulation. Although genetic variation in GPCR coding genes across species and within populations are known to affect GPCR function, how do these mutations confer their effects to affect organismal fitness and evolution remains unclear. To study this relationship at the molecular level, we focused on rhodopsin, a highly adaptive light-activated GPCR that triggers the visual transduction cascade. Our group recently engineered and validated a high-throughput system for measuring rhodopsin function via a fluorescent readout by coupling it to the yeast mating pathway. Here, we employ this system to conduct a deep scanning mutagenesis approach to test the sequence-function relationship for random libraries of rhodopsin variants. Using a combination of fluorescence-activated cell sorting and next generation sequencing, rhodopsin activation was quantified for 1205 rhodopsin mutants. Our results show heterogeneity in mutational tolerance among topological domains in rhodopsin, and that helical density is inversely correlated with mutational tolerance – a previously predicted property of transmembrane proteins but was never observed directly. Our findings suggest that the positional organization of amino acids in proteins influence their mutational tolerance which may affect a protein’s evolutionary potential.

OS2-1

DNA evolution depends on differential methylation patterns in rat speciation

Xiao-Hui Liu[1], Ying Song[1], Ning Li[1], Da-Wei Wang[1], Fei Ling[2], Jin-Min Lian[3], Qi-Ye Li[3], Ya-Bin Jin[2], Zhi-Yong Feng[4], Lin Cong[5], Dan-Dan Yao[4], Chan Luo[5], Ming D. Li[6]

[1]Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China. [2]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China. [3]China National GeneBank, BGI-Shenzhen, Shenzhen, China. [4]Plant Protection Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China. [5]Plant Protection Institute, Heilongjiang Academy of Agricultural Sciences, Harbin, China. [6]The First Affiliated Hospital, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University School of Medicine, Hangzhou, China

Abstract

The fixation of phenotypes and underlying alleles is a typical evolutionary process in speciation. As the primary molecular basis of phenotypic plasticity, epigenetic mechanisms also play an essential role in maintaining phenotypes. However, whether and how DNA evolution was shaped by epigenetic alteration, especially accompanied DNA fixation in speciation, remains unknown. We used sperm methylomes of three rat subspecies as epigenetic markers and screened out genomic regions that experienced distinct differential methylation. To obtain independent results, they were further filtrated according to genomic locations to guarantee that their evolutionary features were not interactively affected by nearby DMRs (differentially methylated regions) of other datasets. By analyzing intraspecies and interspecies phylogenetic relationships, we showed that, in the same genomic regions, the significantly accelerated DNA evolution only occurred in individuals or lineages that experienced differential methylation. Across the same genomes, differential methylation led to a significant increase of F_{ST} only in lineage-specific DMRs and a significant increase of π in both individual-specific and lineage-specific DMRs. Correlations among methylation, π and F_{ST} showed that it was methylation consistency rather than the absolute methylation difference that significantly influenced both π and F_{ST}. The change of both π and DNA fixation depended on the degree of intraspecies methylation consistency. While the breakdown of methylation consistency facilitated the promotion of π, the maintenance of methylation consistency facilitated the acceleration of DNA fixation.

OS2-2

Post-Insemination Selection Dominates Pre-Insemination Selection in Driving Rapid Evolution of Male Competitive Ability

Katja R Kasimatis[1,2], Megan J Moerdyk-Schauwecker[2], John H Willis[2], Patrick C Phillips[2]
[1]University of Toronto, Toronto, ON, Canada. [2]University of Oregon, Eugene, OR, USA

Abstract

Sexual selection drives the evolution of some of the most remarkable phenotypes observed in nature. From a male's perspective, sexual selection can optimize reproductive success by acting on the variance in mating success (pre-insemination selection) as well as the variance in fertilization success (post-insemination selection). Indirect evidence from lifetime reproductive-success experiments indicates that pre-insemination selection can be the dominant determinant of male reproductive fitness. In contrast, evolutionary analyses of seminal fluid proteins show a high opportunity for post-insemination selection. However, the balance between pre- and post-insemination selection has not been examined using a framework that directly isolates and quantifies the effects of post-insemination selection. We used experimental evolution of a uniquely engineered genetic system that allows sperm production to be turned off and on in obligate male/female populations of *Caenorhabditis elegans* to show that enhanced post-insemination competition increases the efficacy of selection and surpasses pre-insemination sexual selection in driving a polygenic response in male reproductive success. We identified a strong, rapid response to selection after 30 generations of experimental evolution. Post-insemination competitive ability increased by 5- to 7-fold in all evolved populations, suggesting a strong underlying pressure on sperm competitive ability. Pre-insemination competitive ability increased to a lesser degree, however, this response was reduced under enhanced pre-mating competition conditions. Furthermore, we found that post-insemination selection resulted in a strong polygenic response at the genomic level. Our work indicates that post-insemination interactions are important in determining male reproductive fitness and likely are genetically complex.

OS2-3

Goldilocks and the Three Genotypes: Characterizing the Prevalence of Overdominance for Adaptive Mutations that Arise in Diploids

Vivian K Chen[1], Monica R Sanchez[2], Tsuya Yokoyama[1], Samuel B Hoelscher[3], Dmitri A Petrov[1], Gavin Sherlock[1]

[1]Stanford University, Stanford, CA, USA. [2]Los Alamos National Labs, Los Alamos, New Mexico, USA. [3]University of California San Francisco, San Francisco, CA, USA

Abstract

There is a strong bias against the establishment of recessive beneficial mutations because new adaptive mutations in diploids tend to be heterozygous (Aa). Thus, adaptive mutations in diploids must be more fit than the homozygous ancestor (AA). This pattern is known as Haldane's sieve. Theoretical predictions suggest that adaptive mutations in diploids should often be overdominant (f(Aa) > f(AA)). Limited empirical data are consistent with these predictions but the number of mutations tested so far is too small to confidently make this claim.

Lack of high-throughput methods has hampered our ability to measure dominance of mutations. We combine barcoding and CRISPR/Cas9 approaches to determine the prevalence of overdominance for 100+ adaptive mutations that arose in diploid evolutions. Diploids were barcoded, and evolved under various experimental conditions. We used barcode lineage tracking to track adaptation and through clone sequencing identified 100+ heterozygous, putative adaptive mutations. Some mutations were tested by segregation and found to be recessive lethal - these are by definition overdominant.

For the non-lethal subset, we completed a pilot combining iSeq 2.0 barcoding with CRISPR/Cas9 to engineer mutations into barcoded haploids. From these haploids, we generated uniquely double-barcoded diploids of all genotypes (AA, Aa, aa) for each mutation. Thus, we can distinguish and measure fitness of all genotypes for all mutants in a single pool. Through this proof of concept pilot, we determined the distribution of fitness effects and the dominance of 17 mutations. We are now greatly expanding this set, and will report our progress.

OS2-4

Evolutionary persistent Y chromosomes protected by meiotic executioners

Paul D Waters[1], Aurora Ruiz-Herrera[2]
[1]UNSW Sydney, Sydney, NSW, Australia. [2]Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

The X and Y chromosomes were once identical, having evolved from an ordinary pair of autosomes via process of Y degradation. As such, the Y chromosome has been touted as a wimpy relic of the X, with its survival dependent on a few critical functions in spermatogenesis and sex determination, the loss of which would signal its demise. Why then has it survived since its origin (~165 MYA) in all but a handful of therian mammal species? This is in stark contrast to the high turnover of sex chromosomes and sex determining switches observed in other vertebrate linages. The mammal Y turns out to be an exception of persistence, rather than the rule. Here we propose a novel explanation for such perseverance: the persistent Y hypothesis. The Y chromosome bears genes that are critical for successful meiotic progression, the expression of these so-called ‘executioner genes’ are tightly regulated (i.e., silenced during the meiotic sex chromosome inactivation - MSCI) to ensure germ cell survival, hence contributing to the offspring. When executioners are translocatedto an autosome they escape this silencing, and being pachytene-lethal, cease meiosis. Thus, the only heritable transposition events are to the X – where they remain subject to meiotic silencing. We propose that in eutherian mammals, *Zfy* genes act as ‘executioner genes’, playing this role as their own judge, jury and executioner. Herein we discuss how this selfishness of meiotic executioners has posed strong evolutionary constraints for the Y chromosome to persist in mammals.

OS2-5

Leveraging quantitative genomics for the conservation of the critically endangered kākāpō

Lara Urban[1], Joseph Guhlin[2], Marissa Le Lec[2], Jana Wold[3], Daryl Eason[4], Lydia Uddstrom[4], Deidre Vercoe[4], Jon Slate[5], Cock Van Oosterhout[6], Hernán E. Morales[7], Alana Alexander[1], Emily Koot[8], Murray Cox[9], Patrick Biggs[9], David Winter[9], Tammy Steeves[3], Neil J. Gemmell[1], Anna Santure[10], Peter Dearden[2], Andrew Digby[4]

[1]Department of Anatomy, University of Otago, Dunedin, New Zealand. [2]Department of Biochemistry, University of Otago, Dunedin, New Zealand. [3]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. [4]Kākāpō Recovery, Department of Conservation, Invercargill, New Zealand. [5]Department of Animal & Plant Sciences, University of Sheffield, Sheffield, United Kingdom. [6]School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom. [7]Section for Evolutionary Genomics, GLOBE institute, University of Copenhagen, Copenhagen, Denmark. [8]The New Zealand Institute for Plant and Food Research, Palmerston North, New Zealand. [9]School of Fundamental Sciences, Massey University, Palmerston North, New Zealand. [10]School of Biological Sciences, University of Auckland, Auckland, New Zealand

Abstract

Across the world, wild populations are declining at an unprecedented rate, culminating in the loss of genetic diversity and adaptive potential. Only few genomic studies have informed conservation decisions, often due to the challenges associated with complementing genomics with fitness or demographic data. Combining genomic and phenotypic data, however, enables the application of quantitative genomic approaches that can directly assess the adaptive potential of populations.

We leveraged cutting-edge quantitative genomic approaches to explore genomic, phenotypic, and environmental factors that affect the persistence of the critically endangered kākāpō (*Strigops habroptilus*). An extraordinarily detailed phenotypic catalogue allowed us to (i) assess heritability and polygenicity, (ii) identify underlying genes and gene pathways, and (iii) evaluate genomic prediction of various traits.

We will report on two exemplary phenotypes, plumage colour and disease susceptibility, to showcase the power of such approaches and their potential to benefit conservation management: We found kākāpō plumage to be a Mendelian trait, allowing us to leverage genomic data to correct errors in the phenotypic catalogue and to accurately predict the plumage of the next generation of kākāpō. As *LHX8*, the potential causative gene, is known to be involved in tissue differentiation, we assessed additional evidence of the plumage colour being caused by structural differences.

We further identified several immune genes that contribute to the species' susceptibility to the exudative cloacitis disease. Additional gene pathway analyses indicate that cloacitis might be caused by retroviral infection, an important clue in the search for the still unknown causative agent of the disease.

OS2-6

Dating Alphaproteobacteria evolution with eukaryotic fossils

Sishuo Wang, Haiwei Luo
The Chinese University of Hong Kong, Hong Kong, Hong Kong

Abstract

Elucidating the timescale of the evolution of Alphaproteobacteria, one of the most prevalent microbial lineages in marine and terrestrial ecosystems, is key to testing hypotheses on their co-evolution with eukaryotic hosts and Earth's system, which, however, is largely limited by the scarcity of bacterial fossils. Here, we incorporate eukaryotic fossils to date the divergence times of Alphaproteobacteria, based on the mitochondrial endosymbiosis that mitochondria evolved from an alphaproteobacterial lineage. We estimate that Alphaproteobacteria arose ~1900 million years (Ma) ago, followed by rapid divergence of their major clades. We show that the origin of Rickettsiales, an order of obligate intracellular bacteria whose hosts are mostly animals, predates the emergence of animals for ~700 Ma but coincides with that of eukaryotes. This, together with reconstruction of ancestral hosts, strongly suggests that early Rickettsiales lineages had established previously underappreciated interactions with unicellular eukaryotes. Moreover, the mitochondria-based approach displays higher robustness to uncertainties in the calibrations compared with the traditional strategy using cyanobacterial fossils. Further, our analyses imply the potential of dating the (bacterial) tree of life based on endosymbiosis events, and suggest that previous applications using divergence times of the modern hosts of symbiotic bacteria to date the bacterial tree of life might need to be revisited.

OS2-7

Genomics of secondarily temperate adaptation in ecologically divergent icefish species

Angel G Rivera-Colón[1], Christina Cheng[1], Julian M Catchen[2]
[1]University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]University of Illinois at Urbana-Champaign, Ur, IL, USA

Abstract

White-blooded Antarctic icefishes display extreme specialization to the narrow temperature and oxygen conditions of the Southern Ocean. Even after this extreme specialization, a single icefish species can be found in a temperate environment. *Champsocephalus esox* evolved from a cold-specialized ancestor and retains many icefish-defining traits yet survives in an environment much warmer and less oxygenated than all other Antarctic icefish. We used comparative genomics to study the evolution of secondarily temperate adaptation, sequencing the genomes of the temperate *C. esox* and its Antarctic sister species *Champsocephalus gunnari*. Genome-wide comparative synteny shows evidence of genomic rearrangements between both species, although the large-scale structure of the genome appears to be conserved. When measuring genetic diversity and divergence, we see that some of the identified structural variants also show elevated divergence between the ecologically divergent species, highlighting their role as putative islands of speciation and divergence. In addition, a coalescent genealogy approach reiterates the complex evolutionary history, origin, and timing of the polar-to-temperate transition in Antarctic icefishes. Studying a naturally occurring secondarily temperate icefish provides a natural experiment for assessing the evolvability and survival of Antarctic marine fauna as they face a rapidly warming Southern Ocean. More broadly, this system serves as a unique example for how ecological specialization is not exclusively an evolutionary dead end, highlighting the mechanisms that allow specialists to diversify and adapt to changing environments.

OS2-8

Cis-regulatory convergence between the extinct Tasmanian tiger and canids

Laura E Cook[1], Charles Y Feigin[2], Irene Gallego Romero[1], Andrew J Pask[1]
[1]University of Melbourne, Melbourne, Victoria, Australia. [2]Princeton University, Princeton, New Jersey, USA

Abstract

Despite last sharing a common ancestor 160 million years ago, the extinct Tasmanian tiger (thylacine) and grey wolf are one of the most striking examples of phenotypic convergence in mammals. Due to the evolutionary distance between them and their extreme morphological similarity, especially in skull shape, these species are an exceptional model for investigating whether species use the same or different molecular mechanisms to produce near-identical phenotypes. We discovered 339 conserved non-coding regions that are under convergent accelerated evolution in the thylacine and the wolf, representing putative regulatory regions underpinning their convergent skull shape. These candidate regions overlap with active cis-regulatory elements (CREs) in mouse embryonic facial prominences from embryonic day (E) 10.5 -15.5. We investigated whether they also correspond to active CREs during marsupial craniofacial development by performing H3K4me3 and H3K27ac ChIP-seq in the developing facial prominences of the fat-tailed dunnart (*Sminthopsis crassicaudata*), another closely related carnivorous marsupial. These data represent the first epigenome for craniofacial tissue of any marsupial. Of our previously identified accelerated regions, 124 overlapped with these marsupial CREs, refining our list to accelerated regions involved in marsupial skull development. In addition to this, we have found that 40% of convergent accelerated regions contain convergently gained or lost transcription factor binding sites. This further suggests that these candidates may alter spatiotemporal gene expression or gene expression levels through altered transcription factor binding. In conclusion we provide preliminary evidence that convergent regulatory evolution is associated with convergent phenotypic evolution in the thylacine and wolf.

OS3-1

Molecular Evolutionand the Decline of Purifying Selection with Age

Changde Cheng[1], Mark Kirkpatrick[2]
[1]St. Jude's Research Hospital, Memphis, Tennessee, USA. [2]University of Texas, Austin, Texas, USA

Abstract

Life history theory predicts that the intensity of selection declines with age, and this trend should impact how genes expressed at different ages evolve. Here we find consistent relationships between a gene's age of expression and patterns of molecular evolution in two mammals (the human *Homo sapiens* and the mouse *Mus musculus*) and two insects (the malaria mosquito *Anopheles gambiae* and the fruit fly *Drosophila melanogaster*). When expressed later in life, genes fix nonsynonymous mutations more frequently, are more polymorphic for nonsynonymous mutations, and have shorter evolutionary lifespans, relative to those expressed early. The latter pattern is explained by a simple evolutionary model. Further, early-expressed genes tend to be enriched in similar gene ontology terms across species, while late-expressed genes show no such consistency. In humans, late-expressed genes are more likely to be linked to cancer and to segregate for dominant disease-causing mutations. Last, the effective strength of selection ($N_e s$) decreases and the fraction of mutations increases with a gene's age of expression. These results are consistent with the diminishing efficacy of purifying selection with age, as proposed by Medawar's classic hypothesis for the evolution of senescence, and provide links between life history theory and molecular evolution.

OS3-2

Higher ancestral rate of evolution of duplicated genes is consistent across diverse lineages

Zoe Vance[1], Laurence D. Hurst[2], Aoife McLysaght[1]
[1]Trinity College Dublin, Dublin, Ireland. [2]University of Bath, Bath, United Kingdom

Abstract

The faster rate of evolution of duplicated genes relative to singletons has been well documented in multiple lineages. This observation has generally been attributed to a release from constraint following creation of a redundant copy by duplication. However it is also possible that duplicated genes were faster evolving even prior to duplication, with this property being correlated with the likelihood of duplication and leading to an enrichment of fast evolving genes among duplicates over time. Previous studies investigating this idea have come to different conclusions and clarity is needed on whether this disagreement relates to differences in methodology or legitimate biological differences between the lineages selected for study. Here, we show that duplicable genes are faster evolving prior to duplication in the *Drosophila* lineage. This supports results in primates showing that ancestrally faster evolving genes are more likely to duplicate, in contrast to results from analysis of paralogs in yeast and nematodes claiming conserved genes are more likely to duplicate. Our findings indicate that this higher tendency to duplicate among fast evolving genes is likely a broadly applicable principle of gene duplication not restricted to a single lineage. This has broad implications for our understanding of the mechanisms which initially create gene duplicates and later select for their retention.

OS3-3

The origin of genes in pieces: Tracing the spread of introns during eukaryogenesis

Julian Vosseberg, Michelle Schinkel, Sjoerd Gremmen, Berend Snel
Utrecht University, Utrecht, Netherlands

Abstract

The emergence of the eukaryotic cell from its prokaryotic ancestors is one of the most enigmatic evolutionary events. It has become clear that the last eukaryotic common ancestor (LECA) already had a complex nature, reflected by its full eukaryotic cellular compartmentalisation and a relatively large genome consisting of many paralogs. The protein-coding genes of LECA were interspersed with introns, a unique feature of eukaryotic genomes. It has remained largely elusive how and in which order the numerous genomic and cellular changes at the origin of eukaryotes occurred. Certain gene duplications predating LECA share ancestrally conserved intron positions. These paralogous intron positions provide a unique and currently not exploited means to shed light on the ordering of crucial events in early eukaryotes. We created multiple sequence alignments of large sets of duplicated families and detected many more paralogous introns than a previous large-scale study (Sverdlov et al., Trends Genet., 2007), which was based on only a handful of genomes. In particular, paralogs involved in signal transduction and the cytoskeleton shared intron positions. These paralogous introns suggest an early origin of introns during eukaryogenesis. Using the information from shared intron positions is a promising method to trace the rise in cellular and genomic complexity during early eukaryotic evolution.

OS3-4

Widespread introgression across a phylogeny of 155 Drosophila genomes

ANTON SUVOROV[1], Bernard Kim[2], Jeremy Wang[1], Ellie Armstrong[2], David Peede[1], Emmanuel D'Agostino[1], Donald Price[3], Peter Wadell[4], Michael Lang[5], Virginie Courtier-Orgogozo[5], Jean David[6,7], Dmitri Petrov[2], Daniel Matute[1], Daniel Schrider[1], Aaron Comeault[8]
[1]University of North Carolina, Chapel Hill, Chapel Hill, USA. [2]Stanford University, Stanford, USA. [3]University of Nevada, Las Vegas, USA. [4]Massey University, Palmerston North, New Zealand. [5]Université de Paris, Paris, France. [6]Université Paris-Saclay, Gif sur Yvette, France. [7]Sorbonne Universités, Paris, France. [8]Bangor University, Bangor, United Kingdom

Abstract

Genome-scale sequence data has invigorated the study of hybridization and introgression, particularly in animals. However, outside of a few notable cases, we lack systematic tests for introgression at a larger phylogenetic scale across entire clades. Here we leverage 155 genome assemblies, from 149 species, to generate a fossil-calibrated phylogeny and conduct multilocus tests for introgression across 9 monophyletic radiations within the genus Drosophila. Using complementary phylogenomic approaches, we identify widespread introgression across the evolutionary history of Drosophila. Mapping gene-tree discordance onto the phylogeny revealed that both ancient and recent introgression has occurred, with introgression at the base of species radiations being particularly common. Our results provide the first evidence of introgression occurring across the evolutionary history of Drosophila and highlight the need to continue to study the evolutionary consequences of hybridization and introgression in this genus and across the Tree of Life.

OS3-5

Giant genomes uncover ecological speciation in the deep ocean

Alexandra A-T Weber[1,2], Adnan Moussalli[2], Andrew F. Hugall[2], Sophie Arnaud-Haond[3], Jozée Sarrazin[1], Matthew D. McGee[4], Timothy D. O'Hara[2]

[1]Ifremer, Brest, France. [2]Museums Victoria, Melbourne, Australia. [3]Ifremer, Sète, France. [4]Monash University, Melbourne, Australia

Abstract

The deep ocean is the largest biome on Earth and yet it is among the least studied environments of our planet. Hence, the genomic mechanisms underlying the formation of species in the deep sea remain virtually unknown. Here we present the assembly of one of the largest invertebrate genomes. We used 10x chromium linked-reads, ONT long reads and transcriptomic reads to generate a draft genome assembly. The abyssal brittle star *Ophiosphalma armigerum* (Echinodermata: Ophiuroidea) genome is 8 Gb large and contains about 52% repeats. We further sequenced 123 *O. armigerum* individuals spanning the whole species geographic (from the North Atlantic Ocean to South Australia) and bathymetric range (from 2200m to 4800m depth). We found significant bathymetric structuring among bathyal (~2500m) and abyssal (~4000m) populations, suggesting ecological speciation along a depth gradient in *O. armigerum.* We then describe genomic islands of speciation and candidate genes for deep-sea adaptation. When focusing on the individuals from abyssal depths, we uncovered high levels of gene flow among localities around Australia. Remarkably, trans-oceanic connectivity was detected as individuals from the North Atlantic Ocean and individuals from South Australia were connected by significant levels of gene flow at abyssal depths. This study provides unprecedented marine invertebrate genomic resources and sheds light on speciation and connectivity mechanisms in the largest biome of the planet.

OS3-6

Population structure limits parallel evolution in sticklebacks

Bohao Fang[1], Petri Kemppainen[1], Paolo Momigliano[1], Juha Merilä[1,2]
[1]University of Helsinki, Helsinki, Finland. [2]The University of Hong Kong, Hong Kong, Hong Kong

Abstract

Population genetic theory predicts that small effective population sizes (N_e) and restricted gene flow limit the potential for local adaptation. In particular, the probability of evolving similar phenotypes based on shared genetic mechanisms (*i.e.* parallel evolution), is expected to be reduced. We tested these predictions in a comparative genomic study of two ecologically similar and geographically co-distributed stickleback species (*viz*. *Gasterosteus aculeatus* and *Pungitius pungitius*). We found that *P. pungitius* harbours less genetic diversity and exhibits higher levels of genetic differentiation and isolation-by-distance than *G. aculeatus*. Conversely, *G. aculeatus* exhibits a stronger degree of genetic parallelism across freshwater populations than *P. pungitius*: 2996 *vs.* 379 SNPs located within 26 *vs.* nine genomic regions show evidence of selection in multiple freshwater populations of *G. aculeatus* and *P. pungitius*, respectively. Most regions involved in parallel evolution in *G. aculeatus* showed increased levels of divergence, suggestive of selection on ancient haplotypes. In contrast, haplotypes involved in freshwater adaptation in *P. pungitius* were younger, and often associated with reduced diversity. In accordance with theory, the results suggest that connectivity and genetic drift play crucial roles in determining the levels and geographic distribution of standing genetic variation, providing evidence that population subdivision limits local adaptation and therefore also the likelihood of parallel evolution.

OS3-7

The genomic signatures of natural selection in admixed human populations

Sebastian Cuadros Espinoza[1,2], Guillaume Laval[1], Lluis Quintana Murci[1,3], Etienne Patin[1]
[1]Institut Pasteur, Paris, France. [2]Collège Doctoral Sorbonne Université, Paris, France. [3]Chair Human Genomics and Evolution, Collège de France, Paris, France

Abstract

Admixture is now recognized as a major evolutionary force in human evolution, yet the degree to which it facilitated adaptation to local environments, a process known as adaptive admixture, remains largely unexplored. Few evolutionary studies include admixed populations, and the power to detect positive selection in admixed populations is unknown. Here, we used extensive forward simulations to evaluate the power of various neutrality statistics to detect adaptive admixture, under realistic assumptions accounting for background selection and complex admixture scenarios. We show that classical neutrality statistics (i.e iHS and F_{ST}) falsely identify neutral mutations as beneficial in the admixed population, because they were adaptive in the source populations. Conversely, admixture specific statistics based on deviations of allele frequencies and local ancestry present high power to detect beneficial mutations in the admixed population, particularly when admixture is ancient and admixture rate is low. We next integrate the latter statistics in genome-wide scans for adaptive admixture in 13 different worldwide populations. We replicate previous studies suggesting that lactase persistence in East Africa and resistance against *Plasmodium vivax* malaria, in Madagascar, North Africa and South Asia were acquired through admixture. Among other novel selection hits, we report a locus associated with immune tolerance to pathogens in Island Southeast Asia. Finally, our analyses identify the *HLA* locus as a worldwide hotspot of adaptive admixture. Collectively, our study provides new evidence that adaptive admixture was prevalent in the evolution of modern humans.

OS3-8

Intra-host genetic diversity and selection in acute SARS-CoV-2 infections

Lele Zhao, Katrina Lythgoe, Matthew Hall, Sebastian Ramos-Onsins, Tanya Golubchik, Luca Ferretti
Big Data Institute, Nuffield Department of Medicine, Oxford, United Kingdom

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has a narrow transmission bottleneck and low levels of diversity, which explain its comparatively slow evolutionary rates. However, SARS-CoV-2 shares some general features of other RNA viruses, namely high mutation rates and rich evolutionary dynamics within hosts. Studying the mutational dynamics and the selective pressures during SARS-CoV-2 infections is key to understand the evolution of the virus and the emergence of Variants of Concern, which currently represent a major risk for public health worldwide.

Here we present an in-depth analysis of the genetic diversity and selection pressure in >1300 deep-sequenced clinical samples from the first COVID-19 wave in the United Kingdom (Lythgoe, Hall et al, Science 2021). We infer genomic signatures of within- and between-host selection through a combination of intra-host frequency spectrum analysis, intra-host polymorphisms versus divergence, as well as comparisons of individual viral swarms. SARS-CoV-2 samples from different individuals and viral lineages are independent - but not necessarily identical - realisations of the evolutionary process during infection, hence we also explore the impact of individual heterogeneities and genomic background.

OS4-1

$2B$ or not $2B$: Can inbreeding depression be explained by recessive deleterious nonsynonymous mutations?

Chris Kyriazis, Kirk E Lohmueller
UCLA, Los Angeles, CA, USA

Abstract

Inbreeding depression is a ubiquitous phenomenon in diploid outbreeding organisms. For example, in humans, it is estimated that first-cousin marriages result in an increase in juvenile mortality of 4.4%. Although there is a growing consensus that inbreeding depression can be explained by recessive deleterious mutations rather than overdominance, the types of mutations that are responsible remain largely unknown. Estimates of the distribution of fitness effects (DFE) of nonsynonymous mutations are available for humans and other species. It remains to be tested whether nonsynonymous recessive deleterious mutations, with selective effects drawn from these distributions, can explain empirical estimates of the inbreeding load ($2B$). Here, we address this question using forward-in-time simulations of human coding sequence with DFEs, mutation rates, and demographic models inferred from the literature, combined with a range of plausible dominance coefficients. We find that, under nearly all parameter combinations, existing DFEs of nonsynonymous mutations can explain observed levels of inbreeding depression in humans. Specifically, our models predict values of $2B$ ranging from 1.2 to 5.8, which overlap with the observed human inbreeding load of 1.4. Our work validates molecular population genetic estimates of the DFE by showing that these DFEs can recapitulate reductions in fitness measured in an orthogonal manner. Further, our results suggest that there may not be a substantial contribution to inbreeding depression from overdominance, non-coding variation, or structural variation. These results have implications for modelling inbreeding depression in other organisms, where empirical estimates of the inbreeding load may not be available.

OS4-2

Analysis of Ancient Pan-Genomes from the Cariogenic Oral Microbe *Streptococcus mutans*

Megan Michel[1,2,3], Aida Andrades Valtueña[1,4], Rodrigo Barquera[1], Tiago Ferraz da Silva[1,5], Evelyn Guevara[6], Alissa Mittnik[7], Kathrin Nägele[4], Elizabeth A. Nelson[1,8], Gunnar Neumann[1,3], Maria Spyrou[1,9], Vanessa Villalba[1], Kirsten Bos[1], Wolfgang Haak[1], Cosimo Posth[1,9], Philipp W. Stockhammer[1,3,10], Alexander Herbig[1], Johannes Krause[4,3]

[1]Archaeogenetics Department, Max Planck Institute for the Science of Human History, Jena, Germany. [2]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [3]Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Munich, Germany. [4]Archaeogenetics Department, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [5]Institute of Bioscience, Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil. [6]Department of Forensic Medicine, University of Helsinki, Helsinki, Finland. [7]Department of Genetics, Harvard Medical School, Boston, MA, USA. [8]Department of Anthropology, University of Connecticut, Storrs, CT, USA. [9]Archaeo- and Palaeogenetics Group, Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany. [10]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilian University, Munich, Germany

Abstract

The gram-positive oral pathobiont *Streptococcus mutans* plays a key role in dental caries pathogenesis, given its ability to form multispecies oral biofilms, ferment diverse dietary carbohydrates, and tolerate acidic metabolic byproducts that demineralize tooth enamel. However, as *S. mutans* is genetically competent and recombines prolifically, modern isolates vary significantly in their repertoire of virulence-associated genes. Over time, human populations have experienced marked variation in caries burden, with generally higher incidence in agricultural populations consuming carbohydrate-rich diets. When *S. mutans* first acquired specific virulence gene clusters and whether these varied across groups practicing different subsistence strategies remains to be explored.

To address this question, we generated the first ancient *S. mutans* genome-wide data from a global set of 80 individuals spanning the Mesolithic to the modern era. After enrichment using a custom in-solution hybridization capture designed to encompass modern *S. mutans* pan-genomic diversity, 60 samples yielded an average genomic coverage of at least 3X. Comparing the gene content of our ancient strains to a modern *S. mutans* pan-genome suggests that the diversity present in modern isolates is remarkably old, with similar accessory components present across thousands of years of evolution. Interestingly, the presence of several loci appears to correlate with host subsistence patterns; strains from six ancient hunter-gatherers from Europe, Africa, and South America lack the ComCDE system, which regulates biofilm formation, bacteriocin production, and competence. Based on these findings, we highlight ancient pan-genomic analysis as a promising avenue for exploring the evolution of virulence phenotypes in *S. mutans*.

OS4-3

Ghost lineages deceive introgression tests and call for a new null hypothesis

Theo Tricou[1], Eric Tannier[1,2], Damien de Vienne[1]
[1]CNRS, Villeurbanne, France. [2]INRIA, Grenoble, France

Abstract

The data that is known and sampled in any evolutionary study is always a small part of what exists, known or not, or what has existed in the past and is extinct. Therefore it is likely that all detected past horizontal gene fluxes, hybridization, introgressions, admixtures or transfers, involve “ghosts”, that is, extinct or unsampled lineages. The presence of these ghosts is acknowledged by all scientists, but almost all wish that and make as if their blurring influence would be low, like a background noise that, with a reasonable approximation, can be ignored. We assess this undervalued hypothesis by qualifying and quantifying the effect of ghost lineages on introgression detection by the popular D-statistics method. We use a genomic dataset of bears to illustrate and circumscribe the possibility of misinterpretation and show on simulated data that under certain conditions, far from unrealistic, most results interpreted from D-statistics, concerning the existence of introgression and the identity of donors and recipients of horizontal gene flows, are erroneous. In particular, the use of a distant outgroup, usually given as a solid ground for these tests, leads in fact to an increase in the error probability, and to false interpretations in a vast majority of the cases. We argue for a switch of the null hypothesis: the results of detection methods for gene fluxes should be interpreted with the full and visible participation of the unknown ghosts.

OS4-4

A large and diverse autosomal haplotype is associated with sex-linked colour polymorphism in the Trinidadian guppy

Josephine R Paris[1], James R Whiting[1], Mitchel J Daniel[2], Joan Ferrer Obiol[3], Paul J Parsons[1], Mijke J van der Zee[1], Christopher W Wheat[4], Kimberly A Hughes[2], Bonnie A Fraser[1]
[1]University of Exeter, Exeter, United Kingdom. [2]Florida State University, Tallahassee, Florida, USA. [3]University of Barcelona, Barcelona, Spain. [4]Stockholm University, Stockholm, Sweden

Abstract

Colour polymorphism provides a tractable trait that can be harnessed to explore the evolution of sexual selection and sexual conflict. Male colour patterns of the Trinidadian guppy (*Poecilia reticulata*) are governed by both natural and sexual selection, and are typified by extreme pattern colour variation as a result of negative frequency dependent selection. Since guppy colour patterns are often inherited faithfully from fathers to sons, it has been historically presumed that colour genes are physically linked to sex determining loci as a ‘supergene’ on the sex chromosome. Yet the actual identity and genomic location of the colour pattern genes has remained elusive. We phenotyped and genotyped four guppy ‘Iso-Y lines’, where colour was inherited along the patriline, but backcrossed into the stock population every generation for 40 generations. Using an unbiased phenotyping method to proportion colour pattern differences between and among the Iso-Y lines, we confirmed that the breeding design was successful in producing four distinct colour patterns. Our analysis of genome resequencing data of the four Iso-Y lines uncovered a surprising genetic architecture for colour pattern polymorphism. Genetic differentiation among Iso-Y lines was repeatedly associated with a large and diverse haplotype (~5Mb) on an autosome (LG1), not the sex chromosome (LG12). Moreover, the LG1 haplotype showed elevated linkage disequilibrium and exhibited evidence of sex-specific diversity when we examined whole-genome sequencing data of the natural source population. We hypothesise that colour pattern polymorphism is driven by Y-autosome epistasis.

OS4-5

Structural complexity and evolution of large inversions underlying ecological adaptation in Atlantic herring

Minal S Jamsandekar[1], Mafalda Sousa Ferreira[2], Mats Pettersson[2], Angela P Fuentes-Pardo[2], Brian W Davis[1], Leif Andersson[2,1]

[1]Texas A&M University, College Station, Texas, USA. [2]Uppsala University, Uppsala, Sweden

Abstract

Atlantic herring is a pelagic fish and one of the most abundant vertebrates in the world. Using a high-quality genome assembly and whole genome sequencing of more than 50 population samples across the entire species distribution, we found hundreds of loci showing strong genetic differentiation between subpopulations due to ecological adaptation. Among these loci, we found four megabase sized blocks of SNPs in strong linkage disequilibrium and with sharp boundaries to flanking SNPs consistent with the presence of inversions. Haplotypes at all four loci showed striking frequency differences between populations spawning in the warmer waters surrounding Ireland and Britain compared with populations spawning further north in the Atlantic Ocean and in the Baltic sea. We therefore named haplotypes as ‘S’ (Southern) and ‘N’ (Northern). Further, we used PacBio HiFi technology to confirm and characterize inversions. We collected six samples each from populations with high frequencies of the S and N haplotypes, respectively, for sequencing. Using these data, we found exact coordinates for proximal and distal breakpoints and characterized structural variations around both breakpoints. Large deletions, retrotransposable elements and repetitive regions were abundant near the breakpoints. Theory suggests that transposable elements and deleterious mutations accumulate within the inversions due to relaxed purifying selection facilitated by suppression of recombination. We are using these data to investigate this hypothesis. This detailed characterization of inversions apparently maintained as balanced polymorphisms in one of the world’s most abundant vertebrates will shed light on the evolution of inversions and how they contribute to ecological adaptation.

OS4-6

Sediment DNA reveals Neandertal population history

Benjamin Vernot[1], Elena I Zavala[1], Asier Gómez-Olivencia[2], Zenobia Jacobs[3], Viviane Slon[1], Fabrizio Mafessoni[1], Frédéric Romagné[1], Alice Pearson[1], Martin Petr[1], Nohemi Sala[2], Adrián Pablos[2], Arantza Aranburu[4], José María Bermúdez de Castro[5], Eudald Carbonell[6], Bo Li[3], Maciej T Krajcarz[7], Andrey I Krivoshapkin[8], Kseniya A Kolobova[8], Maxim B Kozlikin[8], Michael V Shunkov[8], Anatoly P Derevianko[8], Bence Viola[9], Steffi Grote[1], Elena Essel[1], David López Herráez[1], Sarah Nagel[1], Birgit Nickel[1], Julia Richter[1], Anna Schmidt[1], Benjamin Peter[1], Janet Kelso[1], Richard G Roberts[3], Juan-Luis Arsuaga[2], Matthias Meyer[1]

[1]MPI EVA, Leipzig, Germany. [2]UCM-ISCIII, Madrid, Spain. [3]University of Wollongong, New South Wales, Australia. [4]UPV/EHU, Leioa, Spain. [5]CENIEH, Burgos, Spain. [6]IBHES, Tarragona, Spain. [7]Polish Academy of Sciences, Warsaw, Poland. [8]Russian Academy of Sciencs, Novosibirsk, Russian Federation. [9]University of Toronto, Toronto, Canada

Abstract

The study of hominin history has progressed through both archaeological and genetic insights. However, many archaeological sites lack associated hominin fossils, frustrating genetic analyses. Even when fossils are found, they often do not cover the full time-span of a site, or sampling them for DNA may not be possible. Here we present targeted enrichment and sequencing of hominin nuclear DNA from sediments, and insights into human history derived from this DNA. We developed methods to capture hominin DNA even in the presence of homologous faunal DNA, and evaluate the extent of microbial and faunal DNA in our data. We applied these methods to sediment samples from Galería de las Estatuas (Arsuaga et al., 2017), a site in northern Spain, and Denisova (Jacobs et al., 2019) and Chagyrskaya (Kolobova et al., 2020) caves, in the Altai Mountains in southern Siberia, and identified and sequenced Neandertal nuclear DNA in stratigraphic layers spanning 55k – 200 thousand years ago. We then placed each sample on the Neandertal phylogenetic tree, inferring the most likely divergence date from a lineage. In Estatuas we demonstrate a population transition, and associate this change with specific layers. In Chagyrskaya, all layers are associated with a single Neandertal lineage, suggesting a more homogenous occupation. This work demonstrates that detailed genetic analyses may be possible from many more archaeological sites than was previously thought, and is particularly encouraging for time-series studies of single sites, or for sites with a sparse fossil record.

OS4-7

Genome analysis traces regional dispersal of rice in Taiwan and Southeast Asia

Ornob Alam[1], Rafal M Gutaker[2,1], Cheng-chieh Wu[3,4], Karen Hicks[5], Kyle Bocinsky[6], Cristina C Castillo[7], Steven Acabado[8], Dorian Q Fuller[7,9], Jade A d'Alpoim Guedes[10], Yue-ie Hsing[3], Michael D Purugganan[1]
[1]New York University, New York, NY, USA. [2]Royal Botanic Garden, Kew, Richmond, London, United Kingdom. [3]Academia Sinica, Nankang, Taipei, Taiwan. [4]National Taiwan University, Taipei, Taiwan. [5]Kenyon College, Gambier, OH, USA. [6]Crow Canyon Archaeological Center, Cortez, CO, USA. [7]University College London, London, United Kingdom. [8]University of California, Los Angeles, Los Angeles, CA, USA. [9]North-West University, Xi'an, China. [10]University of California, San Diego, San Diego, CA, USA

Abstract

The dispersal of rice (Oryza sativa) following domestication influenced massive social and cultural changes across South, East, and Southeast Asia. The history of dispersal across islands of Southeast Asia, and the role of Taiwan and the Austronesian expansion in this process remain largely unresolved. Here, we reconstructed the routes of dispersal of O. sativa ssp. japonica rice through Taiwan and the northern Philippines using whole-genome re-sequencing of indigenous rice landraces coupled with archaeological and paleoclimate data. Our results indicate that japonica rice found in the northern Philippines diverged from Indonesian landraces as early as 3500 BP. In contrast, rice cultivated by the indigenous peoples of the Taiwanese mountains has complex origins. It comprises two distinct populations, each best explained as a result of admixture between temperate japonica that presumably came from northeast Asia, and tropical japonica from the northern Philippines and mainland Southeast Asia respectively. We find that the temperate japonica component of these indigenous Taiwan populations diverged from northeast Asia subpopulations at about 2600 BP, while gene flow from the northern Philippines occurred before ~1300 years BP. This coincides with a period of intensified trade across the South China Sea. Finally, we find evidence for positive selection acting on distinct genomic regions in different rice subpopulations, indicating local adaptation associated with the spread of japonica rice.

OS4-8

Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2 mRNA vaccinated individuals

Talia Kustin[1,2], Noam Harel[1,2], Uriah Finkel[3], Shay Perchik[3], Sheri Harari[1,2], Maayan Tahor[1], Itamar Caspi[1], Rachel Levy[4], Michael Leschinsky[3], Shifra Ken Dror[5], Galit Bergerzon[5], Hala Gadban[5], Faten Gadban[5], Eti Eliassian[6], Orit Shimron[6], Loulou Saleh[7], Haim Ben-Zvi[7], Doron Amichay[8,9], Anat Ben-Dor[8], Dana Sagas[10], Merav Strauss[10], Yonat Shemer Avni[10,11], Amit Huppert[12,13], Eldad Kepten[3], Ran D. Balicer[3], Doron Nezer[14], Shay Ben-Shachar[3,13], Adi Stern[1,2]

[1]The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [2]Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel. [3]Clalit Research Institute, Innovation Division, Clalit Health Services, Ramat Gan, Israel. [4]Tel Aviv University, Tel Aviv, Israel. [5]Clalit Health Services, Central Laboratories, Haifa and Western Galilee, Israel. [6]Progenin Laboratories, Jerusalem District, Clalit Health Services, Jerusalem, Israel. [7]Microbiology lab, Rabin Medical Center, Beilinson Hospital, Petach Tikva, Israel. [8]Central Laboratory, Clalit Health Services, Tel Aviv, Israel. [9]Department of Clinical Biochemistry and Pharmacology, Faculty of Health Sciences, Ben-Gurion University, Beersheba, Israel. [10]Microbiology Laboratory, Emek Medical Center, Afula, Israel. [11]Faculty of Health Sciences, Ben Gurion University of the Negev, Beersheba, Israel. [12]The Bio-statistical and Bio-mathematical Unit, The Gertner Institute for Epidemiology and Health Policy Research, Chaim Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel. [13]The Sackler Faculty of Medicine, Tel-Aviv University,, Tel Aviv, Israel. [14]Clalit Health Services, Tel Aviv, Israel

Abstract

The BNT162b2 mRNA vaccine is highly effective against SARS-CoV-2. However, apprehension exists that variants of concerns (VOCs) may surmount vaccine protection, following evidence showing reduced neutralization of VOCs B.1.1.7 and B.1.351 in laboratory assays. We performed a matched cohort study that examined the distribution of SARS-CoV-2 variants conditioned on infection of vaccinated individuals (“breakthrough cases”), compared to matched infections of unvaccinated individuals. We hypothesized that if there is lower vaccine effectiveness against one of the VOCs, its proportion among the breakthrough cases should be higher than among unvaccinated cases. Our results show that vaccinees that tested positive at least a week after the second dose were indeed disproportionally infected with B.1.351, as compared with unvaccinated individuals. Those who tested positive between two weeks after the first dose and one week after the second dose, were disproportionally infected by B.1.1.7 suggesting reduced vaccine effectiveness against both VOCs at particular time windows following vaccination. Nevertheless, vaccine effectiveness remains high among those fully vaccinated and the B.1.351 frequency in Israel to-date remains low. From an evolutionary point of view, it is possible that immune evasion mutations of B.1.351 incur a fitness cost in the form of reduced transmissibility, especially as compared to the highly transmissible B.1.1.7. Our results emphasize the importance of tracking viral variants in a rigorous framework and of increasing vaccination, which we conclude is the safest and most effective means of preventing the onwards spread of B.1.351 and other possible future VOCs.

OS5-1

History Repeats Itself: Using evolutionary convergence to reveal adaptations and genome-wide functional networks

Nathan L Clark[1], Wynn Meyer[2], Raghav Partha[3], Jerrica Jamison[4], Jaret Lieberth[1], Jason Presnell[1], Jiaxuan Yang[5,6], Maria Chikina[5]

[1]University of Utah, Salt Lake City, Utah, USA. [2]Lehigh University, Bethlehem, PA, USA. [3]Grail, San Francisco, CA, USA. [4]University of Toronto, Toronto, ON, Canada. [5]University of Pittsburgh, Pittsburgh, PA, USA. [6]Tsinghua University, Beijing, China

Abstract

Life is in a constant state of revision in response to evolutionary pressures such as environmental change. Our goal is to discover genetic changes underlying adaptations by leveraging convergent evolution, a phenomenon in which independent lineages evolve the same phenotype. We employ our novel computational approaches to identify genes and noncoding regions important for convergent phenotypes ranging from aquatic adaptations to long lifespan. We first present evidence that unrelated lineages of marine mammals experienced change in a shared gene set, whose functions are associated with an aquatic lifestyle. Genes with roles in lung and skin experienced convergently accelerated rates of change due to the requirements of diving and pathogens, while sensory system genes experienced convergently relaxed constraint. One gene, Paraoxonase 1, became a pseudogene repeatedly in aquatic mammals, likely to cope with oxidative stress brought by repeated diving, and its loss exposes these species to the neurotoxic effects of organophosphate pesticides. Other studies in subterranean mammals showed convergent accelerations in genes related to eyesight, as expected for these blind species. Since the relaxed selection is strong, it provides an opportunity to identify new genes and regulatory regions associated with ocular development. A genome-wide scan revealed hundreds of noncoding regions that accelerated only in blind species, and which we find to drive retinal expression in zebrafish embryos. Accordingly, we are sequencing patients with congenital eye diseases at these regions to locate causal mutations. In sum, studies of convergent evolution are advancing our understanding of adaptive evolution, gene regulation, and human health.

OS5-2

Lineage-specific genes: Evolutionary novelties or technical artifacts? How to tell and why it matters

Caroline M Weisman, Andrew W Murray, Sean R Eddy
Harvard University, Cambridge, MA, USA

Abstract

“Lineage-specific” genes appear to have homologs only in a restricted group of related species, strikingly absent from the rest of the tree of life. They are often interpreted as novel genes, receiving much interest related to their apparent potential to underlie evolutionary innovation. An alternative, “null” hypothesis is that these genes are not meaningfully novel: they do have existing homolog whose sequences have merely diverged too much to be detected by homology search. Here, we first develop a simple method to test whether a given lineage-specific gene can be explained by this possibility of “divergence beyond detection.” For most lineage-specific genes in animals, insects, and fungi, we find that this is indeed the case: no novelty is required to explain them. We also find some lineage-specific genes in all clades that strongly resist this explanation, flagging them as candidates for true genetic novelty. We then apply our tool to the intriguing case of the evolution of mesoderm, proposed to have evolved convergently in bilateria and ctenophores on the basis of the apparent lineage-specificity of many bilaterian genes involved in mesoderm determination. We show that these genes may actually have homologs in ctenophores missed by standard homology search; identify ctenophore genes that we think may be these missing homologs; and discuss ongoing experiments testing the homology and functions of these genes, with implications for the evolutionary origins of mesoderm.

OS5-3

Mutation rate and spectrum are selectively maintained in yeast

Haoxuan Liu, Jianzhi Zhang
University of Michigan, Ann Arbor, MI, USA

Abstract

Because the vast majority of mutations are deleterious, one wonders why the mutation rate μ has not been reduced by natural selection to zero. The prevailing hypothesis asserts that selection has pushed μ to the minimal level achievable in the presence of genetic drift, or the drift barrier, and therefore predicts an elevation of μ upon the removal of selection. Here we show that, contrasting this prediction, random mutation accumulation (MA) in yeast frequently lowers μ by a substantial fraction, and deleting the newly discovered mutator gene *PSP2* nearly halves μ. Comparing the MA with natural yeast strains demonstrates that μ is maintained at an intermediate level in nature by stabilizing selection. Consistently, μ exceeds the drift barrier in organisms ranging from bacteria to mammals, often by orders of magnitude. Our analysis further indicates that, the mutation spectrum, such as the universal AT mutational bias and transition to transversion mutational bias, is not intrinsic but has been shaped by selection. Together, these findings suggest evolutionary optimizations of mutation rate and spectrum, which blur the separation of mutation from selection as distinct evolutionary forces but open the door to decreasing mutagenesis in various organisms by genome editing.

OS5-4

Substantial and similar fitness effects of synonymous and nonsynonymous mutations in yeast

Xukang Shen, Siliang Song, Jianzhi Zhang
University of Michigan, Ann Arbor, Michigan, USA

Abstract

Synonymous mutations are commonly considered neutral or nearly so. Recent case studies using manipulative experiments, however, found appreciable fitness effects of some synonymous mutations. To investigate the generality of this phenomenon, we created 8348 yeast mutants each carrying one point mutation (22% synonymous) in one of 21 endogenous genes with diverse functions and expression levels, and measured the fitness of these mutants relative to the wild-type in a rich medium. Surprisingly, over 80% of synonymous mutations significantly lower the fitness, and the distribution of fitness effects (DFE) is overall similar between synonymous and nonsynonymous mutations. To uncover the mechanistic basis of the above finding, we measured the mRNA concentration of the mutated gene in each mutant. We found that both synonymous and nonsynonymous mutations frequently alter the mRNA level of the gene and that the expression change caused by a coding mutation positively correlates with its fitness effect. To explain why the vast majority of genes have a nonsynonymous to synonymous substitution rate ratio (d_N/d_S) that is substantially lower than 1 despite having similar DFEs of synonymous and nonsynonymous mutations, we propose nonsynonymous mutations have more variable fitness effects across environments than synonymous mutations; consequently, when the environment varies, a larger fraction of nonsynonymous than synonymous mutations are selectively purged. Our simulation and measurement of DEFs in additional environments support this hypothesis. We conclude that most synonymous mutations are strongly non-neutral, requiring reexamining numerous biological conclusions (e.g., about mutation, selection, and divergence time) relying on assuming neutral synonymous mutations.

OS5-5

Inferring selection against deleterious alleles using the magic of haplodiploidy

Sara E Miller[1], Michael Sheehan[2]
[1]Cornell University, Ithaca, NY, USA. [2]C, I, NY, USA

Abstract

Distinguishing deleterious variants from neutral genetic variation within populations is of major interest both for understanding disease and for predicting the evolutionary response of populations, but rapid identification of deleterious variants remains a challenge. In haplodiploid species, haploid males develop from unfertilized eggs and females develop from fertilized diploid eggs. In essence, the entire haplodiploid genome is comparable to sex chromosomes in diploid species. This naturally occurring variation in ploidy between haplodiploid sexes leads to a long-standing hypothesis that in males, deleterious alleles will be directly exposed to selection, but deleterious alleles can be masked from selection by partial or complete dominance in females. Because of stronger selection against haplodiploid males, we can predict that males will have fewer deleterious variants than females. Identifying which variants are differentially removed in haplodiploid males would therefore give insight into the evolutionary dynamics of deleterious mutations. In this study, we provide the first test of this hypothesis by comparing re-sequenced genomes of populations of male and female northern paper wasps (*Polistes fuscatus*). Matching our predictions, males had fewer deleterious variants than females. Differences in allele frequency between the sexes are non-randomly distributed throughout the genome with males showing fewer nonsense and missense mutations as well as a lower proportion of rare alleles. Our findings suggest that comparing allele frequencies between male and female haplodiploids can be a powerful method for detecting strongly deleterious recessive variants.

OS5-6

Rapid turn-over of centromere sequences in *D. melanogaster* and the *simulans* clade

Cécile Courret[1], Xiaolu Wei[1], Ching-Ho Chang[1], Lucas Hemmer[1], Bryce Santinello[2], Barbara Mellone[2], Amanda M Larracuente[1]
[1]University of Rochester, Rochester, NY, USA. [2]University of Connecticut, Storrs, CT, USA

Abstract

Centromeres are chromosomal structures that are indispensable for faithful genome inheritance during cell division. Centromeres are typically defined epigenetically by the presence of the centromere-specific histone H3 variant, CENP-A. Centromeres form in repeat-rich regions of the genome but the roles of DNA sequences in centromere function are unclear. The highly repetitive nature of centromeres presents a major challenge for genome assembly and thus for understanding centromeric DNA organization.

We recently revealed that all centromeres in *D. melanogaster* correspond to islands of complex DNA enriched in retroelements and flanked by tandem repeats. Each centromere is unique—the only sequence they have in common is the *G2/Jockey-3* retroelement. It is unclear if any of these sequences are important for centromere function. Here we study the evolution of centromere composition to gain insights into the role of DNA sequence in centromere biology. We characterized centromere organization in three sister species: *D. simulans, D. sechellia,* and *D. mauritiana*. We discovered that *simulans* clade centromeres have similar organization as *D. melanogaster*'s: islands of complex repeats flanked by tandem repeats. However, none of the *D. melanogaster* centromere islands are conserved in the *simulans* clade. Instead, the *simulans* cladecentromeres are mainly enriched in complex satellites. *G2/Jockey-3* is enriched in *D. simulans* centromeres, but much less so in *D. sechellia* and *D. mauritiana.* Our results highlight the rapid turnover of centromeric sequences among the *simulans* clade species and *D. melanogaster*. Identifying the functional centromeric DNA will give insights into their roles in chromosome function and evolution.

OS5-7

Inferring Population Size Histories using Coalescent Hidden Markov Models with Tree Height and Total Branch Length as Hidden States

Gautam Upadhya, Matthias Steinruecken
University of Chicago, Chicago, IL, USA

Abstract

Unraveling complex demographic histories of populations is a central problem in population genetics. Past demographic events are of general anthropological interest, but understanding them is also important for establishing accurate null models when identifying adaptive or disease-associated genetic variation. Coalescent Hidden Markov Models (CHMMs) are an important class of tools for inferring past population size changes. These models analyze linkage information in population genomic datasets by using the local genealogies relating sampled individuals as latent states in an HMM framework. Applying these models to large sample sizes is challenging since the number of possible genealogies becomes large. We present CHIMP (CHMM History-Inference ML Procedure), a novel CHMM method for inferring population size histories. It scales well for large sample sizes and only requires unphased genomes as input. We present two implementations that, respectively, use either the height of the genealogical tree (TMRCA) or its total branch length (L) as the latent variable of the CHMM. Requisite transition and emission probabilities are each obtained by numerically solving systems of differential equations derived from the respective ancestral process with recombination or mutation. Population size history parameters are inferred using an Expectation-Maximization algorithm where a composite likelihood scheme allows scaling to large sample sizes. We demonstrate the efficiency and accuracy of our method in a variety of benchmark tests using simulated data and present comparisons to other state-of-the-art methods. The TMRCA implementation of CHIMP performs comparably well and provides accurate population size estimates in intermediate and ancient times.

OS5-8

Strong within-host selection in a maternally inherited obligate symbiont: *Buchnera* and aphids

Julie Perreau[1], Bo Zhang[2,1], Gerald P Maeda[1], Mark Kirkpatrick[1], Nancy A Moran[1]
[1]University of Texas at Austin, Austin, TX, USA. [2]Chinese Academy of Agricultural Sciences, Chinese Academy of Agricultural Sciences, China

Abstract

Numerous animal lineages have maternally inherited symbionts that are required for host reproduction and growth. Endosymbionts also pose a risk to their hosts, due to mutational decay of their genomes through genetic drift, or to selfish mutations that favor symbiont fitness over host fitness. One model for heritable endosymbiosis is the association of aphids with their obligate bacterial symbiont, *Buchnera*. We experimentally established heteroplasmic pea aphid matrilines, containing pairs of closely-related *Buchnera* haplotypes, and used deep sequencing of diagnostic markers to measure haplotype frequencies in successive host generations. These frequencies were used to estimate the effective population size of *Buchnera* within hosts, i. e. the transmission bottleneck size, and the extent of within-host selection. The within-host effective population size was in the range of 10 to 20, indicating strong genetic drift and the potential for fixation of deleterious mutations. Remarkably, closely related haplotypes were subject to strong within-host selection, with selection coefficients as large as 0.5 per aphid generation. In one case, the direction of selection depended on the thermal environment, and went in the same direction as between-host selection. In another, a new mutant haplotype had a strong within-host advantage under both environments, but had no discernible effect on host-level fitness under laboratory conditions. Thus, within-host selection can be strong, resulting in rapid fixation of mutations with little impact on host-level fitness. Together, these results show that within-host selection can drive evolution of an obligate symbiont, accelerating sequence evolution.

Room 1-01

Effects of *EGLN1* haplotypes on hemoglobin concentration in Andean highlanders

Yoshiki Yasukochi[1], Takayuki Nishimura[2], Mayumi Ohnishi[3], Mika Nishihara[3], Juan Ugarte[4], Hideki Fukuda[5], Kiyoshi Aoyagi[3]

[1]Mie University, Tsu, Mie, Japan. [2]Kyushu University, Fukuoka, Fukuoka, Japan. [3]Nagasaki University, Sakamoto, Nagasaki, Japan. [4]Universidad Mayor de San Andres, La Paz, Bolivia, Plurinational State of. [5]4. National Institute of Public Health, Wako, Saitama, Japan

Abstract

Several studies have explored genetic or physiological factors related to high-altitude adaptation in modern humans. It is reported that hemoglobin (Hb) concentrations of Andean highlanders are relatively higher than those of other highlander populations. Recent genome-wide scans have independently detected signals of positive natural selection around the *EGLN1* genomic region in Tibetan and Andean highlanders. However, to the best of our knowledge, there is no evidence suggesting that *EGLN1* single nucleotide polymorphisms (SNPs) actually affect Hb concentrations in the Andeans. Therefore, we investigated the association between four tag SNPs (rs2486740, rs508618, rs12097901, and rs1769792), covering a ~81-kb genomic region around *EGLN1*, and hematological traits in 99 healthy young men and women living in La Paz, Bolivia. Physiological data were obtained from participants of two universities located at altitudes of ~3700 and ~4000 m in 2016. Using genomic DNA extracted from saliva, genotypes of the four SNPs were determined by PCR–direct sequencing or TaqMan assays. To assess combined effects of the four SNPs on Hb levels, we estimated phased haplotypes using an IMPUTE2 program. Our findings indicated that frequencies of the estimated haplotypes increased with Hb concentrations in Bolivians. Based on analyses of 1000 Genomes Project data, we also found that a certain haplotype related to high Hb levels may have rapidly expanded in the ancestral populations of Bolivian and Peruvian individuals.

Room 1-02

Limited introgression from Greater Crested Tern into critically endangered species Chinese Crested Tern

Lanhui Peng
State Key Laboratory of Biocontrol, Department of Ecology, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong Province, China

Abstract

Interspecific hybridization and introgression could be threatening to endangered species. Chinese Crested Tern (*Thalasseus bernsteini*) is a critically endangered species with population fewer than one hundred individuals, breeding within large colonies of Greater Crested Tern (*Thalasseus bergii*). Mating between the two species has been recorded and hybridization has been verified using molecular techniques. To investigate whether there is recent introgression between the two species, we analyzed high-quality whole genome data of 13 Chinese Crested Tern and 21 Greater Crested Tern from Jiushan and Wuzhishan archipelagos, Zhejiang Province, China. Principal component analysis and admixture analysis showed no population structure within sampled individuals of each species. RFMix was used to infer local ancestry structure of sampled individuals. We detected one presumably recent hybrid individual in Greater Crested Tern, with about 6% of genetic material shared with Chinese Crested Tern, possibly having a Chinese Crested Tern ancestral four generations ago. Aside from this individual, very low level of genetic sharing between the two species (<1%) was detected, likely not resulting from recent introgression. To sum up, we confirmed that recent hybridization and backcrossing into Greater Crested Tern has taken place, while recent introgression into Chinese Crested Tern has not been detected. Considering the remarkably different population size of the two species, backcrossing into Chinese Crested Tern is less likely to happen, but the possible introgression from Greater Crested Tern is still a matter of concern. For actual conservation actions, we recommend intensive monitoring on hybrids and prohibit backcrossing into parent populations.

Room 1-03

Two divergent haplogroups of a sacsin-like gene trace back to the origin of Acroporidae corals

Shiho Takahashi-Kariyazono, Yohey Terai
SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa, Japan

Abstract

Reef-building corals are declining due to environmental changes such as ocean acidification and rising sea temperatures. Sacsin is a member of the heat shock proteins and has been reported as a candidate protein associated with the stress response in *Acropora* corals. Recently, high nucleotide diversity and the persistence of two divergent haplogroups of sacsin-like genes in *Acropora millepora* have been reported, but the origin of genetic variation in these haplogroups has not been studied.

In this study, we analyzed a genomic region containing a sacsin-like gene from *Acropora* and *Montipora* species. Higher nucleotide diversity in the sacsin-like gene compard to surrounding regions was also observed in *A. digitifera*. This nucleotide diversity is drived from two divergent haplogroups of a sacsin-like gene that is present in at least three *Acropora* species. The origin of these two haplogroups can be traced back before the divergence of the genera *Acropora* and *Montipora*, and the polymorphism of these haplogroups has persisted in the family Acroporidae. Although the link between genetic variation in sacsin-like genes and functional differences in sacsin-like proteins is not clear, it is possible that divergent haplogroups have responded differently to envionmental stress factors during the evolution of reef-building corals and now serve important roles in the adaptive phsiological ecology of this group of keystone species.

Room 1-04

Adaptation of A-to-I RNA editing in *Drosophila*

Yuange Duan, Shengqian Dou, Shiqi Luo, Hong Zhang, Jian Lu
Peking University, Beijing, China

Abstract

Adenosine-to-inosine (A-to-I) editing is hypothesized to facilitate adaptive evolution by expanding proteomic diversity through an epigenetic approach. However, it is challenging to provide evidences to support this hypothesis at the whole editome level. In this study, we systematically characterized 2,114 A-to-I RNA editing sites in female and male brains of *D. melanogaster*, and nearly half of these sites had events evolutionarily conserved across *Drosophila* species. We detected strong signatures of positive selection on the nonsynonymous editing sites in *Drosophila* brains, and the beneficial editing sites were significantly enriched in genes related to chemical and electrical neurotransmission. The signal of adaptation was even more pronounced for the editing sites located in X chromosome or for those commonly observed across *Drosophila* species. We identified a set of gene candidates that had nonsynonymous editing events favored by natural selection. We presented evidence that editing preferentially increased mutation sequence space of evolutionarily conserved genes, which supported the adaptive evolution hypothesis of editing. We found prevalent nonsynonymous editing sites that were favored by natural selection in female and male adults from five strains of *D. melanogaster*. We showed that temperature played a more important role than gender effect in shaping the editing levels, although the effect of temperature is relatively weaker compared to that of species effect. We also explored the relevant factors that shape the selective patterns of the global editomes. Altogether we demonstrated that abundant nonsynonymous editing sites in *Drosophila* brains were adaptive and maintained by natural selection during evolution. Our results shed new light on the evolutionary principles and functional consequences of RNA editing.

Room 1-05

Assessing population connectivity of New Zealand stream insects using RADseq SNPs and COI markers

Vanessa A. Barbosa[1], Angela McGaughran[1], Ian D. Hogg[1,2], S. Elizabeth Graham[3]

[1]School of Science, University of Waikato, Hamilton, New Zealand. [2]Polar Knowledge Canada, Canadian High Arctic Research Station, Cambridge Bay, Canada. [3]National Institute of Water and Atmospheric Research – NIWA, Hamilton, New Zealand

Abstract

Genetic markers have been increasingly used as an indirect measure of dispersal in freshwater insects. We investigated the population structure of three stream insect species: the mayfly *Coloburiscus humeralis*, the stonefly *Zelandobius confusus* and the caddisfly *Hydropsyche fimbriata*. Using SNP data from ezRAD-seq and mtDNA (cytochrome C oxidase subunit 1, COI), we compared the performance of both marker types. Our analysis included samples from within and among streams from forested and fragmented habitats of mountain ranges of New Zealand's North Island. Both markers indicated a consistent lack of strong population structure for the three species across the study area. Within a single region, there was no clear subdivision of populations between sampling sites or stream localities. Evidence of genetic differentiation was more frequent at larger spatial scales when comparing populations from different regions. At smaller spatial scales (e.g. between neighbouring streams), significant genetic differentiation was found for a few pairwise comparisons in *H. fimbriata* indicating that gene flow might be limited in this species. Together, our data suggest that dispersal ability is greater in *C. humeralis* than *Z. confusus* and *H. fimbriata*. We conclude that RADseq-based SNP will provide a useful method for assessment of fine-scale population structure in aquatic insects.

Room 1-06

The vast majority of mammalian circular RNAs are junk RNAs

CHUAN XU[1,2], JIANZHI ZHANG[3]

[1]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China. [2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Mi, USA. [3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Abstract

Ubiquitous in eukaryotes, circular RNAs (circRNAs) comprise a large class of mostly non-coding RNAs produced by back-splicing, which covalently links a downstream splice-donor site to an upstream splice-acceptor site. Although some circRNAs have demonstrated biochemical activities, whether most circRNAs are functional is unknown. Here we test the hypothesis that circRNA production largely results from splicing error so is deleterious instead of beneficial. The error hypothesis predicts that (1) back-splicing is much rarer than linear-splicing, (2) back-splicing diminishes with the splicing amount because of intensified selection against splicing error, (3) the overall prevalence of back-splicing in a species declines with its effective population size, and (4) circRNAs are evolutionarily unconserved. All of these predictions are verified using RNA sequencing data of 11 tissues each from three mammals, and over 97% of observed circRNA production is estimated deleterious. Hence, circRNAs should generally be considered nonfunctional products of splicing error, or junk RNAs.

Room 1-07

Structome: Exploring the structural neighbourhood of proteins

Ashar Malik[1], Jane Allison[2], Anthony Poole[2]

[1]Bioinformatics Institute, Agency for Science, Technology and Research, Singapore, Singapore. [2]School of Biological Sciences, University of Auckland, Auckland, New Zealand

Abstract

Evolutionary relationships are conventionally uncovered using protein sequences. Protein structure as opposed to the sequence can hold evolutionary signals over longer timescales and can therefore prove useful towards uncovering deep evolutionary relationships. To this end, Structome has been developed as a resource for researchers to quickly determine the structural neighbourhood of a query structure. The structural neighbourhood comprises protein structures within a certain user selected structural similarity cutoff. This resource, therefore allows the inspection of the neighbourhood of a query protein structure from which inferences can be made about the evolutionary relationships, through the use of phylogenetic networks. Domain annotation from SCOP and CATH databases are also provided, to allow users to validate their observations, along with sequence similarity. Covering ~70% of the proteins in RCSB PDB, Structome is a comprehensive tool for the analysis of the protein structure landscape.

Room 1-08

Selection and introgression shape heterogeneous genomic differentiation patterns in two incipient Chinese shorebird species

Wenjun Zhou, Hongzhou Lin, Jie Tu, Chenqing Zheng, Yang Liu
State Key Laboratory of Biocontrol, School of Life Sciences/School of Ecology, Sun Yat-sen University, Guangzhou, Guangdong, China

Abstract

Deciphering causes of differentiated genomic patterns has been one of the most important tasks in population and comparative genomics studies. In recent years, rapidly accumulating evidence has shown that introgression events alone cannot explain heterogeneous genomic landscape, but with many other forces such as selection and variation of recombination rates acting together. However, to what extent each of aforementioned forces have happened to shape the genomic patterns of species at an early stage of divergence is rarely testified. In this study, we performed whole genome sequencing for 58 individuals of Kentish plover (KP, *Charadrius alexandrinus*) and 43 individuals of White-faced plover (WFP, *Charadrius dealbatus*), two closely related shorebird species that are still at an early diverging stage. Using principal component analysis and ADMIXTURE, we show extensive introgression between the coastal populations of KP and WFP both along the coastal and Taiwan Island populations, which is subsequently confirmed by ABBA-BABA tests. Divergent estimation using F_{ST} and d_{xy} calculations shows regions of relatively high divergence level on chromosome 1 and chromosome Z. Using *nSL* test, we found evidence that selection on different parts of the genomes may be also acting as a force to shape heterogeneous genomic landscapes. Together, our research performed analysis on two closely related shorebird species, tested the existence and traces of introgression events and potential selection forces, and provided an insight into the formation of genomic islands in early divergent species.

Room 1-09

Shifts in mutation spectra enhance access to beneficial mutations

Mrudula Sane[1], Gaurav D Diwan[1,2], Bhoomika A Bhat[1,3], Lindi M Wahl[4], Deepa Agashe[1]
[1]National Centre for Biological Sciences, Bangalore, India. [2]University of Heidelberg, Heidelberg, Germany. [3]Indian Institute of Science, Bangalore, India. [4]Western University, Ontario, Canada

Abstract

Biased mutation spectra are pervasive, with widely varying direction and magnitude of mutational biases that influence genome evolution and adaptation. Why are unbiased spectra rare, and how do such diverse biases evolve? Our experiments show that changing the mutation spectrum allows populations to sample previously under-sampled mutational space. The resulting shift in the distribution of fitness effects is advantageous: the beneficial mutation supply and beneficial pleiotropy increase, and deleterious load reduces. More broadly, adaptive walk simulations indicate that the evolution of a mutational bias in an unbiased ancestor is selectively neutral; but reversing the direction of a long-term bias is always selectively favoured. Indeed, spectrum changes in the bacterial phylogeny occur frequently, typically involving reversals of ancestral bias. Thus, shifts in mutation spectra evolve under selection, and can directly alter outcomes of adaptive evolution by facilitating access to beneficial mutations.

Room 1-10

Characterization of the amphioxus insulin-like peptide (Ilp) gene – insights into evolution of the vertebrate insulin/IGF pathway

Yichen Dai[1], Guang Li[2], Xuming Zhou[1]

[1]Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [2]Xiamen University, Xiamen, China

Abstract

Insulin and insulin-like growth factors (IGFs) are essential signaling molecules responsible for regulating essential processes such as dietary metabolism and body growth. Vertebrate insulin and IGF proteins belong to the same protein family, but the evolutionary origin of these signaling molecules has remained unclear. Here, we focus on the cephalochordate amphioxus (‘lancelets’), living members of a vertebrate sister lineage. Previous studies have identified two *insulin-like peptide* (*Ilp*) genes in the Florida lancelet (*Branchiostoma floridae*) genome, of which *Ilp1* expression is detected in the embryo and persists into adulthood. We generated a transgenic Florida lancelet line carrying a frameshift mutation in *Ilp1*, and we show that $Ilp1^{-/-}$ individuals cannot survive past 16 days post fertilization and are smaller in body size compared to $Ilp1^{+/-}$ and $Ilp1^{+/+}$ siblings. However, a phenotypic difference is not observable in the embryo, and we modified a single cell sequencing protocol to perform transcriptomic sequencing of single embryos obtained from a cross between a $Ilp1^{+/-}$ female and $Ilp1^{+/-}$ male. In $Ilp1^{-/-}$ embryos, there is significant up-regulation of genes involved in lipid metabolism at the early and mid-neurula stage compared to $Ilp1^{+/+}$ embryos, and in 1- and 3-gill slit $Ilp1^{-/-}$ individuals, genes involved in lipid metabolism, DNA damage and oxidative stress response are expressed at different levels compared to $Ilp1^{+/+}$ individuals. These results are similar to patterns observed in other animals under low insulin drive: a metabolic shift towards lipid oxidation accompanied with cellular stress response, suggesting that lancelet *Ilp1* is functionally similar to vertebrate insulin and IGF.

Room 1-11

Hsp90 in Mutational Robustness

Fu-Hsuan Ko[1], Po-Hsiang Hung[2], Jun-Yi Leu[1]

[1]Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan. [2]Department of Genetics, Stanford University, Stanford, California, USA

Abstract

Mutational robustness is a property that the phenotype of biological systems can remain constant regardless of genetic variation. One proposed mechanism of robustness is via molecular chaperones, which assist protein folding to make mutated proteins retain functions. Hsp90, a molecular chaperone, has been found to buffer natural genetic variation in multiple model organisms. However, which mutations can be buffered and the underlying molecular mechanisms still remain elusive. Here, using budding yeast Saccharomyces cerevisiae as a model, we applied EMS mutagenesis to generate Hsp90-buffered mutant strains. By combing NGS-based bulk segregant analysis and CRISPR-Cas9 gene editing, we identified that the deleterious phenotypic outcome of *acs2*, encoding acetyl-CoA synthetase, can be partially buffered by Hsp90. In the normal condition, *acs2* leads to decreased fitness; while in Hsp90 inhibition or high-temperature condition, *acs2* is lethal. Our research provides direct evidence that Hsp90 plays a role in robustness against the effects of harmful mutations.

Room 1-12

Genetic variation of olfactory receptor repertoire across human populations

Muhammad S Akhtar[1], Ryuichi Ashino[1], Yoshihito Niimura[2], Kazushige Touhara[3], Amanda D Melin[4,5], Shoji Kawamura[1]

[1]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan. [2]Department of Veterinary Sciences, Faculty of Agriculture, University of Miyazaki, Miyazaki, Miyazaki, Japan. [3]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Tokyo, Japan. [4]Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada. [5]Department of Anthropology and Archaeology, University of Calgary, Calgary, Alberta, Canada

Abstract

The human genome contains roughly 400 intact and 440 pseudogenized olfactory receptor (OR) genes. However, the intact/pseudogene composition of the entire OR gene family among human populations remains unclear. This is largely due to reliance on whole genome sequence data, which are often incomplete, especially for multigene families. We employed a target capture approach, probing OR genes followed by massive-parallel sequencing. Study populations included 17 ethnic groups from African-, European- and Asian-origins with diverse historical subsistence. We designed probes for 554 OR genes based on the human reference genome hg38 and the chimpanzee genome Pantro3.0. Targeted genes are 398 annotated intact OR genes in hg38, 4 “alt” intact OR sequences in hg38, 99 OR “fresh” pseudogenes in hg38, 53 chimpanzee intact OR genes of which orthologs are absent in hg38. An OR gene is considered intact if it has an ORF $\geq$ 250 amino acids and can make a 7-transmembrane structure. Target capture improved intact/pseudogene distinction and revealed population specific intact/pseudogene polymorphism in 134 OR genes. Out of these 134genes, 5 genes are annotated pseudogenes in human reference genome hg38 and 1 OR gene is absent in hg38 before but intact in chimpanzee. In addition to the intact/pseudogene polymorphism, many genes showed copy number variation across populations. Many of the nucleotide and copy number variations were found to be associated with either a single population or within one geographic region.

Room 1-13

Evolutionary responses to codon usage of horizontally transferred genes in *Pseudomonas aeruginosa*

Martijn Callens[1], Celine Scornavacca[2], Stéphanie Bedhomme[1]
[1]CEFE, Montpellier, France. [2]ISEM, Montpellier, France

Abstract

Prokaryote genome evolution is characterized by the frequent gain of genes through horizontal gene transfer (HGT). For a gene, being horizontally transferred can represent a strong change in its genomic and physiological context. If the codon usage of a transferred gene deviates from that of the receiving organism, the translation rates and fitness benefits it provides can be reduced due to a mismatch with the expression machinery. Consequently, transferred genes with a deviating codon usage can be selected against or elicit evolutionary responses that might enhance their translation such as gene amelioration and compensatory evolution. Within bacterial species, the extent and relative importance of these different mechanisms has never been considered altogether. We used a phylogeny-based method to investigate the occurrence of these different evolutionary responses in *Pseudomonas aeruginosa*. Selection on codon usage of genes acquired through HGT was observed over evolutionary time, with the overall codon usage converging towards that of the core genome. Gene amelioration, through the accumulation of synonymous mutations after HGT, did not seem to systematically affect transferred genes. This pattern therefore seemed to be mainly driven by selective retention of transferred genes with an initial codon usage similar to that of the core genes. Additionally, variation in the copy number of tRNA genes was often associated with the acquisition of genes for which the observed variation could enhance their translation. This provides evidence that compensatory evolution might be an important mechanism for the integration of horizontally transferred genes.

Room 1-14

Evolution of gene-network involved in drought tolerance of wild tomato Solanum chilense

Kai Wei[1], Gustavo Silva-Arias[1], Saida Sharifova[2], Aurélien Tellier[1]

[1]Professorship for Population Genetics, Department of Life Science Systems, School of Life Sciences, Technical University of Munich, Freising, Germany. [2]Khazar University, Baku, Azerbaijan

Abstract

The wild tomato species *Solanum chilense* occurs at the southern edge of the entire tomato clade distribution range covering habitats from sea level up to 3500 m of altitude surrounding the Atacama desert. S. chilense populations are challenged by prolonged drought making it an ideal model for the discovery of mechanisms to tolerate drought stress. We performed a two-pronged approach to infer the evolution of drought resistance gene-networks in *S. chilense.* First, we characterized gene-networks using RNA-seq expression analyses from leaf and shoot apex 16 libraries from a coastal population (LA1963) grown under normal watering (control) and drought stress experimental conditions. We identified 2,484 differentially expressed genes (DEGs) showing highly significant values. Weighted gene co-expression network analysis revealed two different gene modules significantly correlated with the control/drought conditions and enriched metabolism to water deprivation and genetic-environment response GO terms or KEGG pathways. We identified transcription factors (TFs) and predicted their respective transcription factor binding sites (TFBSs) for each of the drought-related modules. We inferred specific gene co-expression networks as sets of TFs as hub (key) genes that regulate the expression of downstream set of co-expressed genes with TFBSs targeted by same key TFs. We further studied the evolution of the identified gene networks using whole-genome data by comparing positive selection signals between genes and promoter regions considering their position in the gene-networks. Our results suggested that integrating transcriptomics and genomics analyses can facilitate the understanding of the emergence of mechanisms for drought tolerance.

Room 1-15

A differentiated methylation region (DMR) acts as a potential cis-regulatory element in controlling seasonal expression of androgen receptor gene in the Norway rats

Yaqi Ying, Yufang Zhou, Ning Li, Dawei Wang, Xiao-Hui Liu, Ying Song
Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

Abstract

The Norway rats show an apparent seasonal reproduction mode in North China, but not in South China. The androgen receptor (AR) signaling is essential for spermatogenesis and testicular development in male animals. By comparing the sperm methylome of the Norway rats between the North and South populations in China, we identified a differential methylation region (DMR) in the first intron of the AR gene. We inferred one (AT) and two haplotypes (GA and AT) in the South and North populations, respectively. The two haplotypes showed significantly different regulatory activities in the double luciferase gene reporter system. By analyzing the expression of Ar and the methylation status of the DMR in different sizes of rat testes in spring and autumn in North China, we found a significant seasonal difference in Ar expression in rats with small testes (<0.3 g) but not with larger testes, in which Ar was expressed at the highest level in autumn. Our further data showed the spermatogenesis was inhibited during the first meiosis of the small testes in autumn. Analysis of the relationship between the DMR haplotypes and Ar expression suggested a significant positive correlation between the methylation level and expression level in the rats with GA haplotypes, indicating a potential function of the GA haplotype as a suppressor in regulating the Ar seasonal expression. Our findings suggest the Ar plays a vital role in the inhibition of testes in the non-breeding season, and the regulatory activities of the DMR region of Ar depend on its haplotypes.

Room 1-16

Extracting Gene Neighbourhoods from Complex Metagenomic Assembly Graphs

Somayeh Kafaie, Robert Beiko, Finlay Maguire
Dalhousie University, Halifax, Nova Scotia, Canada

Abstract

Antimicrobial resistance (AMR) is one of the biggest global health challenges of our time with a role in an estimated ~700,000 annual deaths. The genomic context of AMR genes such as mobile genetic elements, gene clusters, and regulatory elements play a key role in their evolution and transmission via lateral gene transfer. Metagenomic sequencing bypasses culturing and directly profiles the genetic content of a microbial community, offering a rapid assessment strategy for AMR profiling. Assembly is an important step in metagenome profiling, but assembled contigs present an oversimplified view that can miss valid genomic context information. In this study, we explored the structure of assembly graphs around target AMR genes and developed efficient algorithms to extract the genomic neighborhood of AMR genes from metagenomic data as a list of graph paths upstream and downstream of the AMR gene. Furthermore, we defined metrics to remove invalid paths that arise from assembly errors and only keep the valid paths. We validated our method on several simulated as well as real metagenomic datasets. Our results show that it outperforms contigs and the recently developed tools Metacherchant and Spacegraphcats in terms of precision and sensitivity. Based on our experimental results, our proposed method is a promising tool to extract the AMR context from metagenomic samples accurately and study the composition of resistomes leading to a better reconstruction of the AMR evolution.

Room 1-17

Origins and diversification of germline specific histone H2B variants

Pravrutha Raman[1], Callie Rominger[1,2], Janet Young[1], Antoine Molaro[3], Toshio Tsukiyama[1], Harmit Malik[1,4]

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Whitman College, Walla Walla, WA, USA. [3]Genetics, Reproduction and Development (GReD) Institute, Université Clermont Auvergne, Clermont-Ferrand, France. [4]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Histones and their post-translational modifications facilitate diverse chromatin functions in eukaryotes. Whereas replication-dependent histones (H2A, H2B, H3 and H4) primarily repackage genomes after genome replication, replication-independent histone variants have evolved to promote specialized chromatin functions including gene expression, genome stability and epigenetic inheritance. Some histone variants are conserved across eukaryotes while others are lineage-specific, providing unique opportunities to study innovations in chromatin function. We carried out comprehensive phylogenomic analyses in mammalian genomes to reveal the presence of at least six H2B variants that are expressed exclusively in the germline of mammals. In addition to three previously described variants (H2B.1, subH2B, and H2B.W), we discovered two new oogenesis-expressed variants — H2B.L and H2B.N — that originated in vertebrates and mammals respectively and are retained in most mammals including humans. Another variant H2B.M is paralogous to and undergoing recurrent gene conversion with H2B.W. All H2B variants display dramatic expansion and pseudogenization across mammals suggesting high evolutionary turnover. Although some H2B variants evolve more rapidly than ancestral H2B across mammals, all variants show strong purifying selection suggesting that H2B variants may have vital but potentially redundant functions. Sequence based structure analyses of new variants show H2B.L contains a conserved histone fold domain, whereas H2B.N has diverged significantly with a truncated C-terminus, predicting unconventional chromatin packaging and histone function. Together, our analyses reveal that unusual chromatin environments in the germline and zygote may have led to the selection of an expanded H2B repertoire whose functional dissection can reveal specialized chromatin function in mammalian lineages.

Room 1-18

The evolution of duplicated genes in the transcription-translation space reveals insights on their retention mechanisms

Simon Aubé, Lou Nielly-Thibault, Christian R Landry
Université Laval, Québec city, Québec, Canada

Abstract

Gene duplication is a major source of biological novelty. When both resulting paralogs are preserved, it frequently leads to two diverged proteins with different molecular functions. Early expression changes are often the first step of this divergence and have been extensively studied. However, few studies have considered their transcriptional and translational components. To better understand how duplicates diverge at these two levels, we leveraged a published dataset of transcription and translation rates for the yeast *Saccharomyces cerevisiae*. Comparisons within 500 high-confidence paralog couples showed that relative divergence is 1.5 times greater in transcription than in translation. To elucidate which constraints could explain this divergence bias, we constructed a minimal model of post-duplication evolution in which the cumulative expression of two identical paralogs is under selection, while mutations randomly affect the transcription or translation rates of either copy. Stochastic simulations of this model showed that it can faithfully reproduce the empirical divergence patterns if combined with a mutational target size difference of about three-fold between transcription and translation rates. To improve the minimal model, we added constraints associated with the minimization of transcription cost as well as of expression noise. This impaired the simulations' ability to reproduce the empirical divergence, with or without the transcriptional bias of mutations. These results suggest that expression divergence with maintenance of cumulative protein abundance – a process known as absolute dosage subfunctionalization – has played a major role in the evolution of yeast paralogs and may have shaped their functional divergence.

Room 1-19

N-terminal extensions using non-AUG start codons identified under nutrient limited growth conditions.

Pieter Spealman[1], Carolina Santana[2], David Gresham[1]
[1]New York University, New York, NY, USA. [2]Universidade Federal da Bahia, Salvador, Bahia, Brazil

Abstract

N-**t**erminal **e**xtensions (**NTE**s) are protein isoforms generated by **a**lternative **t**ranslation **i**nitiation **s**ites (**aTIS**s) located upstream and in-frame to protein-coding genes. Because they lack an intervening stop codon translation results in additional peptides added to downstream proteins. NTEs are known to alter patterns of subcellular localization, post-translational regulation, and substrate affinity. While the exact biological function of NTEs is variable, several lines of evidence suggest a role in cellular adaptation to stress. Because aTISs can be non-canonical (non-AUG) start codons they are difficult to identify by analysis of genome sequences alone.

To test the effects of stress on NTE induction, we grew Saccharomyces cerevisiae under homeostatic nitrogen-limiting conditions. We performed RNAseq, Riboseq, and mass spectrometry. These were compared to publicly available data that includes both rich media and acute stress induced by rapamycin.

Here, we identify NTEs using NTEseqr, a reading frame sensitive analysis that searches transcript leaders upstream of protein coding genes to identify significant patterns of ribosome footprints. We found NTEs at over 5 times higher frequency in the nitrogen limited condition (66 NTEs) compared to rich media (13) and 3.5 times higher than acute stress (19). Under stress we find NTEs predominantly feature non-AUG start codons (UUG, AUA, AUU), consistent with altered start codon affinity as part of regulatory adaptation to stress. Genes with NTEs were significantly enriched in metabolic processes associated GO terms. We also identified peptides in the paired mass spectrometry data that map uniquely to NTEs, providing further evidence in support of their translation.

Room 1-20

Polymorphism and evolution of microRNA genes in Arabidopsis halleri.

Flavia Pavan, Chloé Beaumont, Sophie Gallina, Clément Mazoyer, Mathieu Genete, Eléonore Durand, Firas Louis, Vincent Castric, Sylvain Legrand
Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France

Abstract

MicroRNAs (miRNAs) are a class of small non-coding RNAs that play important regulatory roles in plant and animal genomes. They are produced from a primary transcript forming a short foldback structure from which a mature miRNA is processed and negatively regulates a series of mRNA targets. Some miRNA genes are conserved over long evolutionary time scales, while others have apparently emerged more recently. To compare the functional constraint over the evolution of these two classes of miRNA genes, we performed a polymorphism analysis of 72 Arabidopsis halleri individuals from across the species range. Our resequencing strategy was based on targeted sequence capture of 185 miRNA genes and their predicted target sites. We identified the homologous miRNAs in 88 plant species and determined the time since they emerged. We observed that the mature miRNA sequence presented two fold lower nucleotide diversity than the other parts of the hairpin (stem, terminal loop). The miRNA-targeted region on the CDS was also less polymorphic than the neighboring regions, with a mean reduction of about 20%. This suggests that evolution of the mature miRNA sequence and its complementarity target sequence are constrained by purifying selection. In addition, the evolutionarily conserved miRNA hairpins and their mature miRNA sequences presented less nucleotide diversity than the more recently emerged ones, suggesting that younger miRNA genes generally evolve under weaker selective constraint than older ones. These findings contribute to a broader understanding of how new regulatory elements integrate into regulatory networks over the course of evolution.

Room 2-01

Functional investigation of evolutionarily young transcription start sites in human genome

Liang Wu
Sun Yat-sen University, Guangzhou city, Guangdong province, China

Abstract

The transcription start sites (TSSs) play important roles in regulation of gene expression. Previous studies revealed high turnover rate of TSSs in primate genomes and identified many evolutionarily young TSSs in human genome. However, the functions of these evolutionarily young TSSs in human genome remain poorly understood. To investigate functions of young TSSs, we identified previously reported regulatory variants (e.g. regulatory QTLs, allele-specific transcription factor binding, chromatin accessibility etc.) that are located in or around young TSSs. Next, we studied potential functional effects of these variants on young TSSs by integrating multiple sources of data. We identified seven regulatory variants associated with functions of seven young TSSs with high confidence. Among these TSSs, one from gene *CCDC122* is of particular interest, because it is reported to be involved in leprosy in genome-wide association studies. We performed further analysis to reveal putative roles of this TSS in development of leprosy in human. We also compared the frequencies of regulatory variants associated with young TSSs in different populations and discussed the implications. In sum, our integrative analysis with multi-omics data provides novel insights into functions of evolutionarily young TSSs.

Room 2-02

A high-quality bonobo genome refines the analysis of hominid evolution

Yafei Mao
University of Washington, Seattle, Washington, USA

Abstract

The divergence of chimpanzee and bonobo provides one of the few examples of recent hominid speciation. Here we describe a fully annotated, high-quality bonobo genome assembly, which was constructed without guidance from reference genomes by applying a multiplatform genomics approach. We generate a bonobo genome assembly in which more than 98% of genes are completely annotated and 99% of the gaps are closed, including the resolution of about half of the segmental duplications and almost all of the full-length mobile elements. We compare the bonobo genome to other great apes and identify more than 5,569 fixed structural variants that specifically distinguish the bonobo and chimpanzee lineages. We focus on genes that have been lost, changed in structure or expanded in the last few million years of bonobo evolution. We produce a high-resolution map of incomplete lineage sorting and estimate that around 5.1% of the human genome is genetically closer to chimpanzee or bonobo and that more than 36.5% of the genome shows incomplete lineage sorting if we consider a deeper phylogeny including gorilla and orangutan. We also show that 26% of the segments of incomplete lineage sorting between human and chimpanzee or human and bonobo are non-randomly distributed and that genes within these clustered segments show significant excess of amino acid replacement compared to the rest of the genome.

Room 2-03

Diverse molecular mechanisms contribute to differential expression of human duplicated genes

Colin Shew, Paulina Carmona-Mora, Daniela C. Soto, Mira Mastoras, Elizabeth Roberts, Joseph Rosas, Dhriti Jagannathan, Gulhan Kaya, Henriette O'Geen, Megan Y. Dennis
University of California, Davis, Davis, CA, USA

Abstract

Emerging evidence links genes within human-specific segmental duplications (HSDs) to traits and diseases unique to our species. Despite being nearly identical by sequence (>98.5%), paralogous HSD genes are differentially expressed across human cell and tissue types, though the underlying mechanisms have not been examined. We first considered the evolutionary trajectories of HSD gene expression by comparing cross-tissue mRNA levels of 75 HSD genes from 30 families between humans and chimpanzees. These data indicated expression patterns consistent with relaxed selection on derived paralogs, while ancestral paralogs exhibited greater expression conservation with chimpanzee orthologs. Exceptions to this pattern suggested certain derived paralogs may retain or supplant ancestral functions. Within humans, we also quantified HSD gene expression in a population-scale panel of lymphoblastoid cell lines (LCLs, N=462). We found no relationship between paralogous expression divergence and sequence divergence or copy number variation, and a limited influence of post-transcriptional regulation. We next identified novel candidate *cis*-regulatory elements (cCREs) in duplicated regions through reanalysis of publicly available chromatin immunoprecipitation sequencing (ChIP-seq) data from the LCL GM12878, as well as large-insert ChIP-seq libraries targeting active chromatin features (H3K27ac, H3K4me3, H3K4me1, RNA PolII). Finally, we used luciferase reporter assays to compare HSD cCRE activity in LCLs and HeLa cells and found some duplicated cCREs (2/3 promoters and 5/7 distal elements) drove differential reporter activity, suggesting they may contribute to divergent *cis*-regulation of paralogous genes. This work provides evidence that *cis*-regulatory divergence contributes to novel expression patterns of very young (<6 million years) gene duplicates in humans.

Room 2-04

Genetic diversity in chimpanzee transcriptomics data does not represent wild populations

Navya Shukla[1,2], Bobbie Shaban[2], Irene Gallego Romero[1,2]

[1]School of BioSciences, University of Melbourne, Melbourne, Victoria, Australia. [2]Melbourne Integrative Genomics, University of Melbourne, Melbourne, Victoria, Australia

Abstract

Comparative transcriptomic studies between humans and non-human primates are used to characterize differences in gene expression that could underlie observed phenotypic differences. Chimpanzees, as one of our closest extant relatives, have been particularly useful in these studies. Although critically endangered, 4 distinct subspecies of chimpanzees—Western (Pan troglodytes verus), Central, (Pan troglodytes troglodytes) Eastern (Pan troglodytes schweinfurthii) and Nigeria-Cameroon (Pan troglodytes ellioti)—are recognised, each of them containing substantial genetic diversity. Ideally, transcriptomic studies would sample as much of this genetic diversity as possible, but the majority of tissue samples in existing RNA-seq datasets are derived from existing cell lines or collected post mortem from captive research animals. Additionally, samples are often inconsistently labelled across studies, making it difficult to ascertain the true number of chimpanzees from which transcriptomic data is available.

To address these questions we genotyped 468 RNA-seq samples available in the public NCBI Sequence Read Archive (SRA) and evaluated genetic diversity across them, using a reference dataset of wild-born animals. 452 samples had significant levels of Western ancestry; 276 were exclusively Western. At the individual level, we used Identity-by-State clustering and available metadata information to identify 135 unique individuals in our dataset. We then assessed relatedness between these individuals, and inferred 24 parent/offspring, 5 sibling, and 108 second-degree pairs, some previously unknown, in 9045 total sample pairs. Altogether, our results show that current transcriptomic data of chimpanzees does not represent extant genetic diversity in chimpanzees, and provide important context to current comparative transcriptomics research.

Room 2-05

Drivers of copy number variation in adaptive immune receptor V gene families of the great apes

Hao H Yiu, Philip LF Johnson
University of Maryland, College Park, MD, USA

Abstract

The vertebrate adaptive immune system uses a repertoire of immune receptors generated by somatic recombination of genes to recognize diverse antigens. These genes evolved through duplication, deletion, and diversification, resulting in gene families showing copy number and allelic variation which enhances the diversity of the immune repertoire. The variable (V) genes of the immunoglobulin (IGHV) and T cell receptor beta (TRBV) loci contribute the bulk of antigen specificity to their respective receptors. High polymorphism of many immune genes may have been driven by pathogen-mediated selection, however, the germline evolution of the IGHV and TRBV loci remain understudied in part due to difficulties in distinguishing between orthologs and paralogs. Here we investigate the evolutionary history of these gene families throughout the great ape phylogeny using a combination of data from human immune-focused sequencing projects and whole genome shotgun sequencing from the Great Apes Genome Project. We assess both within- and between- species diversity and estimate rates of V gene family expansion. Considering the extensive within-genome diversity, we observe surprisingly modest differences between different populations of the same species and between the great ape species. While most V gene families in the great apes expanded or contracted without signatures of selection, we find limited evidence of selection driving a Nigerian Chimpanzee-specific expansion of several IGHV and TRBV gene families. Together, these results suggest that pathogen selection pressure has played only a limited role in the on-going evolution of germline V gene loci, despite their key role in vertebrate adaptive immunity.

Room 2-06

Reconciling methods for detecting adaptation from inter- and intra-specific data

Thibault Latrille, Nicolas Lartillot
LBBE, Lyon, France

Abstract

Adaptation in protein-coding sequences can be detected using multiple sequence alignments across species (inter-specific data). This method uses phylogenetic codon models, classically formulated in terms of the ratio of synonymous and non-synonymous substitution rates. However, because of the background of purifying selection, these models are potentially limited in their sensitivity. Recent developments have led to more sophisticated mutation-selection codon models aiming at making a more detailed quantitative assessment of the interplay between mutation, purifying and positive selection, leading to potentially more powerful and more quantitative detection of adaptation. However, these alternative codon models have not yet been assessed more extensively on empirical data. In this study, we conducted a large-scale analysis on placental mammals protein-coding sequences, and assessed the performance of mutation-selection codon models to detect proteins and sites under adaptation. Mutation-selection models detected up to 4% of proteins to be under ongoing adaptation. Proteins under adaptation were strongly enriched in ontology terms related to immune processes. Finally, adaptation in protein-coding sequences can also be assessed by combining divergence and polymorphism (intra-specific data) in so-called McDonal & Kreitman tests. Taking advantage of this independent approach and leveraging intra-specific polymorphism from Chlorocebus Sabaeus (Green monkey), we integrated inter- and intra-specific data across the entire exome, and showed that proteins and sites detected to be under adaptation at the phylogenetic scale are also under adaptation at the population-genetic scale. Altogether, our exome-wide analysis shows that phylogenetic mutation-selection codon models and population-genetics test of adaptation (McDonal & Kreitman test) can be reconciled and are congruent.

Room 2-07

Allelic origin of human L and M opsin genes predating catarrhine-platyrrhine split

Yuka Matsushita[1], Naoko Takezaki[2], Amanda D Melin[3], Shoji Kawamura[1]

[1]Dept. Integrated Biosci., Grad. Sch. Frontier Sci., Univ. Tokyo, Kashiwa, Chiba, Japan. [2]Faculty of Med., Kagawa Univ., Kita-gun, Kagawa, Japan. [3]Dept. Anthropol. Archaeol. and Dept. Med. Genet., Univ. Calgary, Calgary, Alberta, Canada

Abstract

Routine trichromatic color vision is a characteristic feature of catarrhine primates including humans. This is enabled by L and M opsin genes arrayed on the X chromosome. On the other hand, most of platyrrhines and some species of lemuriforms have polymorphic color vision achieved by spectrally distinct L/M opsin alleles of a single locus gene. However, it remains elusive whether the L and M opsin gene loci in catarrhines and the L/M opsin alleles of platyrrhines (and lemuriforms) originated independently or not. The L and M opsin genes of catarrhines are susceptible to gene conversion and homogenization between them, which obscures their evolutionary history. To obtain entire coding sequences of the L/M opsin genes from catarrhines, platyrrhines, tarsiers and strepsirrhines, we applied the target capture and massive parallel sequencing methods to their genomic DNA samples. To lessen the homogenizing effect on reconstruction of the phylogenetic relationships of L and M opsins, we used only gene regions (282 bp in total) which are highly divergent between the L and M opsin genes in catarrhines. Using the maximum likelihood method, we showed that the L and M sequences of catarrhines split before catarrhines and platyrrhines split. Hence, the L and M opsin genes of catarrhines originated from juxtaposition of already differentiated L and M alleles and not from duplication of identical sequences and subsequent divergence.

Room 2-08

Rapid Evolution of Bacterial Evasion Through Domain Shuffling of a Primate Housekeeping Protein.

EmilyClare Baker, Ryan Sayegh, Kristin Kohler, Wyatt Borman, Matthew Barber
University of Oregon, Eugene, Oregon, USA

Abstract

Vertebrate epithelial surfaces are often the site of initial contact between hosts and pathogens. As such, interactions at this interface play an outsized role in determining successful colonization or resistance. Proteins dedicated to host defense may have significant scope in how they may evolve at pathogen binding surfaces. For proteins involved in broader host functions it is unclear if they have the same evolutionary flexibility when pathogens target functional surfaces. The vertebrate CEACAM family of proteins are widely expressed across epithelial surfaces and function in a variety of cellular processes including cellular adhesion, cell-cycle regulation, and immune signaling. CEACAMs are a frequent target of bacterial adhesin proteins, facilitating surface attachment and pathogenesis by bacterial cells. We have found that the multifunctional protein binding domain of a subset of CEACAM proteins show evidence of rapid evolution in primates. Binding experiments demonstrate that divergence between primate species in this domain determines binding with a genetically and structurally diverse set of adhesins from pathogenic bacteria. Additional analyses found that the evolutionary history of bacterially antagonized CEACAMs has been characterized by extensive concerted evolution of the protein binding domain facilitated by repeated episodes of gene conversion. Examination of human population data found evidence that gene conversion continues to shape CEACAM variation in humans. We further show that this variation can influence binding to bacterial adhesins. In all, this work demonstrates how gene conversion among paralogs can drive rapid host protein evolution and alter binding to pathogenic bacterial adhesins.

Room 2-09

Ape-specific ATF4 retrocopies may act to regulate parental ATF4 activity and integrated stress response outcomes

Hans M Dalton, Katie G Owings, Nels C Elde, Clement Y Chow
University of Utah, Salt Lake City, UT, USA

Abstract

Multiple stress signals trigger the integrated stress response (ISR), including endoplasmic reticulum stress and viral infection, causing strong inhibition of protein synthesis machinery while repair occurs. Certain genes contain upstream Open Reading Frame (uORF) regulatory sequences in their 5'UTR region that allow, or even upregulate, translation during the ISR. One of the most upregulated uORF-containing transcripts is *ATF4* - a transcription factor that activates pro-survival or apoptotic genes.

We have found conservation of multiple *ATF4* retrocopies - labeled in humans as pseudogenes *ATF4P1-P4*. *hATF4P1* and *P2* are highly truncated retrocopies that date to the common ancestor between apes and old world monkeys, while *hATF4P3* and *P4* are ape-specific. *hATF4P3* is a full-length copy that maintains all domains and uORF regulatory regions with 95% amino acid (AA) identity with the parent *ATF4* gene; of the 5% AA that are diverged, none are in the DNA binding domain. *hATF4P4* has a truncation event removing its DNA binding domain yet maintains 90% AA identity with the parent gene in the truncated half. *hATF4P4* is well conserved in humans, chimps, bonobos, and gorillas (Homininae). Strikingly, independent retrocopies with similar truncation events have occurred in orangutan, gibbon, and gorilla genomes.

We have cloned the full-length (*P3*) and truncated (*P4*) *ATF4* retrogenes into human cell lines and tested for viability under ISR activation. Intriguingly, overexpression of *hATF4P4* under ISR activation causes loss of cell viability. We hypothesize that *hATF4P4*-like retrocopies may regulate the parent *ATF4* gene by interaction with its regulatory proteins.

Room 2-10

Evolutionary diversity of bitter taste receptor gene repertoire compared among cercopithecid and platyrrhine monkeys

Min Hou[1], Masahiro Hayashi[1], Ryuichi Ashino[1], Amanda D Melin[2], Shoji Kawamura[1]
[1]University of Tokyo, Kashiwa, Chiba, Japan. [2]University of Calgary, Calgary, Alberta, Canada

Abstract

The bitter taste receptor gene (*TAS2R*) family is comprised of roughly 20~30 intact genes in mammals. However, it remains elusive how its diversity is shaped during evolution. It is of interest whether the diversity is influenced by other senses. Primates consist of taxonomic groups which are diverse in many aspects including color vision. In this study we focus on cercopithecid and platyrrhine monkeys of which only a few species with the whole-genome sequence (WGS) data publicly available have been studied for *TAS2R* gene repertoire. The color vision of cercopithecids is characterized as the routine trichromacy, while that of platyrrhines is diverse ranging from routine trichromacy in howler monkeys, trichromacy/dichromacy polymorphism in many genera, and to monochromacy in nocturnal owl monkeys. Both groups are also diverse in their diets. We employed the target capture (TC) method specifically probing *TAS2R*s followed by massive-parallel sequencing to improve retrieval and identification of gene sequences which are often problematic in WGS for multigene families. We will discuss whether TC is more effective than WGS and how TC can be improved in retrieving gene sequence and distinguishing intact and disrupted genes. We will also discuss whether color vision has any impact on *TAS2R* repertoire in these monkeys.

Room 2-11

Population genomics of the pseudoautosomal region in hominids

Juraj Bergman, Mikkel H Schierup
Aarhus University, Aarhus, Denmark

Abstract

The pseudoautosomal region (PAR) is a 2.7 Mb region in the telomeres of sex chromosomes with a crucial role in pairing and proper inheritance of X and Y. Paradoxically, it is also a dynamic region, prone to fast nucleotide evolution and high recombination rates. We will present the evolutionary dynamics of the PAR across great apes, as well as modern and extinct Homo species. We estimate PAR divergence rates to be $1.14\text{-}1.21\times10^{-9}$ substitutions per site and year, approximately 40-65% larger than autosomal rates. Compared to the ancestral sequence, the chimpanzee PAR is most highly diverged (3.2%), while the evolution of the orangutan PAR is slowest (2.8% diverged) and AT-biased. Approximately 50% of PAR mutations are due to non-replication origins, and likely attributable to recombination-associated mutation. The strength of GC-biased gene conversion is strongest for CpG transitions and generally twice as strong in the PAR compared to the autosomes. We show that the clock-like behaviour of CpG mutations in the PAR is highly disrupted due to their elevated turnover rate (~20% of ancestral CpG sites have diverged along the hominid phylogeny). We show that the PAR has a unique repeat structure comprising the largest proportion of its length (~25%) compared to other telomeric regions, which may play a role in shaping its unique recombination landscape. We also characterise the potential impact of PAR recombination on PRDM9 motif evolution in Homo species. Lastly, we compare female and male diversity patterns in the PAR, but find no evidence for sexual antagonism.

Room 2-12

Cell-type and cytosine context-specific evolution of DNA methylation in the human brain

Hyeonsoo Jeong[1], Isabel Mendizabal[1], Stefano Berto[2], Paramita Chatterjee[1], Thomas Layman[1], Noriyoshi Usui[2], Kazuya Toriumi[2], Connor Douglas[2], Devika Singh[1], Iksoo Huh[1], Todd M. Preuss[3], Genevieve Konopka[2], Soojin V. Yi[1]

[1]Georgia Institute of Technology, Atlanta, GA, USA. [2]UT Southwestern Medical Center, Dallas, TX, USA. [3]Emory University, Atlanta, GA, USA

Abstract

Recent studies of human brain evolution have largely focused on gene and protein expression differences comparing humans to other non-human primates. However, it is increasingly appreciated that epigenetic modifications including regulatory changes play an important role in human evolution. Despite the fundamental importance of epigenetics in regulatory and developmental processes, how human epigenetic patterns differ from those of other species remains poorly understood. Also, due to the heterogeneity of bulk tissue, reliable identification of human-specific epigenetic modifications remains a major hurdle. To address this challenge, we generated whole-genome methylomes from neurons and oligodendrocytes from the frontal cortex of humans, chimpanzees, and rhesus macaques to demonstrate dynamic evolutionary trajectories of DNA methylation in cell-type and cytosine-context specific manner. We show that the two brain cell types have distinct DNA methylation patterns, yet the majority of the differential methylation between cell types is highly conserved across the three species. Remarkably, DNA methylation in non-CG context has increased (hypermethylation) in neuronal gene bodies during human brain evolution, contributing to human-specific down-regulation of gene expression particularly in early development and neuronal subtypes. In contrast, DNA methylation in CG context shows pronounced reduction (hypomethylation) during evolution in the human brain, which significantly contributed to cell-type-specific active regulatory regions. Also, the human-specific neuron-hypomethylation carries a significant genetic risk for schizophrenia. These findings indicate substantial reprogramming of epigenomic landscapes during human brain evolution that has shaped the cellular regulatory landscape and contributed to the increased vulnerability to neuropsychiatric diseases.

Room 2-13

Non-invasive sampling to investigate local genetic adaptation in chimpanzees

Harrison J Ostridge[1,2], The Pan African Programme: The Cultured Chimpanzee Consortium[3]
[1]UCL, London, London, United Kingdom. [2]ZSL, London, London, United Kingdom. [3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Saxony, Germany

Abstract

Chimpanzees are our closest living relatives and are endangered with numbers in continuous decline. They inhabit a diversity of habitats in sub-Saharan Africa, from deep forest to woodland savannah. These environmental differences provide the opportunity for genetic adaptations that affect both the genome and phenome. Because environmental differences do not follow subspecies boundaries, studying adaptation to the environment requires an analysis that is fine-scale at the geographic level. Generating this type of dataset is extremely difficult for a species that is endangered, protected and for which only non-invasive sampling is possible. Here we overcome this challenge by using faecal samples from wild individuals across the range of the species. As part of the Pan African Programme, we captured and sequenced 800 exomes from non-invasive samples of wild chimpanzees across all four subspecies. We show that this process produces a valuable genetic dataset that allows us to identify the characteristic signatures of local adaptation. Further, for each sampled population a wealth of information on habitat, climate and behaviour was also collected. As a second step in our analysis, we will combine these unprecedented genetic and environmental datasets to jointly identify fine-scale local adaption together with the likely associated selective factors. This will provide the first map of local adaptation in chimpanzees which, besides having obvious interest to understand the species, can inform conservation efforts by identifying populations that exhibit adaptive genetic diversity and key environmental pressures that influence chimpanzee survival and reproduction.

Room 2-14

Selection in the Great Ape Face: A Balanced View.

Alfie Gleeson, Aida ANDRÉS, Aida Gómez-Robles
University College London, London, United Kingdom

Abstract

Facial morphology is a diverse phenotype in humans and other great apes. Although levels of cranial diversity largely mirror neutral genomic diversity, the extent to which selection may have shaped craniofacial morphology is not understood. Some studies suggest that balancing selection maintains facial diversity in humans, and potentially other great apes. In the past decade GWAS have identified SNPs associated with variations in facial morphology in humans. We used these SNPs to identify candidate “face morphology genes”, allowing us to investigate signatures of balancing selection associated with face morphology across the great apes. We built a set of 159 candidate genes and investigated their signatures of balancing selection using genome-wide values of the NCD statistic and empirical p values. Fifty face morphology genes show signatures of balancing selection in at least one of the great apes. *Pan troglodytes schweinfurthii* showed the most (n=13, 8.2% of total candidates), much higher than human populations (7 populations, 2-5 genes each, mean= 4, ~2.5% of total candidates). After accounting for ascertainment bias, we observe no enrichment in targets of balancing selection among candidate genes. Interestingly, when categorising the candidate genes into their original associated phenotype, we find a moderate enrichment for the lips and eyes, and a deficit in the general face phenotype. Our results suggest that although selection may maintain high diversity in a small number of genes associated with facial morphology in great apes, there is insufficient evidence to suggest that balancing selection is a strong force in generating facial diversity.

Room 2-15

De novo emergence of functional microproteins in the human lineage

Nikolaos Vakirlis[1], Kate Marie Duggan[2], Aoife McLysaght[2]
[1]B.S.R.C. "Alexander FLeming", Vari, Greece. [2]Trinity College Dublin, Dublin, Ireland

Abstract

Recent work has identified hundreds of short, noncanonical Open Reading Frames (ORFs) in the human genome that are translated and are essential for cellular growth. Most of these functional coding sequences had so far been misannotated because of their short length and lack of conservation. We investigated the evolutionary origin of hundreds of these ORFs in unprecedented detail, by reconstructing their ancestral genomic sequences and their ancestral transcriptional status across the vertebrate phylogeny. Many originated at the root of the vertebrate tree and have conserved their short length ever since. Crucially however we were also able to find clear cases of human-specific, de novo emerged ORFs that formed on both ancient and more recent noncoding transcripts. Thus, our findings demonstrate the recent de novo birth of microproteins with essential cellular functions, unique to the human genome.

Room 2-16

Sex-specific population structure in baboons revealed by hundreds of complete X-chromosomes

Erik F Sørensen[1], Jeffrey Rogers[2], Christian Roos[3], Dietmar Zinner[3], Sascha Knauf[3], Clifford J Jolly[4], Jane Phillips-Conroy[5], Julia Fischer[3], Idrissa Chuma[6], Julius Keyyu[7], Kasper M Terkelsen[1], Primate Sequencing and Conservation Initiative[3,8]

[1]Aarhus University, Aarhus, Denmark. [2]Baylor College of Medicine, Houston, Texas, USA. [3]German Primate Center, Göttingen, Germany. [4]New York University, New York, New York, USA. [5]Washington University, St. Louis, Missouri, USA. [6]Tanzania National Parks, Arusha, Tanzania, United Republic of. [7]Tanzania Wildlife Research Institute, Arusha, Tanzania, United Republic of. [8]Baylor College of Medicine, Houston, USA

Abstract

Baboons (Papio) contain multiple hybridizing species and their divergence is similar in timeframe to the divergences among Homo species. The system thus presents an opportunity to study species dynamics similar to those acting among archaic and modern humans in the past.

Baboons have a large variation in male reproductive success and highly varied social structure, with four species having male-biased dispersal and two having female-biased dispersal. Both of these circumstances lead to strong signals on the X chromosome, which experience vastly different dynamics than the autosomes both concerning the histories of population size and admixture events. These factors allow us to investigate an interesting system of the dynamics present when genes flow across species boundaries.

We sequenced 223 wild baboons at 30X depth, primarily sampled from locations in East Africa. The number of samples together with the breadth of sampling locations allows for investigation of the degree of variation present in each species, as well as signals of population structure and admixture.

We use PCA to visualize the clusters present on autosomes and the X chromosome, summary statistics based on the degree of heterozygosity to ascertain diversity, as well as statistics based on private and shared variants in the species to ascertain possible admixture.

We find that the heterozygosity on the X chromosome is depressed compared to that on autosomes when allowing for the lower population size and mutation rate on X. We further find that admixture signals are weaker on the X, suggesting some amount of reproductive isolation.

Room 2-17

How sequence context leads to variable rates of mutation across the genome

Madeleine Oman, Aqsa Alam, Rob Ness
University of Toronto, Toronto, Canada

Abstract

Genetic mutations introduced during DNA repair and replication contribute all the raw genetic variation for evolution and also underlie a range of genetic diseases. The rate at which mutations occur can vary from site to site across the genome by up to 100-fold. The drivers of this variation are a major area of research for evolutionary, medical, and applied genomics researchers. Numerous mechanisms that may increase mutability, however the strongest predictor of mutation rate is the sequence itself. Studies examining variation in the mutation rate among the 64 trinucleotides have shown that across life the most mutable trinucleotides mutate at rates nearly 30-fold higher than the least mutable. The idea that DNA sequence drives its own mutability raises a simple prediction; highly mutable sequences will mutate more frequently and eliminate themselves in favor of sequences with lower mutability, ultimately leading to a lower equilibrium state. However, while this might be true in neutral sequence, purifying selection may impede progress toward lower mutation states, causing higher rates of mutation in the genome's most important regions. Here we conduct an evolutionary simulation using human mutation data to test the predictions that (1) DNA will evolve to relatively low equilibrium mutation rate and (2) Purifying selection will constrain the change of sequence resulting in higher equilibrium mutation rates in coding sequence. In support of our predictions, we show that sequence may evolve in simple processes in the genome, which may have important implications for how we model evolution and susceptibility to disease.

Room 2-18

Proteome-wide scan reveals targets of sexual selection in primates

Brianna L Ports, Michael Jensen-Semaan
Duquesne University, Pittsburgh, PA, USA

Abstract

Polygynandry is a mating system where multiple males mate with an individual female during a single estrus cycle. In this system, competition for mates often occurs after ejaculation where males with characteristics increasing reproductive success will fertilize more eggs and pass those characteristics onto their offspring. This varies distinctly from polygynous and monogamous species where females mate with only one male and sperm competition is diminished. The wide range of mating types that are present between closely related primate species serves as an excellent model to examine the molecular mechanisms that drive these complex mating systems. Here we focus on detecting genes that have evolved in response to similar selective pressures induced by differences in mating systems using testes-to-body weight ratios as a proxy. To investigate these differences, we identified residual testes sizes for 22 primate species and collected 19,975 aligned protein sequences. Using maximum likelihood, we identified gene trees for each individual gene with branch lengths representing the rate of nonsynonymous substitutions. A linear regression model was then used to compute the relative evolutionary rate of each gene with corrections for average branch lengths. This methodology allowed us to detect genes that are evolving rapidly in polygynandrous species from distinct lineages, but not in monogamous or polygynous species. Further downstream likelihood analyses in PAML helped to identify the strength and direction of selection on each rapidly evolving gene as well as individual sites being selected upon.

Room 2-19

The role of p63 gene in primate including human tooth development

Adam M Wandzura[1], Sara V Good[2], Julia C Boughner[1]

[1]University of Saskatchewan, Department of Anatomy, Physiology & Pharmacology, Saskatoon, SK, Canada. [2]University of Winnipeg, Department of Biology, Winnipeg, MB, Canada

Abstract

Funding: NSERC Discovery Grants to SVG and JCB, USRA to AW

The evolutionarily ancient transformation-related protein 63 (p63) gene is widely expressed in vertebrates including primates. This p63 transcription factor is integral to the development of many body parts including teeth. Our team recently has shown that p63 regulates a network of genes vital to tooth formation in mouse and fish. We now seek to probe for downstream members of this network in primates, but ethical factors make samples of developing primate teeth scarce. As an alternate first step, we conducted a comparative evolutionary analysis to test the relative degree of evolutionary conservation of p63 domains across extent primate species encompassing humans, monkeys and apes. The four domains [DNA Binding, Oligomerization, Transactivation, SAM] of p63 were analyzed, and molecular phylogenies calibrated to model evolutionary conservation. Results show that the p63 oligomerization domain (involved in transcription factor activation) was least conserved across primates relative to the gene's other three domains. Our novel findings suggest that evolutionary constraint on p63's oligomerization domain was conducive to the selection of novel downstream targets and formation of a gene regulatory network.

Room 2-20

Experimental and Bioinformatic Analyses of Coevolution of Primate Seminal Proteins and HIV/SIV

Emine F Kahveci[1], Feng Hsiao[2], Thomas Mack[3], Ludger Ständker[3], Jan Münch[3], Nadia R Roan[2], Michael I Jensen-Seaman[1]

[1]Duquesne University, Pittsburgh, PA, USA. [2]University of California, San Francisco, CA, USA. [3]Ulm University, Ulm, Germany

Abstract

Male reproductive proteins are among the most rapidly evolving proteins in mammals, commonly attributed to sexual selection in the form of sperm competition. However, defense against sexually transmitted pathogens such as HIV/SIV might also be a contributing factor. Previous studies have demonstrated that amyloid fibrils formed from peptides derived from the abundant human seminal proteins semenogelin 1, semenogelin 2, and prostatic acid phosphatase are used by HIV to dramatically increase the infectivity rate. Considering the much longer history of SIV prevalence in nonhuman primates compared to humans and the large documented fitness costs associated with SIV infection in chimpanzees, we aim to test whether these proteins in primates have been evolving in response to HIV/SIV. We used maximum likelihood-based sequence analysis to identify codons evolving under positive and purifying selection in primates grouped by their presence or absence of SIV. The regions of these proteins that form fibrils do not appear to be evolving significantly differently than the non-fibril forming regions. Empirically, we tested synthetic peptides corresponding to the fibril-forming regions of nine primate species for their ability to form fibrils *in vitro* and to enhance HIV infectivity using the TZM-bl reporter cell line. Human homologs consistently showed a greater ability to form fibrils and enhance infection, most notably compared to those from chimpanzee and gorilla. The inclusion of hypothetical ancestral peptides allowed us to infer the direction of change. Furthermore, using phylogenetic approaches to map character states we extended our prediction of ancestral phenotypes more broadly throughout primates.

Room 3-01

Are transposons responsible for the streamlined genomes of prokaryotes?

Bram van Dijk[1], Frederic Bertels[1], Paul Rainey[1], Nobuto Takeuchi[2]
[1]Max Planck Institute for Evolutionary Biology, Plön, Germany. [2]University of Auckland, Auckland, New Zealand

Abstract

Eukaryotes and prokaryotes have very distinct genome architectures, with astonishing differences in the ratio of coding/non-coding DNA, the abundance of transposable elements (TEs), the number of paralogous genes, and in complexity in general. Multiple hypotheses have been proposed to explain why prokaryotes have small and streamlined genomes, that is to say, consist mostly of essential DNA. In this modelling study, we present a novel hypothesis based on the co-evolution of TEs and their host genomes. We show that TEs, DNA elements that can self-amplify within genomes, can actually promote the reduction of genome size. The mechanism behind this emergent form of genome streamlining is analogous to abortive infection as observed for bacteriophages, where a lineage hampers the spread of selfish genetic elements by means of early death. Although this suicidal defence strategy is not beneficial to the individual, it confers protection to the lineage by preventing the accumulation of TEs in the environment. Interestingly, this streamlining only emerges when TEs spread through horizontal gene transfer (HGT), and does not happen in a sexual population. In short, the co-evolution of TEs and their host can elegantly explain why prokaryotes have streamlined genomes, and why eukaryotes do not.

Room 3-02

Purifying selection in the long-term evolution experiment with Escherichia coli

Rohan Maddamsetti[1], Nkrumah Grant[2]
[1]Duke University, Durham, USA. [2]University of Idaho, Moscow, USA

Abstract

Purifying selection maintains organismal structure and function over evolutionary time. Despite its importance for understanding the evolutionary dynamics of natural microbial communities, purifying selection has been little studied in microbial evolution experiments. In a series of recent papers, we report that simple methods, applied to genomic and metagenomic time series of Lenski's long-term evolution experiment with *Escherichia coli* (LTEE), are able to discover several interesting patterns of purifying selection, which in some cases, are universal across the tree of life. First, abundant proteins evolve slowly in mutator populations of the LTEE. Specifically, the density of all observed mutations per gene, significantly anti-correlates with mRNA abundance, protein abundance, and degree of protein-protein interaction. The same pattern holds for nonsynonymous mutation density. However, synonymous mutation density, measured across the LTEE mutator populations, positively correlates with protein abundance. Second, we find evidence of purifying selection on the *E. coli* protein-protein interaction (PPI) network, such that evolved PPI networks are more resilient than expected by chance. Third, we developed a simple test to infer mode of selection (STIMS) which is able to recover a signal of purifying selection on essential genes. Many of these essential genes show evidence of both positive and purifying selection under LTEE conditions. Finally, we used STIMS to test for purifying selection on central and superessential metabolic enzymes, and find that patterns of purifying selection on metabolic enzymes in the LTEE are largely idiosyncratic and population-specific.

Room 3-03

Bacteria Don't Remain Neutral on the Matter: Differential Contribution of Adaptive Evolution Across Bacterial Species

Ellis L Torrance, Awa Diop, Louis-Marie Bobay
University of North Carolina at Greensboro, Greensboro, NC, USA

Abstract

The pervasiveness of neutral versus adaptive evolution (i.e. positive selection) is a cornerstone debate in the study of evolutionary and population genetics. However, the impact of adaptive relative to neutral evolution on shaping bacterial genomic and evolution dynamics remains largely unknown. As of yet, the impact of positive selection (α) on a gene-by-gene basis has only been assessed in few bacterial species using a small number of related genomes. In this analysis, we quantified the contribution of adaptive evolution to bacterial genome dynamics by comparing hundreds of genomes across multiple groups of closely related bacterial species. Results revealed a significant contribution of adaptive evolution to bacterial evolution; however, the impact of adaptive evolution relative to neutral evolution was found to vary substantially across species. The estimated proportion of adaptive selection for each gene also demonstrated variations along chromosome location. Overall, this study provides a first estimate of the contribution of adaptive evolution—and its variation—across many bacterial species under a unified framework.

Room 3-04

Gene Flow and Introgression are Pervasive Forces Shaping Prokaryote Evolution.

Awa Diop, Ellis L Torrance, Caroline M Stott, Louis-Marie Bobay
Microbial Genomics and Evolution Lab, Department of Biology, University of North Carolina at Greensboro, Greensboro, North Carolina, USA

Abstract

Bacteria and archaea are asexual organisms, and their evolution is thought to be primarily dominated by clonal evolution. In this study, we assessed species boundaries by comparing traditional taxonomic classification schemes to the patterns of gene flow in the genomic backbone of bacteria and archaea, i.e. the core genome. We analyzed over 2,500 species and >100,000 genomes and delimited biological species in Bacteria and Archaea based on their patterns of gene flow within and between species. Our results show that very few populations appear to be truly clonal, where only 2.2 to 17.4% of prokaryotic species do not present signs of gene flow, indicating that truly asexual organisms are very rare in nature. We also found that sequence thresholds, which are routinely used to define species borders in microbes, are poor predictors of species borders when defined with gene flow. Finally, we found that many species engage in various levels of introgression, although this does not preclude their classification into distinct biological species. Overall, our findings indicate that prokaryotic evolution is shaped by similar forces driving the evolution of sexual organisms and gene flow appears to play a predominant role across the Tree of Life.

Room 3-05

The effect of antibiotic resistance plasmids on biofilm formation in clinical enterobacterial isolates

Laura Brülisauer, Ricardo Leòn-Sampedro, Fabienne Benz, Alex Hall
ETH, Zürich, Switzerland

Abstract

Plasmids contribute crucially to the spread of antibiotic resistance (AR) via horizontal gene transfer among bacterial pathogens. Besides conferring resistance, AR plasmids can provide various other beneficial traits to their hosts, enhancing their success colonizing the human gut. In clinical settings, biofilms are one of the main concerns due to their potential to further increase resistance against antimicrobial substances by physical and chemical protection. Some evidence shows that this lifestyle can promote plasmid endurance. However, there is little consensus about how AR plasmids specifically affect biofilm formation and the success of clinically relevant bacteria. Here, we study the influence of AR plasmids on biofilm formation in clinical enterobacteria isolated from hospitalized patients. We use four *Escherichia coli* isolates belonging to the four most relevant phylogroups, each harbouring a different AR plasmid encoding extended-spectrum β-lactamase or carbapenemase genes.

First, we tested the effect of AR plasmids in two *E. coli* recipients under different growth conditions and found plasmid-mediated changes in biofilm formation. In ongoing work, using a CRISPR/Cas9 based system we are curing the plasmids of the clinical isolates for comparison of the plasmid-mediated effects on biofilm formation. To compare these effects in the different clinical isolates, we introduce each plasmid into each cured host isolate. Finally, we measure biofilm formation and assess the plasmid fitness cost for each plasmid-host combination (including the cured hosts) under different growth conditions. Our study provides new information about how AR plasmids influence biofilm formation and other phenotypic traits in epidemiologically successful, clinical bacteria.

Room 3-06

Domestic Sewage Discharges Induce Large-scale Changes in Prokaryotic Communities of Mangrove Sediments in Serinhaém Estuary, Brazil

Carolina O de Santana[1], Pieter Spealman[2], Vania M. M. Melo[3], David Gresham[2], Taise B. de Jesus[4], Eddy J. F Oliveira[4], Fabio A. Chinalia[1]

[1]Federal University of Bahia, Salvador, Bahia, Brazil. [2]New York University, New York City, NY, USA. [3]Federal University of Ceará, Fortaleza, CE, Brazil. [4]State university of Feira de Santana, Feira de Santana, BA, Brazil

Abstract

Mangroves are tropical ecosystems with strategic importance for climate change mitigation local and global scales. They are also under considerable threat due to fragmentation, degradation, and urbanization. However, the complete understanding of how anthropogenic actions can affect microbial biodiversity and functional adaptations is still lacking. In this study, we carried out 16S rRNA gene sequencing analysis using sediment samples from two distinct mangrove areas located within the Serinhaém Estuary, Brazil. The first sampling area was located around the urban area of Ituberá, impacted by domestic sewage, while the second was an environmentally conserved site. Our results show significant changes in the structure of the communities between impacted and conserved sites. Biodiversity, along with functional potentials for the cycling of carbon, nitrogen, phosphorus and sulfur, were significantly increased in the urban area. We found the environmental factors of organic matter, pH, salinity, dissolved oxygen and lead were significantly correlated with the observed shifts in the communities. Contributions of specific taxa to the functional potentials was negatively correlated with biodiversity, such that fewer number of taxa in the conserved area contributed to the majority of the metabolic potential. The results suggest that the contamination by domestic sewage generated an eutrophic environment that may have led to the extinction of taxa observed at the conserved site. In their place we find that the impacted site is enriched in prokaryotic families that are known human and animal pathogens, a clear negative effect of the urbanization process.

Room 3-07

Asexual experimental evolution of yeast does not curtail transposable elements

Piaopiao Chen, Jianzhi Zhang
University of Michigan, Ann arbor, Michigan, USA

Abstract

Compared with asexual reproduction, sex facilitates the transmission of transposable elements (TEs) from one genome to another, but boosts the efficacy of selection against deleterious TEs. It is thus unclear whether sex has a positive net effect on TE's proliferation. A study concluded that sex is at the root of TE's evolutionary success because the yeast TE load was found to decrease rapidly in approximately 1000 generations of asexual but not sexual experimental evolution. However, this finding contradicts the maintenance of natural yeast TEs where sexual reproduction occurs extremely infrequently. Here we show that the purported TE load reduction during asexual experimental evolution was an artifact of low genomic sequencing coverages. We observe stable TE loads in both sexual and asexual experimental evolution from multiple yeast datasets with sufficient coverages. To understand the evolutionary dynamics of yeast TEs, we turn to asexual mutation accumulation (MA) lines that have been under virtually no selection. We find that both TE transposition and excision rates per generation, but not their difference, tend to be higher in environments where yeast grows more slowly. However, the transposition rate is not significantly higher than the excision rate and the variance of the TE number among natural strains is close to its neutral expectation, suggesting that selection against TEs is at best weak in yeast. We conclude that the yeast TE load is maintained largely by a transposition-excision balance and that the influence of sex remains unclear.

Room 3-08

Genomics and metabolic capabilities of *Klebsiella pneumoniae* isolates from Grey-headed flying foxes

Ben Vezina[1], Louise M Judd[1], Fiona K McDougall[2], Wayne SJ Boardman[3], Michelle L Power[2], Kathryn E Holt[1,4], Kelly L Wyres[1]

[1]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, VIC, Australia. [2]Department of Biological Sciences, Macquarie University, Macquarie University, NSW, Australia. [3]School of Animal and Veterinary Sciences, University of Adelaide, University of Adelaide, SA, Australia. [4]London School of Hygiene and Tropical Medicine, London, London, United Kingdom

Abstract

The Grey-headed flying fox (*Pteropus poliocephalus*) is an endemic Australian fruit bat, known to carry pathogens with zoonotic potential. We recently showed they harbour bacterial pathogen *Klebsiella pneumoniae* and closely related species in the *K. pneumoniae* Species Complex (*Kp*SC). *Kp*SC diversity was low within wild flying fox populations and almost uniform within captive populations, exhibiting little antimicrobial resistance and few human virulence factors. However, high-resolution genomic analysis has not been performed.

We leveraged existing draft genomes for 39 *Kp*SC isolates from Grey-headed flying foxes and generated long read data for each of thirteen distinct *Kp*SC sequence types (STs) present in the population, to construct high-quality completed genomes, to facilitate genomic comparison, to track isolate's relationships and plasmid transmission. We generated *in silico* genome scale metabolic models to predict and compare substrate usage to 59 external *Kp*SC isolates from human and environmental sources. No substrates were overrepresented in either group.

Of 13 STs comprising the flying fox isolates, five were overrepresented. High-resolution comparison indicated three STs were internally clonal. In one case, two clonal *Klebsiella africana* isolates (ST4938) were found at two flying fox populations located within flying distance of each other. Interspecies horizontal plasmid transmission between *K. pneumoniae* and *K. africana* was observed for a single plasmid, pFF1003_1, across multiple STs.

Together our data strongly indicate that *Kp*SC are able to transmit within the flying fox population or via a common nutrient source, and that these isolates can harbour plasmids with similarity to those found in human derived *Kp*SC.

Room 3-09

Exploring the adaptation of Escherichia coli under prolonged starvation: growth or stress response?

Pabitra Nandy[1,2], Savita Chib[1], Aswin Sai Narain Seshasayee[1]

[1]National Centre for Biological Sciences, Bangalore, Karnataka, India. [2]Max Planck Institute for Evolutionary Biology, Plön, Germany

Abstract

In nature, bacterial populations face constant nutrient limitation and competition for existing resources. It is thought that the bacterial genome has been shaped by a “feast-or-famine” lifestyle.

Bacterial cultures in the laboratory can survive for years without the requirement of externally supplied nutrients in a phase called the ‘Long Term’ Stationary Phase (LTSP). Studying the evolution of bacteria under prolonged starvation can shed light on their adaptation strategies in nature, as it approximates the ‘feast-or-famine’ paradigm. In LTSP, the genetic diversity of the initial, homogenous inoculum increases, and different subpopulations compete with each other for survival with new, fitter mutations constantly outcompeting older populations.

To explore bacterial adaptation in LTSP, we evolved *Escherichia coli* isolates under nutrient starvation for two months. From this experiment, we isolated a small-colony forming mutant(SCV) after one month of starvation and genetically and phenotypically characterized this mutant.

Using whole-genome sequencing we identified a single point mutation in the beta-prime subunit of RNA polymerase - ***rpoc*^{A494V}** to cause the slow growth phenotype. We also found that the slow-growing mutant outcompeted the ancestral populations depending on the environmental conditions. The competitive advantage of the mutant was also controlled by the activity/expression status of the stationary-phase sigma factor (σ^S) in the background.

In E. *coli*. rapid growth and stress response are regulated by two transcription factors- σ^D and σ^S respectively. Our work suggests that the fine-tuning of activity between these two global regulators is an active strategy to generate diverse mutants and determine their adaptive fitness across LTSP.

Room 3-10

Comparative genomics of Priestia bacteria isolated from air

Sam Spence[1], Stephan Schuster[2,3], Hielim Kim[1,2]

[1]Asian School of the Environment, Nanyang Technological University, Singapore, Singapore. [2]Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, Singapore. [3]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

Abstract

The air microbiome is an understudied ecosystem that is a medium for global dispersal and also harbours opportunistic human pathogens. Spore-forming bacteria are well equipped to survive the stresses of transport through air and have been found in samples of Singapore air – in particular, the species *Priestia megaterium* and *Priestia aryabhattai*, which have been argued to be the same species under two names.

We aimed to reveal the evolutionary relationships between the two species to resolve this taxonomic dispute, and to identify their functional differences. Using whole genome sequencing data of bacteria cultures isolated from air, our whole-genome phylogenetic analysis clearly demonstrates that *P. aryabhattai* forms a distinct clade from *P. megaterium*, suggesting that *P. aryabhattai* is a subspecies of *P. megaterium*. Interestingly, the strains isolated from air almost all belonged to the *P. aryabhattai* clade. Further, several genes were identified which differ between the two clades by copy number variations or amino acid substitutions; genes with protein functions related to sporulation and protection from oxidative stress.

These results suggest that *P. aryabhattai* has adapted to survive better while airborne compared to *P. megaterium*. To support this, we are analysing metagenomic data produced from air samples collected at different altitudes, from 1m to 3,500m. Our hypothesis is that *P. aryabhattai* will have lower abundance than *P. megaterium* at low altitudes, but higher abundance at high altitudes. We will report the result of this study.

Room 3-11

Horizontal Gene Transfer Inference: A Comparative Study Reveals Little Agreement across Methods

Swastik Mishra[1], Carolina Fanalista[2], Martin Lercher[1]
[1]Heinrich Heine University, Düsseldorf, Germany. [2]University of Bologna, Bologna, Italy

Abstract

Prokaryotes adapt to novel and changing environments predominantly through Horizontal Gene Transfer (HGT). While current research on prokaryotic evolution employs a wide variety of computational algorithms for inferring HGTs, the relative merits of these methods are unclear. A systematic comparison is complicated by a lack of representative data of known HGT events that could be used for validation. Artificial datasets created through computer simulations provide an alternative, but they rely on simplifying assumptions that may favor certain methods over others. In this study, we compare popular state-of-the-art methods used for HGT inference, examining their performance using a common dataset of sequence and phylogenetic data. We find a very low overlap in the identity of inferred HGT events even for very similar algorithms. Moreover, individual methods differ significantly in the extent to which the genes inferred to have undergone HGT agree with known biases of transferred genes, e.g., the enrichment for specific biological functions, scarcity of protein-protein interactions, and small gene length. We aim to develop this study into a benchmark of existing and future HGT inference methods, highlighting potential methodological biases and forming a rational basis for method choice.

Room 3-12

The genomic instability and evolutionary potential of interspecific yeast hybrids

Devin P Bendixsen, Rike Stelkens
Stockholm University, Stockholm, Sweden

Abstract

Hybrid lineages and hybridization is now more fully appreciated as more than just evolutionary dead ends. Hybridization generates novel genetic and potentially phenotypic diversity which can be exploited by evolutionary forces. Due to the significant level of genetic and evolutionary divergence within the *Saccharomyces* genus and the surprising ability of species within this genus to hybridize, *Saccharomyces* yeast are a promising model of hybridization. Interspecific hybrids show varying levels of success (fertility) and this success correlates with genetic distance. The genomes of hybrids often show substantial deviations from the features of the parent genomes. This includes genomic instabilities characterized by chromosomal rearrangements, gains and losses. It is yet unclear if there are any general patterns in the extent and prevalence of genomic variation and instability recurring across hybrid genomes with different genetic backgrounds. Here, we analyzed the genomic architecture of ~200 previously published *Saccharomyces* yeast hybrids. We quantified several aspects of genomic instability, including the loss and gain of chromosomes and mitochondrial inheritance. We found that genomic instability was widespread in hybrids with elevated aneuploidy and that mitochondrial content does not follow the same inheritance pattern as the nuclear genome. The plastic architecture of hybrid genomes is a recipe for their fitness and evolutionary success, generating new genetic and phenotypic diversity, the colonization of new ecological niches, and higher rates of adaptation to stress. Scrutinizing patterns of genomic instability in yeast hybrids may also serve as a model to better understand drug resistance and cell line disorders.

Room 3-13

Within and among-species variation in floral nectar microbiome composition in *Rhododendron catawbiense* and *Lobelia cardinalis* across elevation and environmental gradients.

Daniel A. Barker
East Tennessee State University, Johnson City, TN, USA

Abstract

Recent studies have uncovered novel associations between plants and microorganisms, particularly those of floral microbiome. Specifically, the nectar microbiome can qualitatively and quantitatively alter floral rewards, thus affecting pollination and plant reproductive success. However, most studies have focused on evaluating drivers of interspecific variation in nectar microbiome and how changes in environmental conditions across space (e.g. elevation) can affect the nectar microbiome is little studied. Here we explored within- and between species variation in nectar microbiome composition, diversity, and abundance and how this varies across elevation and environmental gradients in two species of plants, *Rhododendron catawbiense* (bee-pollinated) and *Lobelia cardinalis* (hummingbird-pollinated). We collected nectar samples, plated, recorded and isolated nectar fungi colonies for each species at two different elevations. Fungal ITS sequences were obtained from individual colonies by Sanger sequencing and from nectar microbiome by Nanopore NGS. Nanopore sequences were clustered by mapping against our Sanger-based reference and the UNITE database. We found significant within-species variation in fungal colony abundance (>5 times between locations) in both species. Among plant species, *R. catawbiense* had significantly more colonies compared to *L. cardinalis*, even though the latter had higher fungal diversity. Interestingly, plant species and locations shared few fungal species reflecting high levels of plant-fungi specificity. In *R. catawbiense* high elevation locations showed lower OTU diversity and high abundance of Metschnikowiaceae OTUs. Our study helps further understanding of the potential drivers of inter- and intraspecific variation in nectar microbiome composition and of its potential consequences for pollination and plant success.

Room 3-14

Evidence for selection in the abundant accessory gene content of the *Pseudomonas spp.* pangenome.

Fiona J Whelan[1], Rebecca J Hall[2], James O McInerney[1]
[1]University of Nottingham, Nottingham, United Kingdom. [2]University of Birmingham, Birmingham, United Kingdom

Abstract

Pangenomes describe the totality of genes (both core and accessory) present in a set of evolutionarily related strains. When a pangenome contains an excess of accessory genes it is said to be *open*. We hypothesize that open pangenomes are maintained by selection, and that, as such, gene gain and loss will not be random, but instead will reflect functional patterns within the pangenome. If this hypothesis is correct, we would expect to see gene associations (i.e. gene-gene co-occurrence) and dissociations (i.e. gene-gene avoidance) between functionally related genes. To test this hypothesis, we examined 209 *Pseudomonas spp.* genomes and identified gene-gene association and dissociation pairs and the connected components (i.e. gene sets) which they subsequently form. We find that the vast majority (86.7%) of abundant accessory genes (present in >5% of genomes) form statistically significant association and/or dissociation patterns with other genes. We found that, when compared to random gene pairs, co-occurring and avoiding gene pairs shared more functionality with each other, were more often expressed together, and are more likely to produce proteins that interact with each other. These results are not due to genes sharing a gene neighbourhood (i.e. being co-localized) on the chromosome. Together, these findings suggest that, overall, the abundant accessory gene content of the *Pseudomonas spp.* pangenome is structured under the constraints of natural selection. Because the *Pseudomonas spp.* pangenome shares traits with other open pangenomes, we suspect that these results are generalizable to other open prokaryote pangenomes.

Room 3-15

Impact of geographic isolation on the pangenomes of horizontally transmitted symbionts

Devani Romero Picazo[1], Tal Dagan[1], Nicole Dubilier[2], Anne Kupczok[3]
[1]Christian-Albrechts University, Kiel, Germany. [2]Max Planck Institute for Marine Microbiology, Bremen, Germany. [3]Wageningen University & Research, Wageningen, Netherlands

Abstract

Gene content diversity is described by the species pangenome and plays a crucial role in the adaptation of bacteria to their environment. Horizontal gene transfer (HGT) contributes to generate and maintain this diversity. While horizontally transmitted symbionts might have a wider potential for HGT compared to vertically transmitted symbionts, the impact that dispersal and colonization among individual hosts has on the frequency of HGT has been understudied. Here we analyze multiple deep-sea *Bathymodiolus brooksi* mussel individuals, that are co-colonized by two horizontally transmitted chemosynthetic bacterial symbiont species -one methane-oxidizing and one sulfur-oxidizing species. By using high-resolution metagenomics methods, we assess symbiont population diversity, reconstruct population pangenomes, assign genes to strains and detect strain-specific genes. We find that the pangenomes of the two symbiont populations are different in size and shape, where the methane-oxidizing symbiont has a smaller accessory genome and a larger mobilome. Additionally, we observe that gene content diversity is related to nucleotide diversity and strain composition for each species. Moreover, we find that the accessory genomes in both species are enriched in functions related to defense mechanisms, mobile elements and DNA repair. Notably, the orthologous genes between the two symbionts are mostly depleted from these categories. We conclude that the geographic isolation among symbiont populations from individual mussels limits the exposure of the symbionts to mobile genetic elements and suggest that the methane-oxidizing symbiont population is more recently associated to the mussel host than that of the sulfur-oxidizing symbiont.

Room 3-16

Innovation in the cyanobacterial OCP photo-damage protection system by neo-functionalization of the horizontally acquired FRP

Niklas Steube[1], Marcus Moldenhauer[2], Thomas Friedrich[2], Georg KA Hochberg[1]
[1]Max Planck Institute for Terrestrial Microbiology, Marburg, Germany. [2]Technische Universität Berlin, Berlin, Germany

Abstract

Horizontal gene transfer (HGT) events are common and known to provide novelty among bacteria. Complete functional units ranging from single enzymes up to multi-component molecular machines can be shared rapidly even between distantly related species, because they supply all components necessary to perform a new function. However, the acquisition of isolated novel parts to work as pieces of existing systems seems uncommon, perhaps because integrating proteins into existing functions is difficult. In this study, we show innovation in the cyanobacteria by the horizontal acquisition of the fluorescence recovery protein (FRP) into the existing orange carotenoid protein (OCP) driven photo damage protection system. We illustrate that FRP was horizontally transferred into cyanobacteria from hosts in which it fulfilled a cell-pole associated function, totally unrelated to OCPs. By applying ancestral sequence reconstruction along with biophysical characterization of protein interactions of ancestral FRPs that existed shortly before and after the HGT event with current OCP, we show that FRP could fortuitously already interact with and regulate OCPs long before the horizontal transfer event that primarily brought together these two proteins into the same proteome. Our findings suggest that new, beneficial protein-protein interactions can be created in an instant through horizontal transfer events and that this may be an underappreciated source of evolutionary innovation. They also indicate that intricate protein-protein interactions are not necessarily built by gradual evolution processes, but can arise in dramatic leaps that are driven by random transfer events.

Room 3-17

Evolution of an Adhesin Gene Family in Independently Derived *Candida* Yeast Pathogens

Bin Z He[1], Jan Fassler[1], Rachel Smoak[2], Lindse Snyder[3]

[1]University of Iowa, Biology Department, Iowa City, Iowa, USA. [2]University of Iowa, Civil and Environmental Engineering, Iowa City, Iowa, USA. [3]University of Iowa, Interdisciplinary Graduate Program in Genetics, Iowa City, Iowa, USA

Abstract

The ability to colonize and infect humans have arisen multiple times in the Ascomycete yeasts. Known by the genus name Candida, they were found to belong to different phylogenetic groups separated by close, low pathogenic potential relatives. Despite the different evolution origin, there is a shared genomic trait in the Candida pathogens, namely a rich repertoire of cell wall proteins that include the adhesins, which is crucial for biofilm formation and is a known virulence factor. In this study we characterized the evolution of a putative adhesin gene family in a newly emerging, multidrug resistant Candida pathogen, C. auris, and investigated its evolution in the separately evolved C. albicans group. In C. auris, we found that the sequence features of several homologs in this family are highly similar to known fungal adhesins. Furthermore, the predicted structure of the N-terminal domain showed conformational similarity to a group of unrelated bacterial adhesins, strongly implicating a role as bona fide adhesins. Phylogenetically, we found that this family is yeast-specific and has undergone massive expansions independently in the C. auris and the C. albicans groups. In contrast to the relatively conserved N-terminal domain, the C-terminal of the protein homologs showed rapid evolution in sequence composition and tandem repeat structures, with potential implications for functional differences. In summary, we identified a novel candidate adhesin gene family in C. auris, and showed that the same protein family was repeatedly expanded in two separate Candida species groups; albeit the C-terminal region seemed to have either evolved separately or diverged significantly.

Room 3-18

Novel empirical substitution models for specific HIV-1 proteins

Roberto Del Amparo, Miguel Arenas
University of Vigo, Vigo, Galicia, Spain

Abstract

The HIV-1 protease (PR) and integrase (IN) are common molecular targets of HIV-1 antiretroviral therapies. However, HIV-1 often evolves rapidly, leading to resistance against therapies, and thus its evolution should be considered to develop effective therapies. By contrast, and despite the HIV-1 is the most sequenced organism so far, only two empirical substitution models (HIVb and HIVw), based on the entire viral proteome, have been developed. Here we inferred new empirical substitution models to mimic the HIV-1 PR and IN evolution considering thousands of protein sequences and we found that these models more accurately describe (in terms of likelihood) the evolutionary process of HIV-1 PR and IN than currently available substitution models. The results also suggest that efforts should be made in developing protein-specific substitution models in order to improve the accuracy of modeling protein evolution.

Room 3-19

Genomic changes and the importance of the chaperonin GroEL in the evolution of highly bottlenecked bacterial populations

Beatriz Sabater-Munoz[1], Roser Montagud-Martinez[2,3], Mario A. Fares[4,5], Christina Toft[6]

[1]Integrative and Systems Biology lab, Department of Plant stress, Institute of Plant Molecular and Cellular Biology, Spanish National Research Council (IBMCP CSIC-UPV), Valenci, Spain. [2]Integrative and Systems Biology lab, Department of Plant stress, Institute of Plant Molecular and Cellular Biology, Spanish National Research Council (IBMCP CSIC-UPV), Valencia, Spain. [3]BioSystems Design lab, Program for Systems Biology of Molecular Interactions and Regulation, Institute for Integrative Systems Biology (I2SysBio CSIC-UV), Paterna, Spain. [4]Posthumous author, Dublin, Ireland. [5]Integrative and Systems Biology lab, Department of Plant stress, Institute of Plant Molecular and Cellular Biology, Spanish National Research Council (IBMCP CSIC-UPV), Paterna, Spain. [6]Panomics and Evolutionary Systems Microbiology lab, Program for Systems Biology of Molecular Interactions and Regulation, Institute for Integrative Systems Biology Spanish National Research Council (I2SysBio CSIC-UV), Paterna, Spain

Abstract

Chaperones are involved in the folding of nascent client proteins, the prevention of polypeptides aggregation, the rescue of unfolded clients due to environmental stresses and act as an evolutionary driver due to their mutational buffering capacities.

Major insect lineages have independently acquired bacterial species, mainly from Gamma-proteobacteria and Bacteroidetes class. These bacterial species could act as nutritional mutualistic factories, facultative mutualists that protect against biotic and abiotic stresses, or reproductive manipulators. Common trade among them is an increased level of genetic drift due to the small population size and the continuous population bottlenecking at each generation, processes that have shaped their genome, proteome, and morphology. Depending on the nature of the relationship, the degree of genome plasticity varies, i.e., obligate nutritional mutualistic symbionts have extremely small genomes lacking mobile elements, bacteriophages, and/or recombination machinery. Under these conditions, endosymbionts face high mutational pressures that may lead to extinction or symbiont replacement. How do they then survive for such a long evolutionary time, and why do they show genome stasis? Here we will focus on the genome changes suffered by these endosymbionts, by comparing them to their free-living relatives, and on the mutational robustness mechanisms, including the moonlighting chaperone GroEL that could explain their long prevalence from an evolutionary perspective by using experimental evolution of *E. coli* to simulate the effect of high *groEL* overexpression and strong genetic drift.

Room 3-20

RCandy: Rapid, simple, and flexible interactive visualisation of homologous recombination events in bacterial genomes in the context of phylogeny and strain metadata

Chrispin Chaguza, Stephen D Bentley
Wellcome Sanger Institute, Cambridge, United Kingdom

Abstract

Homologous recombination is a critical evolutionary process, which increases genomic diversity and adaptation in bacteria. Parallel advances in genomic sequencing technology and the development of efficient computational methods and tools have advanced our understanding of the population evolutionary biology of bacterial populations. To date, a wide range of tools exists for the identification of genomic regions with signatures of putative genetic exchanges through homologous recombination. However, few tools exist for interactively exploring recombination events in bacterial genomes to infer their potential impact on the bacteria. Here, we present RCandy, an R package for rapid, simple, flexible, and interactive visualisation of recombination events in genomes in the context of taxon metadata and phylogenetic trees. RCandy is an R package freely available for use. Source code together with a detailed tutorial is available on GitHub at https://github.com/ChrispinChaguza/RCandy.

Room 4-01

Modeling the genomic determinants of recent selection in human genes

Diego Francisco Salazar-Tortosa[1], Yi-Fei Huang[2,3], David Enard[1]

[1]University of Arizona, Department of Ecology and Evolutionary Biology, Tucson, AZ, USA. [2]Pennsylvania State University, Department of Biology, University Park, PA, USA. [3]Pennsylvania State University, Huck Institutes of the Life Sciences, University Park, PA, USA

Abstract

Multiple genomes including the human genome exhibit strong variations in local gene density, GC content, recombination rate and multiple other parameters. This heterogeneity complicates the interpretation of genome scans for recent positive selection, and the prevalence of recent positive selection in the human and other genomes is still debated as a result. For example, recombination rate strongly influences the detection of positive selection. Other genomic factors may also influence the occurrence and/or our ability to detect positive selection. Here, we introduce mixture distributions for the study of the determinants of recent selection in the human genome. This approach provides a regression model based on fitting multiple Gaussian distributions to the data of interest. We use mixture distributions to model the association between a well-known selective sweep statistic (iHS) and multiple genomic factors likely to affect positive selection if the latter was common enough. We found that a model assuming a mixture of two Gaussian distributions provided an excellent fit to iHS in the human genome, with one of the two distributions likely capturing the positively selected part of the genome. We found several factors associated with iHS, including the recombination rate, the density of regulatory elements in testis, GC-content, gene expression in immune cells, the density of mammal-wide conserved elements, and the distance to the nearest virus-interacting gene. These results support that positive selection was not uncommon in recent human evolution and highlight mixture distributions as a powerful tool to interpret recent genomic adaptation.

Room 4-02

Redefining replication in multi-ancestry genome-wide association studies

Samuel Pattillo Smith[1], Wei Cheng[1], Sahar Shahamatdar[1], Selena Zhang[1], Joseph Paik[1], Misa Graff[2], Christopher Haiman[3], T.C. Matise[2], Kari E North[2], Ulrike Peters[4], Eimear Kenny[5], Christopher Gignoux[6], Genevieve Wojcik[7], Lorin Crawford[1,8], Sohini Ramachandran[1]

[1]Brown University, Providence, RI, USA. [2]UNC - Chapel Hill, Chapel Hill, NC, USA. [3]University of Southern California, Los Angeles, CA, USA. [4]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [5]Icahn School of Medicine at Mount Sinai, New York City, NY, USA. [6]University of Colorado - Denver, Denver, CO, USA. [7]Johns Hopkins University, Baltimore, MD, USA. [8]Microsoft Research New England, Cambrodge, MA, USA

Abstract

Since 2005, genome-wide association (GWA) datasets have been largely biased toward sampling European ancestry individuals, and recent studies have shown that GWA results estimated from European ancestry individuals apply heterogeneously in non-European ancestry individuals. Here, we argue that enrichment analyses which aggregate SNP-level association statistics at multiple genomic scales—to genes and pathways—have been overlooked and can generate biologically interpretable hypotheses regarding the genetic basis of complex trait architecture. We illustrate examples of the insights generated by enrichment analyses while studying 25 continuous traits assayed in 566,786 individuals from seven self-identified human ancestries in the UK Biobank and the Biobank Japan, as well as 44,348 admixed individuals from the PAGE consortium including cohorts of African-American, Hispanic and Latin American, Native Hawaiian, and American Indian/Alaska Native individuals. By testing for statistical associations at multiple genomic scales, enrichment analyses also illustrate the importance of reconciling contrasting results from association tests, heritability estimation, and prediction models in order to make personalized medicine a reality for all.

Room 4-03

Revisiting the out of Africa event with a novel deep learning approach

Francesco Montinaro[1,2], Vasili Pankratov[2], Burak Yelmen[2], Luca Pagani[2,3], Mayukh Mondal[2]
[1]University of Bari, Bari, Italy. [2]University of Tartu, Tartu, Estonia. [3]University of Padova, Padova, Italy

Abstract

Anatomically modern humans evolved around 300 thousand years ago in Africa. Modern humans started to appear in the fossil record outside of Africa about 100 thousand years ago though other hominins existed throughout Eurasia much earlier. Recently, several researchers argued in favour of a single out of Africa event for modern humans based on whole-genome sequences analyses. However, the single out of Africa model is in contrast with some of the findings from fossil records, which supports two out of Africa, and uniparental data, which proposes a back to Africa movement. Here, we used a novel deep learning approach coupled with Approximate Bayesian Computation and Sequential Monte Carlo to revisit these hypotheses from the whole genome sequence perspective. Our results support the back to Africa model over other alternatives. We estimated that there are two successive splits between Africa and out of African populations happening around 60-90 thousand years ago and separated by 13-15 thousand years. One of the populations resulting from the more recent split has to a large extent replaced the older West African population while the other one has founded the out of Africa populations.

Room 4-04

Inferring Demographic History from Allele Frequency Spectra with Supervised Machine Learning

Linh N Tran, Connie K Sun, Ryan N Gutenkunst
University of Arizona, Tucson, AZ, USA

Abstract

In population genetics, current methods for inferring population demographic history are often challenged by the problem of intractable likelihood computation. Previously, our group had developed dadi as a software for inferring demographic history and selection, but computational expense challenges remain and significantly limit scalability. In this work, we aim to employ supervised ML to improve upon the existing dadi inference pipeline. We specifically tested the scikit-learn Random Forest and Multi-layer Perceptron Regressors for demographic parameter estimation, using AFS simulated with dadi under 1-population and 2-population models. We explored how the parameter space and variance of the training data sets can affect learning of the different ML algorithms and compared the efficiency of the ML-based approach to the original one. Some of our preliminary findings include identification of realistic and useful parameter specifications for each demographic model, as well as how different ML models have different sensitivity to variance in training data.

Room 4-05

Dissecting genomic determinants of positive selection with an evolution-guided regression model

Yi-Fei Huang
Pennsylvania State University, University Park, PA, USA

Abstract

In evolutionary genomics, it is fundamentally important to understand how characteristics of genomic sequences, such as the expression level of a gene, determine the rate of adaptive evolution. While numerous statistical methods, such as the McDonald-Kreitman test, are available to examine the association between genomic features and positive selection, we currently lack a statistical approach to disentangle the direct effects of genomic features from the indirect effects mediated by confounding factors. To address this problem, we present a novel statistical model, the MK regression, which augments the McDonald-Kreitman test with a generalized linear model. Analogous to the classical multiple regression model, the MK regression can analyze multiple genomic features simultaneously to distinguish between direct and indirect effects on positive selection. Using the MK regression, we identify numerous genomic features responsible for positive selection in chimpanzees. These features include well-known ones, such as local mutation rate, residue exposure level, tissue specificity, and immune system genes, as well as new features not previously reported, such as gene expression level and metabolic genes. In particular, we show that highly expressed genes may have a higher rate of adaptation than their weakly expressed counterparts, even though a higher expression level may impose stronger negative selection. Also, we show that metabolic genes may have a higher rate of adaptation than their non-metabolic counterparts, possibly due to recent changes in diet in primate evolution. Overall, the MK regression is a powerful approach to elucidate the genomic basis of adaptation.

Room 4-06

Detecting Signals of Natural Selection via Deep Learning

Xinghu Qin[1], Charleston Chiang[2], Oscar Gaggiotti[1]

[1]Center for Biological Diversity, Sir Harold Mitchell Building, University of St Andrews, St Andrews, Fife, United Kingdom. [2]Center for Genetic Epidemiology, Keck School of Medicine & Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California, USA

Abstract

Detecting signatures of natural selection is of fundamental interests because it is a major evolutionary force underlying species local adaptation. However, existing genome scan methods mainly identify alleles exhibiting linear gradients with environmental factors, lacking the power to detect signatures under non-monotonic selection. Although deep learning has shown impressive power in molecular biology, such as variant calling and gene expression prediction, the application of deep learning to detect natural selection is still in its infancy. We proposed a pair of machine learning-based approaches to decipher the signatures associated with arbitrary and non-monotonic selection patterns. Our framework first inferred the highly accurate geographic genetic structure by Kernel Local Fisher Discriminant Analysis of Principal Components (KLFDAPC), which we showed to significantly improve the accuracy of predicting the geographic origin of individuals compared to Principal Components Analysis (PCA). Based on this, we then devised *DeepGenomeScan*, which can identify the outlier loci that contribute the most to the prediction of either geographic coordinates or KLFDAPC genetic features. We showed through both simulations and application to real human genetic data that our framework had significantly greater detection power than PCA and redundancy analysis (RDA)-based genome scans. *DeepGenomeScan* can pinpoint signals of selection under coarse and non-linear environmental gradients, and accurately recapitulate signals of geographic selection in European populations, including identifying a number of novel candidate genes. The *DeepGenomeScan* framework can be extended to various Omics-based genome-wide association studies, providing a general deep learning-based framework for the study of natural selection across a range of species.

Room 4-07

Sequence evolution modelling using variational self-supervised learning framework

Amine Remita, Abdoulaye Baniré Diallo
Université du Québec à Montréal, Montréal, Québec, Canada

Abstract

The challenging task of estimating molecular evolutionary parameters is a ground foundation of several applications such as prediction of mutation effect on functions, protein and vaccine design, recombination detection, viral co-infection studies and surveillance of emergent viruses. Most of the applied computational methods to estimate these evolutionary parameters rely on reconstructing phylogenies from a multiple sequence alignment.

Here we introduce a phylogeny-free probabilistic framework that models ancestral states and substitution model parameters as latent variables, and approximates their joint posterior density with variational Bayesian inference, which leverages new advanced techniques using deep learning. The framework incorporates a self-supervised learning algorithm with an efficient estimator that allows to jointly infer evolutionary related-latent variables (branch lengths and substitution rates), reconstruct ancestral states and generate a distribution of sequence alignments. Similar to phylogenetic reconstruction, we explicitly integrate a Markov chain substitution model (including JC69, HKY and GTR) into our learning algorithm.

The proposed algorithm architecture consists of two main sub-models: 1) a variational encoder that infers the parameters of evolutionary-latent-variable distributions and allows sampling; and 2) a generative model that computes probability transition matrices from sampled latent variables and generates a distribution of sequence alignments from reconstructed ancestral states. The model training is based on maximizing the evidence lower bound on the marginal likelihood of the alignment. Both sub-models are trained jointly using gradient back-propagation with stochastic gradient descent.

We highlight several solutions to challenges affecting the performance such as the variance of the estimator, neural encoder types, convergence and prior distributions.

Room 4-08

Addressing local ancestry, past demography and selection with genealogies and ancient human DNA

Alice Pearson[1], Richard Durbin[1], Eske Willerslev[1,2], GeoGenetics Meso-Neo Consortium[2]
[1]University of Cambridge, Cambridge, United Kingdom. [2]University of Copenhagen, Copenhagen, Denmark

Abstract

Recently, two new approaches have transformed our understanding of human population history. Firstly, sequencing of ancient genomes gives us a snapshot of past genetic variation. We can therefore make inferences from observed genetic signatures present before historical events such as population bottlenecks and natural selection have obscured them from the modern gene pool. Ancient DNA has thus revealed what cannot be determined from modern genomes alone. Secondly, the development of methods that aim to reconstruct population genealogies from genetic variation data. Together with an understanding of how evolutionary processes alter genealogies this has allowed inference of historical and ongoing processes in real world populations.

The latest updates in these approaches now allow us to combine the two to infer genealogies involving both present-day and ancient individuals. I will discuss a method, using machine learning and tree sequences built from ancient and present-day genomes from Europeans and West Asians. The method allows us to infer local ancestry along each sample chromosome and subsequently explore time-resolved histories of selection and population size in a structured population.

Room 4-09

Approximate Bayesian computation untangles signatures of contemporary and historical hybridization between two endangered species

Hannes Dittberner[1], Aurélien Tellier[2], Juliette de Meaux[1]
[1]University of Cologne, Cologne, Germany. [2]Technical University Munich, Munich, Germany

Abstract

Contemporary gene flow, when resumed after a period of isolation, can have crucial consequences for endangered species, as it can both increase the supply of adaptive alleles and erode local adaptation. Determining the history of gene flow and thus the importance of contemporary hybridization, however, is notoriously difficult. Here, we focus on two endangered plant species, *Arabis nemorensis* and *A. sagittata*, which hybridize naturally in a sympatric population located on the banks of the Rhine. Using reduced genome sequencing, we determined the phylogeography of the two taxa but report only a unique sympatric population. Molecular variation in chloroplast DNA indicated that *A. sagittata* is the principal receiver of gene flow. Applying classical D-statistics and its derivatives to whole-genome data of 35 accessions, we detect gene flow not only in the sympatric population but also among allopatric populations. Using an Approximate Bayesian computation approach, we identify the model that best describes the history of gene flow between these taxa. This model shows that low levels of gene flow have persisted long after speciation. Around 10 000 years ago, gene flow stopped and a period of complete isolation began. Eventually, a hotspot of contemporary hybridization was formed in the unique sympatric population. Occasional sympatry may have helped protect these lineages from extinction in spite of their extremely low diversity.

Room 4-10

Introgression Detection from Genomic Data: A Supervised Learning Approach

Yining Liu[1], Miriam Miyagi[2], Andrew J. Blumberg[1]
[1]Columbia University, New York, NY, USA. [2]Harvard University, Boston, MA, USA

Abstract

We describe a supervised machine learning approach to detecting introgression directly from genomic data via classification of a low-dimensional representation. ILUSV (Introgression Locating Using Support Vectors) represents each genomic feature using ordered pairwise divergence and employs a support vector machine to label each genomic region with an estimate for the probability of introgression. We show that ILUSV outperforms standard introgression probability estimation techniques (e.g., those using D-statistics, FST , or Gmin) on genomic windows via coalescent simulations. Interestingly, preliminary experiments suggest that ILUSV is robust against certain model misspecification and can be applied in a wide range of demographic parameters.

Room 4-11

A probabilistic model for indel evolution: differentiating insertions from deletions

Gil Loewenthal[1], Dana Rapoport[1], Oren Avram[1], Asher Moshe[1], Elya Wygoda[1], Alon Itzkovitch[1], Omer Israeli[1], Dana Azouri[1], Reed A. Cartwright[2], Itay Mayrose[1], Tal Pupko[1]
[1]Tel Aviv University, Tel Aviv, Israel. [2]Arizona State University, Tempe, Arizona, USA

Abstract

Insertions and deletions (indels) are common molecular evolutionary events. However, probabilistic models for indel evolution are under-developed due to their computational complexity. Here we introduce several improvements to indel modeling: (1) while previous models for indel evolution assumed that the rates and length distributions of insertions and deletions are equal, here, we propose a richer model that explicitly distinguishes between the two; (2) We introduce numerous summary statistics that allow Approximate Bayesian Computation (ABC) based parameter estimation; (3) We develop a model-selection scheme to test whether the richer model better fits biological data compared to the simpler model. Our analyses suggest that both our inference scheme and the model-selection procedure achieve high accuracy on simulated data. We further demonstrate that our proposed indel model better fits a large number of empirical datasets and that, for the majority of these datasets, the deletion rate is higher than the insertion rate.

Room 4-12

DNADNA: Deep Neural Architectures for DNA, a toolbox for population genetics tasks

Jean Cury, Théophile Sanchez, Erik Madison Bray, Pierre Jobic, Guillaume Charpiat, Flora Jay
LISN, U. Paris Saclay, INRIA, CNRS, UMR 9015, Orsay, France

Abstract

In recent years, deep learning gained lots of traction in different fields, due to advances in computational power, new algorithms and data deluge. The population genetic field has not escaped these trends, and many researchers have applied deep learning techniques to population genetic tasks. These deep learning methods have been used to infer selection, demography, recombination rate or detection of introgression (see for instance Sheehan et al, 2016, Flagel et al 2018 or Sanchez et al 2020). However, most deep learning methods are currently hard to reuse or to adapt for non-expert users, despite being open source. Here we present dnadna, a toolbox for using deep neural architecture directly on DNA or genetic data.
dnadna will allow researchers to focus on their research project, be it the analysis of population genetic data or building new methods, without the need to focus on proper development methodology (unit test, continuous integration, documentation, etc.). Results will thus be more easily reproduced and shared. Having a common interface will also decrease the risk of bugs. dnadna can also be used as a tool for teaching, since it allows in a short amount of time, to reuse a network, or build a new one while saving time on technical requirements necessary for coding properly an end-to-end deep learning approach.

Room 4-13

Seeing the (random) forests for the genes: a machine learning approach to pangenome structure and evolution

Maria Rosa Domingo-Sananes[1], James O McInerney[2]

[1]Nottingham Trent University, Nottingham, United Kingdom. [2]University of Nottingham, Nottingham, United Kingdom

Abstract

Gene content varies across different evolutionary scales, from species to life's domains. Many factors (phylogeny, genome structure, adaptation and variable rates of gene gain and loss) contribute to shaping the patterns of gene presence and absence in genomes. In order to tease out the reasons why a gene might be present in a particular genome, we ask if we can predict whether or not a gene is likely to be present, based on the gene content of the rest of the genome. The ability to predict gene presence can imply the existence of interactions between genes, or shared selective pressures that act on those genes. To test this idea, we trained random forest classifiers to predict the presence of 2,592 variable genes whose presence is scattered across a phylogeny, in 409 genomes of the nitrogen-fixing bacterium *Rhizobium leguminosarum*. The presence of some genes is unpredictable, while for 131 we obtained high-performing classifiers (accuracy and F-score >0.95 for five independent runs of the classifier fitting). Using the training results from the classifiers we extract feature importances to construct a weighted, directed network that indicates how genes contribute to explain the presence of each other. We find groups of tightly connected genes that could signify biologically relevant relationships. We expect that this approach can be easily extended to include further predictive/explanatory variables, such as phenotypic and environmental information as well as core genome sequence variation.

Room 4-14

Statistical method for inferring indel dynamics and its application on conserved domains

Elya Wygoda, Gil Loewenthal, Tal Pupko
Tel Aviv University, Tel Aviv, Israel

Abstract

Insertions and deletions (indels), together with substitutions and genome rearrangement events are the fuel that drives genome evolution among organisms across the tree of life. Deep understanding of indel evolutionary dynamics is thus of great importance for reconstructing the tree of life and for inferring ancestral sequences. Most of the studies concerning indel dynamics reported a deletion bias, meaning that deletions are more ubiquitous than insertions. In a previous study of multiple sequence alignments from the Conserved Domain Database (CDD), it was claimed that in ancient domain families there is an insertion bias. This and other studies of indels dynamics were inferred by ad-hoc methods that involve gap counting in alignments. However, there is no one-to-one relation between gaps and indels, as gaps can be the outcome of more than one event. Thus, counting methods do not account for these scenarios, and therefore these methods introduce biased results, similar to parsimony methods for substitutions. To remedy this limitation, we developed a statistical methodology based on Approximate Bayesian Computation for inferring indel parameters. We also use a model-selection scheme that allows testing if deletions and insertions have different dynamics in a given dataset in question. We apply our methodology to the CDD data for which it was previously claimed that an insertion bias exists. We show that when more rigorous methods are used - an insertion bias is no longer supported.

Room 4-15

Impact of contrasting scenarios of evolution on virus tree shape

Lucia Paiva Barzilai, Carlos Eduardo Guerra Schrago
Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Abstract

Phylodynamic is the research field that investigates infectious diseases behaviour based on their evolutionary and ecological process. RNA viruses are intensively studied as they accumulate genetic differences in an observable time scale, enabling to retrieve information from molecular evolutionary methods and phylogenetic tree shape. Particularly, HIV phylogenetic trees can vary their shape between intra and interhost analyses, which are frequently associated with the action of immune selection, being responsible for the asymmetry observed in intrahost phylogenies. To test this hypothesis, it would be interesting to elucidate the role of a null model of population dynamics trees topology. In this work, we use simulated data to account to the real impact of selection on tree shape and search for features on sequence data and topologies that can be used to distinguish different viruses and scenarios of evolution, using machine learning approaches. We used SANTA-SIM software to generate simulated data in two contrasting scenarios of evolution based on an HIV infection. Sequence data and tree shape analyses were conducted using typical metrics available in R software. Our initial results do not show differences on values distribution in any of the tree shape metrics. However, we could observe differences on sequence data enough to distinguish both scenarios using a support vectors machine. Further, we are developing a forward in time simulation program that accounts for both intra and interhost scenarios, with pathogens evolution during an epidemic, in other to analyse other evolutionary and epidemiological parameters that can retrieve the observed tree shape patterns.

Room 4-16

Modeling satellite DNA organization

Sherif Negm, John Sproul, Tuan Pham, Songeun Lee, Amanda Larracuente
University of Rochester, Rochester, New York, USA

Abstract

Repetitive DNAs comprise large portions of eukaryotic genomes. Satellite DNAs (satDNAs) are abundant tandemly repeated DNA sequences found near centromeres, telomeres, and on sex chromosomes. SatDNAs originate through polymerase slippage, recombination between repeat elements, or TE-mediated mechanisms. Arrays of satDNA repeats are highly dynamic over short periods of evolutionary time: they vary in copy number and organization through unequal exchange, mutation, and other processes. The expansion of satDNA arrays is thought to decrease organismal fitness but the relative importance of processes shaping satDNA evolution in natural populations is poorly understood. Models of unequal crossing over and selection on satDNA arrays mainly focused on copy number number changes. We developed a model that tracks structural changes and mutations in satDNA arrays. Our model simulates both sequence and copy number evolution for a population of size n over x generations. We also designed a new method to determine the site of recombination breakpoint and incorporate random nucleotide mutation and natural selection on copy number. Here we simulate the effects of recombination, drift, and selection on satDNA arrays in populations. We established a Bayesian-inference procedure to infer recombination rate from population satDNA data. Introducing sequence composition and organization to models of satDNA evolution may prove useful for estimating the impact of recombination and natural selection on satDNA arrays from empirical data.

Room 4-17

Archetypal Analysis for Population Genetics

Julia Gimbernat[1], Daniel Mas[2], Carlos Bustamante[2], Alex Ioannidis[2]
[1]Imperial College London, London, United Kingdom. [2]Stanford University, Stanford, California, USA

Abstract

The estimation of individual ancestry clusters using genomic data has many applications in population genetics, such as the identification and characterization of admixed individuals; and is expected to play an important role in personalized medicine. Statistical models developed for this task are often supervised, which require reference individuals that are assigned to homogeneous ancestral populations. In this study, we explore archetypal analysis as an unsupervised clustering approach for associating individuals to populations. This is important for performing analyses that are free from predefined ethnic or racial labels that have often been considered social constructs and can fail to capture genomic variation. We find that archetypal analysis has a similar performance accuracy to previous state-of-the-art clustering methods such as ADMIXTURE, while providing more flexibility such as enabling users to work with lower-dimensional compact representations of genetic data. These advantages will be illustrated using a collection of whole genomes from dogs and from humans.

Room 4-18

Deep residual neural networks resolve quartet phylogenies with heterogeneous sequence evolution

Zhengting Zou[1], Hongjiu Zhang[2], Jianzhi Zhang[3]

[1]Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [2]Microsoft, Inc., Bellevue, USA.

[3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, USA

Abstract

Phylogenetic inference based on molecular sequences has become a fundamental and routine task in evolutionary and other biological studies. Although many phylogenetic methods have been developed to explicitly take into account substitution models of sequence evolution, such methods could fail due to model misspecification or insufficiency, especially when there are heterogeneities in substitution processes across sites and among lineages. In this study, we propose to infer topologies of four-taxon trees by deep residual neural networks, a machine learning approach needing no explicit modeling of the sequence evolution process and having succeeded in solving complex nonlinear inference problems. We train residual networks on simulated protein sequence data with extensive amino acid substitution heterogeneities. We show that the well-trained residual network predictors can outperform existing state-of-the-art inference methods such as the maximum likelihood method on diverse simulated test data, especially under extensive substitution heterogeneities. Reassuringly, residual network predictors generally agree with existing methods in the trees inferred from real phylogenetic data with known or widely believed topologies. Furthermore, when combined with the quartet puzzling algorithm, residual network predictors can be used to reconstruct trees with more than four taxa. We conclude that deep learning represents a powerful new approach to phylogenetic reconstruction, especially when sequences evolve via heterogeneous substitution processes. We present our best trained predictor in a freely available program named Phylogenetics by Deep Learning (PhyDL, https://gitlab.com/ztzou/phydl)

Room 4-19

Neural Networks as Optimal Estimators of the Mutation Rate or the Effective Population Size for Variable Recombination Rates

Klara E Burger[1], Peter Pfaffelhuber[2], Franz Baumdicker[1]
[1]University of Tuebingen, Tuebingen, Germany. [2]University of Freiburg, Freiburg, Germany

Abstract

Although ML methods in population genetics face multiple specific challenges, it is meanwhile clear that they are a promising technique to build more powerful inference tools, especially for problems that are hard to tackle with classical methods. In contrast, in population genetics, many theoretical results have been developed within the last decades and enabled us to identify the best inference technique for some scenarios. For example, the estimation of mutation rate, or equivalently effective population size, is well understood, at least if recombination is low or high. In these two scenarios the best estimation methods, namely Fu's and Watterson's estimator, are known and well understood. For intermediate recombination rates, the theoretical development of optimal estimators however is much more involved. ML tools could help to develop good estimators in these involved scenarios, but it is difficult to assess how well suited these tools are for different applications in population genetics if no benchmark is available. Here we investigate simple feed-forward neural networks for the estimation of the mutation rate and compare their performance with the frequently used optimal estimators. We find that neural networks can reproduce the known optimal estimators if provided with the appropriate features. Remarkably, only one hidden layer is necessary to obtain a single estimator that performs almost as well as the optimal estimators for low and high recombination rates and provides a superior estimation method for intermediate recombination rates simultaneously. This framework also enables us to identify phylogenetic aspects of features learned by various ML techniques.

Room 4-20

Molecular Basis of the Emergence of Eusociality in Bees and Wasps

Katherine Geist[1], Emeline Favreau[2], Sandra Rehan[3], Seirian Sumner[2], Amy Toth[1]
[1]Iowa State University, Ames, IA, USA. [2]University College London, London, United Kingdom. [3]York University, Toronto, ON, Canada

Abstract

The evolution of eusociality requires individuals to forgo some or all their own reproduction to assist the reproduction of others in their group, such as a primary egg-laying queen. The emergence of eusociality has been challenging to explain, especially at early stages, because individuals retain autonomy regarding group membership and reproductive choice. Although inclusive fitness theory may explain the evolution of eusociality through indirect fitness benefits to the helper, it is less clear how often eusociality has emerged via similar genes and pathways. Here, we examine the mechanisms of molecular convergence across two Hymenopteran families – bees and wasps. Within each family, we assess four independent origins of eusociality, two at the family-level in the bees and two at the sub-family level in wasps. Using six species across four independent origins of eusociality, we compare previously published transcriptomes in a meta-analytical framework. Across the six species, social life history occurs in fairly small groups of individuals that vary from facultative to obligately social. Previous work has shown transcriptome-level differences between reproductive and helper phenotypes based on comparisons to more highly eusocial species. We compare patterns of brain transcriptome-wide gene expression between primary reproductives and helpers across these origins at the gene, functional, and network levels. We also apply a machine learning approach to look for more nuanced differences in gene expression between reproductive and helper phenotypes across origins. We use this information to assess the relative importance of phylogenetic relatedness and social life history in shaping the evolution of brain gene expression patterns.

Room 5-01

Bioarchaeological evidence of one of the earliest Islamic burials in the Levant

Megha Srigyan[1], Héctor Bolívar[2,3], Irene Ureña[2], Jonathan Santana[4], Andrew Petersen[5], Eneko Iriarte[6], Emrah Kırdök[2,7], Nora Bergfeldt[2], Alice Mora[8], Mattias Jakobsson[1], Colin Smith[6,8], Juan José Ibañez[9], Anders Götherström[2], Torsten Günther*[1], Cristina Valdiosera*[6,8]

[1]Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden. [2]Centre for Palaeogenetics, Stockholm, Sweden. [3]Instituto del Patrimonio Cultural de España, Madrid, Spain. [4]Department of Historical Sciences, Universidad de Las Palmas de Gran Canaria, Las Palmas de G.C, Spain. [5]University of Wales Trinity Saint David, Lampeter, United Kingdom. [6]Laboratorio de Evolución Humana, Departamento de Historia, Geografía y Comunicación, Universidad de Burgos, Burgos, Spain. [7]Department of Biotechnology, Mersin University, Mersin, Turkey. [8]Dept. Archaeology and History, La Trobe University, Melbourne, Australia. [9]Archaeology of Social Dynamics, Milà i Fontanals Institution, Spanish National Research Council (CSIC), Barcelona, Spain

Abstract

The Middle East plays a central role in human history harbouring a vast diversity of ethnic, cultural and religious groups. However, much remains to be understood about past and present genomic diversity in this region. We present a multidisciplinary bioarchaeological analysis of two individuals dated to late 7th and early 8th centuries from Tell Qarassa in modern-day Syria. Radiocarbon dates and burial type are consistent with one of the earliest Islamic Arab burials in the Levant during the Late Antiquity period. Interestingly, we found genomic similarity to a group of modern-day Bedouins and Saudi rather than to most neighbouring Levantine groups. Our study represents the first genomic analysis of a secondary use site with characteristics consistent with an early Islamic burial in the Levant. We discuss our findings and possible historic scenarios and their possible interaction with religious and cultural processes (including diet and subsistence practices) as well as disease.

Room 5-02

Two pheasant hybrid zones on Tibetan Plateau

Pengcheng Wang[1], Zhengwang Zhang[2], Nan Wang[3]
[1]Chinese academy of Sciences, Beijing, China. [2]Beijing Normal University, Beijing, China. [3]Beijing Forest University, Beijing, China

Abstract

Hybrid zone is a natural laboratory for investigating speciation process. Here, we integrated population genomic data (19,960 SNPs), environmental variations (temperature, precipitation, slope, altitude, rivers) and robust sampling (154 samples) of Tibetan Eared Pheasant (*Crossoptilon harmani*) and White Eared Pheasant (*C. crossoptilon*) to evaluate hybrid zone, investigate demographic history, explore evolutionary driver of the lineage divergence involved in hybrid zone, and infer the potential evolutionary future of hybrid zone. The population genetic data suggest *C. crossoptilon* had two genetic groups, and we classify them as *C. c. drouynii* and *C. c. crossoptilon*. We document two hybrid zones, one involved in *C. c. drouynii* and *C. harmani* located in Eastern Tibetan Plateau, the other involved in *C. c. drouynii* and *C. c. drossoptilon* located in Western Sichuan. Demographic history show secondary contact is the reason for forming hybrid zones. The generalized dissimilarity model uncovered geographical distance, not environmental variations or landscape resistance, best correlated with genetic allele frequency, suggesting environmental selection along the environmental gradients has less effect on the fitness divergence. The lineages involved in hybrid zone who have similar fitness may merge into one lineage. Thus, we propose the potential evolutionary future of these two pheasant hybrid zones may be fusion. To our knowledge, this is the first study to document pheasant hybrid zone with population genomic data and infer its potential evolutionary future, which will deep our understanding of biodiversity. The findings will pave the way for a detailed investigation of ecological and genetic basis for speciation.

Room 5-03

Inference of admixture history in South Asians using mutation spectrum information

Luke Zhu, Kelley Harris
University of Washington, Seattle, Washington, USA

Abstract

Human population groups have distinctive mutation spectra, for instance, it was found that the proportion of TCC>TTC mutations is elevated in Europeans, and to a lesser extent, in South Asians (SAS). This can be explained if the elevated TCC>TTC signal first arose in Europeans and was later introduced to the SAS population through admixture. To infer the timing of this event, we build upon a class of methods that utilizes the association between pairs of loci in the admixed population. In general, admixture leaves behind admixture linkage disequilibrium (ALD) signals. This signal, amplified with appropriately chosen weights, has been shown to follow an exponential decay as a function of genetic distance, and the decay rate constant gives the age of admixture. Here, we propose a novel weighting scheme by leveraging the 3-mer mutation spectrum information. We first infer a 3-mer mutation signature associated with the elevated TCC>TTC signal and take the difference between this inferred signature and the 3-mer spectrum of modern SAS, the result of which is a 96-component weight vector. Each single nucleotide polymorphism (SNP) in our analysis is then assigned a corresponding weight depending on its 3-mer mutation context.

This method will allow us to infer the timing of admixture in SAS without explicitly defining and using genetic data from the ancestral European population, and can be potentially helpful in the future to determine the original source European population for the TCC>TTC signal given the scarcity of ancient ancestral DNA.

Room 5-04

Insights into *Mus musculus* subspecies population structure across Eurasia revealed by whole-genome sequence analysis

Kazumichi Fujiwara[1], Yosuke Kawai[2], Kazuo Moriwaki[3,4], Toyoyuki Takada[3], Toshihiko Shiroishi[3], Naruya Saitou[4], Hitoshi Suzuki[1], Naoki Osada[1]

[1]Hokkaido University, Sapporo, Hokkaido, Japan. [2]National Center for Global Health and Medicine, Shinjuku, Tokyo, Japan. [3]RIKEN BioResource Research Center, Tsukuba, Ibaraki, Japan. [4]National Institute of Genetics, Mishima, Shizuoka, Japan

Abstract

Mus musculus is composed of three major subspecies: The South Asian subspecies (*M. m. castaneus*: CAS), the North Eurasian subspecies (*M. m. musculus*: MUS), and the West European subspecies (*M. m. domesticus*: DOM). They are presumed to have originated in the Indian subcontinent and diverged into subspecies at about the same time about 1 million years ago. The habitation of the mice expanded extensively with human migration within 10,000 years after the onset of agriculture, so hybridization between subspecies is thought to have occurred relatively recently. A large number of laboratory strains of *M. musculus* have been established as experimental model animals and are widely used by researchers around the world. Although these laboratory mice have historically been well studied, yet the genetic background and evolutionary history of wild mice remains unclear.

In this study, we newly sequenced the whole genomes of 98 wild mice collected mainly from Eurasia and analyzed the patterns of genetic differentiation among the subspecies. The results of the genome-wide analysis showed that MUS, CAS, and DOM could be divided into three major groups, supporting the hypothesis that had been considered. Although, the patterns of MUS and CAS suggested that there is more inter-subspecies hybridization than previously thought, especially in the large geological regions of China and the Japanese archipelago. Validation using *f*-statistics allowed us to choose the least hybridized samples among the three major subspecies. Furthermore, simulations of population models using genome-wide SNPs suggest that the sister clades of DOM-MUS diverged about 250,000 years ago.

Room 5-05

Genomic diversity and post-admixture adaptation in the Uyghurs

Yuwen Pan[1], Chao Zhang[1], Yan Lu[1], Zhilin Ning[1], Donesheng Lu[1], Yang Gao[1], Xiaohan Zhao[1], Yajun Yang[2], Yaqun Guan[3], Dolikun Mamatyusupu[4], Shuhua Xu[1]

[1]CAS-MPG Partner Institute for Computational Biology, Shanghai, Shanghai, China. [2]Fudan University, Shanghai, Shanghai, China. [3]Xinjiang Medical University, Urumqi, China. [4]Xinjiang University, Urumqi, China

Abstract

Population admixture results in genome-wide combinations of genetic variants derived from different ancestral populations of distinct ancestry, thus providing a unique “natural experiment” for understanding the genetic determinants of phenotypic variation in humans. Here, we used whole-genome sequencing of 92 individuals with high coverage (30–60×) to systematically investigate genomic diversity in the Uyghurs living in Xinjiang (XJU), an admixed population of both European-like and East-Asian-like ancestry. The XJU population shows greater genetic diversity, especially a higher proportion of rare variants compared with their ancestral source populations, corresponding to greater phenotypic diversity of XJU. Admixture-induced functional variants in *EDAR* were associated with the diversity of facial morphology in XJU. Interestingly, the interaction of functional variants between *SLC24A5* and *OCA2* likely influences the diversity of skin pigmentation. Notably, selection has seemingly been relaxed or canceled in several genes with significantly biased ancestry, such as *HERC2–OCA2*. Moreover, signatures of post-admixture adaptation in XJU were identified, including genes related to metabolism (e.g., *CYP2D6*), digestion (e.g., *COL11A1*), olfactory perception (e.g., *ANO2*), and immunity (e.g., *HLA*). Our results demonstrated population admixture as a driving force, locally or globally, in shaping human genetic and phenotypic diversity as well as in adaptive evolution.

Room 5-06

Reduction and Loss of Mitochondrial Function in Novel Anaerobic Microbial Eukaryotes

Shelby K. Williams[1], Jon Jerlström Hultqvist[1], Yana Eglit[1], Russell Orr[2], Dayana Salas-Leiva[1], Bruce Curtis[1], Alastair G. B. Simpson[1], Andrew J. Roger[1]

[1]Center for Comparative Genomics & Evolutionary Bioinformatics, Halifax, Nova Scotia, Canada.

[2]Bryozoan Lab for Ecology, Oslo, Norway

Abstract

The taxonomic super group Metamonada contains a variety of both free-living and parasitic protists that inhabit highly diverse anaerobic environments, from human guts to estuarine sediments. Metamonads do not possess canonical mitochondria. Instead, they contain a diversity of mitochondrion-related organelles (MROs), which are adapted to anaerobiosis. These organelles lack typical aerobic mitochondrial functions but possess anaerobic metabolic systems, some of which were acquired by lateral gene transfer.

Here we investigate the genomes of five previously uncultured anaerobic protist species informally named the ‘fornifriends’ to better understand the evolutionary mechanisms underpinning mitochondrial reduction. Genomes for these protists were assembled into high-quality draft sequences using a mixture of both long-read (Oxford Nanopore) and short-read (Illumina) sequencing technologies. Gene predictions were aided by RNA-Seq data gathered from each species. Our phylogenomic analyses reveal that the fornifriends form a robust clade within Metamonada.

We also searched the gene predictions and scaffolds of each organism for genes encoding canonical mitochondrial markers and anaerobic metabolic enzymes. There was significant variation in mitochondrial function amongst the fornifriend species with some organisms appearing to have hydrogen producing MROs, while others appear to function only in iron-sulfur synthesis. For one of the species, no putative mitochondrial markers were found at all. Loss of canonical mitochondrial systems within some of the lineages of this group appears to be related to the acquisitions of new functions by LGT, including the SUF iron-sulfur cluster system.

Room 5-07

A modified fluctuation assay reveals a natural mutator phenotype that drives mutation spectrum variation within Saccharomyces cerevisiae

Pengyao Jiang[1], Anja R Ollodart[1], Vidha Sudhesh[1], Alan J Herr[1], Maitreya J Dunham[1], Kelley Harris[1,2]
[1]University of Washington, Seattle, WA, USA. [2]Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Mutations are the source of genetic variation and a prerequisite for evolution. Despite their fundamental importance, however, their rarity makes them expensive and difficult to detect, which has limited our ability to measure the extent to which mutational processes vary within and between species. Here, we use the 1011 *Saccharomyces cerevisiae* collection to measure variation of mutation rates and spectra among strains isolated from a variety of natural and human-related environments. The mutation spectra of variants segregating in different *S. cerevisiae* populations exhibit differences in the relative numbers of transitions and transversions. To directly test how much of the observed mutation spectrum variation is caused by genetic differences between extant strains of *S. cerevisiae*, we developed an experimental pipeline to assay *de novo* mutation rates and spectra of individual strains, using the reporter gene *CAN1*. We found a 10-fold mutation rate variation among 16 haploid strains surveyed. While many strains exhibit similar mutation spectra, two related strains from the “Mosaic beer” clade, known as AEQ and AAR, share a distinctive mutation spectrum enrichment for C>A mutations. This C>A enrichment found through our experimental pipeline mirrors an enrichment of C>A mutations in rare variants segregating throughout the genomes of AEQ and AAR with additional strains. Such enrichment is also found in the diploid ancestor. We further identified a mutator allele in the *OGG1* that showed an elevated mutation rate. Combining evidence from both natural variants and *de novo* mutations, we were able to identify a naturally occurring mutator allele.

Room 5-08

High-throughput functional analysis of natural variants in yeast

Chiann-Ling C Yeh[1], Andreas Tsouris[2], Joseph Schacherer[2], Maitreya J Dunham[1]
[1]University of Washington, Seattle, Washington, USA. [2]University of Strasbourg, Strasbourg, Grand Est, France

Abstract

How natural variation affects phenotype is difficult to determine given our incomplete ability to deduce the functional impact of polymorphisms detected in a population. Despite advances in computational and experimental tools that predict and measure allele function, currently no assay does so in a high-throughput manner while also representing alleles found in natural populations. Here we present such an assay that measures the fitness of hundreds of natural alleles of a given gene without site-directed mutagenesis or DNA synthesis. With a large collection of over 1,000 diverse *Saccharomyces cerevisiae* isolates, we piloted this technique using the gene *SUL1*, a high-affinity sulfate permease that, at increased copy number, can improve the fitness of cells grown in sulfate-limited media. We cloned and barcoded all alleles from natural isolates *en masse* and paired barcodes with their respective variants using PacBio long-read sequencing and a novel error-corrected algorithm. We transformed the reference S288C strain with this library and used barcode sequencing to track growth ability in sulfate limitation of lineages carrying each allele. We show that we can measure fitness conferred by each allele and stratify functional, intermediate, and nonfunctional alleles on a species-wide scale. Integrating fitness results on a phylogenetic tree, we observe how often loss-of-function occurs and whether or not there is an evolutionary pattern to our observable phenotypic results. Our method complements classic genotype-phenotype mapping strategies and demonstrates a high-throughput approach for understanding the effects of polymorphisms on a species-wide scale which can greatly propel future investigations into quantitative traits.

Room 5-09

Geographic and genetic variation in isolated island populations of *Salmonella enterica* serovar Typhi

Zoe A. Dyson[1,2,3,4], Sebastian Duchene[5], Gordon Dougan[3], Stephen Baker[3], Kathryn E Holt[1,2]
[1]London School of Hygiene & Tropical Medicine, London, United Kingdom. [2]Monash University, Melbourne, Australia. [3]University of Cambridge, Cambridge, United Kingdom. [4]Wellcome Sanger Institute, Cambridge, United Kingdom. [5]University of Melbourne, Melbourne, Australia

Abstract

Salmonella Typhi is the aetiological agent of typhoid fever. Each year >10 million cases occur worldwide of which >100,000 are associated with mortalities. While the population structure of *S.* Typhi in continental locations throughout Asia and Africa has been well characterised, the evolution of geographically isolated island populations has been largely overlooked. Comparative analysis of *S.* Typhi genomes from island and continental locations throughout Sub-Saharan Africa, Southeast Asia, South Asia, and the Pacific revealed that island populations were typically dominated by one or two genotypes, harboured fewer genotypes overall (median 7 [IQR 3-11] vs. 10 [IQR 8-12]), and demonstrated higher levels of phylogeographical clustering within the global tree compared to sequences from continents. Island populations had significantly fewer unique SNVs (median 2 [IQR 0-7] vs.11 [IQR 3-36], p=0.0001) and were more genetically differentiated (median Hendrick's G_{ST} 0.502 [IQR 0.466-0.820] vs. 0.106 [IQR=0.0595-0.154], $p=7.125x10^{-6}$). Island populations also displayed significantly lower pangenome diversity (median Jaccard gene content distance 0.0202 [IQR 0.00623-0.0285] vs. 0.0244 [IQR=0.0149-0.0415], $p<2.2x10^{-16}$). Further, phylogenetic analysis of common *S.* Typhi pHCM2 and IncHI1 plasmids revealed the evolution of unique plasmid variants associated with the islands of Samoa and Zanzibar. Taken together, these data suggest that the evolution of island *S.* Typhi populations and associated mobile genetic elements appears to have been driven by less frequent intercontinental transmission events and access to a more limited gene pool, consistent with allopatry.

Room 5-10

The distribution of beneficial and essential genes on plasmids

Yiqing Wang[1], Tanita Wein[1,2], Tal Dagan[1]
[1]Kiel University, Kiel, Schleswig-Holstein, Germany. [2]Weizmann Institute of Science, Rehovot, Israel

Abstract

Plasmids are autonomously replicating extrachromosomal elements that reside in prokaryotes. Plasmids are commonly reported in the literature as drivers of rapid adaption to growth-limiting conditions or a new ecological niche. Here we set out to test how often plasmids observed in *Escherichia* isolates may be assumed to have a benefit for their host. Our search includes two categories of plasmid-encoded genes: genes code for structural proteins or basic metabolic functions that are essential for the host, and genes code for beneficial functions that are advantageous to the host depending on the environmental conditions, e.g. antibiotic resistance gene(ABR gene). We found that essential genes are rarely located on plasmids, except functionally divergent genes e.g. plasmid-derived *ssb* and *groEL/S* gene are located on large plasmids. Further evolutionary experiment shows essential gene acquisition can lead to plasmid extinction due to dose-effect. Furthermore, ABR genes are rarely found in small plasmids and their distribution shows a divergence between chromosomes and plasmids. A group of ABR genes such as *sul2, blaTEM-1,* and *tetA/R* is overrepresented on plasmids. Another group of ABR genes like *oqxA/B* is rarely on plasmid but abundant on chromosomes. Our results show that small plasmids should be considered more as autonomously replicating entities. Large plasmids are more likely to encode host-related genes. Essential genes on plasmid are highly like the result of plasmid acquiring new chromosomal genes from a distant host. Such plasmids may consequently become an essential plasmid or repurpose an essential gene into a beneficial gene.

Room 5-11

Evolutionary constraints and the distribution of beneficial mutational effects in Saccharomyces vineyard adaptation

Emery Longan, Justin C Fay
University of Rochester, Rochester, New York, USA

Abstract

Evolutionary constraints can hinder adaptation by natural selection. The distribution of mutational effects (DME) can impose such constraints if available beneficial mutations are rare, of small effect, or very costly. When exposed to the same environment, related species may have different adaptive potentials due to differences in the DME. *S. cerevisiae*, unlike its sister species, *S. paradoxus,* has adapted to the anthropogenic oenological stressors copper and sulfite. To test whether adaptation to the vineyard environment can be explained by differences in the DME, we performed genome wide saturation mutagenesis on wild isolates of both species and recovered hundreds of mutants displaying increased resistance to copper and sulfite. We then subjected the mutants to a high-throughput phenotyping assay to precisely measure their mutational effect size and pleiotropic costs. By sequencing the genomes of a large subset of mutants (N=150), we identified causal variants in both species. We find evidence for gene level parallel adaptation for both stressors, however, we also find systematic differences in effect size and pleiotropic costs between species. Among the copper mutants, we also find many more aneuploid strains in *S. paradoxus* than *S. cerevisiae*. Together, these results point to effect size and pleiotropic costs being more labile components of the DME than mutational target size. These results also highlight the impact that the DME can have on the probability of parallel adaptation. Thus, whether parallel adaptation occurs in related taxa can be informative in terms of mutational constraints and the DME.

Room 5-12

Exploring population structure in testate amoebae: analyses of single-cell transcriptomic data from uncultivable lineages

Mattia Greco[1], Quinn White[1], Clara Malekshahi[1], Robin Sleith[1], Agnes Weiner[1], Laura A. Katz[1,2]
[1]Smith College, Northampton, Massachusetts, USA. [2]University of Massachusetts Amherst, Amherst, Amherst, Massachusetts, USA

Abstract

Dating back to the Neoproterozoic (~750 mya), shelled amoebae of the order Arcellinida are among the earliest eukaryotic organisms in the fossil record. Species in this clade occur in high abundance in bogs and fens, and their shells are widely used in biomonitoring as well as paleoecological analyses. Yet, due to both the difficulties in culturing these organisms and their very large/complex genomes, to date little is known about the biology of Arcellinda and fundamental questions on the process driving the diversity of this group are still unanswered. Population genetics can provide insights into the evolutionary processes shaping species diversity but its application in non-model microbial organisms is hampered by the lack of reference genomes. Furthermore, most of the assumptions underlying population genetic software (e.g., known ploidy, presence/absence of sex) are not met in protist lineages. Here we explore population structure in two Arcellinida species, Hyalnosphenia elegans, and Hyalosphenia papilio, using single-cell transcriptomes from 93 individuals sampled across New England bogs and fens, and we include a single transcript from Europe for each species. Our approach leverages both population and phylogenomic analyses, using a contamination-free reference generated with our in-house tool PhyloToL. Surprisingly we observe a very low variability and evidence of cryptic species within both H. papilio and H. elegans, with the two cells from France clustering together with open bog samples from New England. These data suggest a combination of low effective population size and high gene flow shapes the diversity within these amoebae species.

Room 5-13

The origin and evolution of cell-intrinsic antibacterial defenses in eukaryotes

Daniel J Richter[1], Tera C Levin[2]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain. [2]University of Pittsburgh, Pittsburgh, PA, USA

Abstract

To survive in a world dominated by bacteria, eukaryotes have evolved numerous self-defense strategies. While some defenses are recent evolutionary innovations, others are ancient, with roots early in eukaryotic history. Here, we look across the eukaryotic tree to identify the most ancient modes of antibacterial defense. We also highlight the diversification of pattern recognition receptors used for bacterial detection. We propose that studying antibacterial responses across diverse eukaryotes can reveal novel modes of defense, while highlighting the critical innovations that occurred early in the evolution of our own immune systems.

Room 5-14

Resource uptake and the evolution of moderately efficient enzymes

Florian Labourel, Etienne Rajon
Laboratoire de Biométrie et Biologie Evolutive, VILLEURBANNE, France

Abstract

Since Life was born, its Evolution has created an exceptional diversity of entities spanning an extravagant range of sizes from tiny microscopic molecules to the giant organisms that embody Megafauna. This broad variability, which exists both between and within classes of biological entities has often been theoretically explained by assuming the existence of biological trade-offs and specific niches (eg. two different nutrients in the environment). However, how these trade-offs build up at the cellular level has mostly remained elusive because models of specialisation overlook the very mechanistic underpinnings of cells, that is to say how they actually work. Here, we develop a model where the fitness of cells emerges from a sequence of enzyme-substrate reactions that each produce a specific metabolite, and show that accounting for constraints sheds light on the reasons why enzyme properties resemble a zoo although they seemingly evolve under a similar directional selective pressure – and should thus, at first glance, all look the same. Our results point to drift and effective population size playing an important role, along with the kinetics of nutrient transporters, the tolerance to high concentrations of intermediate metabolites, the reversibility of reactions and enzyme levels of expression. Based on these metabolic fitness landscapes, we further outline how such models should help understand how new niches may arise owing to some of these cellular constraints and how it should help predict which nutrients are more prone to trigger cross-feeding both *in vitro* and *in natura*.

Room 5-15

The brown bear (*Ursus arctos*) demography on Hokkaido Island, Japan, based on whole-genomic sequence analysis

Yu Endo[1], Naoki Osada[1], Tsutomu Mano[2], Ryuichi Masuda[1]
[1]Hokkaido University, Sapporo, Japan. [2]Institute of Environmental Sciences, Hokkaido Research Organization, Sapporo, Japan

Abstract

Previous studies of the brown bear (*Ursus arctos*) on Hokkaido Island, Japan, showed three allopatrically distributed mitochondrial lineages and gene flow between the lineages due to male-biased dispersal. In this study, we determined whole-genomic sequences for six Hokkaido brown bears and analyzed these data along with previously published genomic sequences from 17 brown bears from other parts of the world. We found that the Hokkaido population is genetically distinct from other populations, with higher genetic diversity than in the endangered populations in Western Europe. A reconstruction of historical demography using the pairwise sequential Markovian coalescent (PSMC) model showed no increase in population size for the Hokkaido population during the Eemian interglacial period (130,000–114,000 years ago). In a phylogenetic analysis of the autosomal data, the Hokkaido population emerged a clade distinct from North American and European populations, showing that it had maintained genetic diversity independently from continental populations following geographical isolation on the island. This autosomal homogeneity contrasts with the geographically separate mitochondrial lineages on Hokkaido, and indicates the occurrence of male-driven gene flow between subpopulations. In addition, *f*4 statistics and genetic structure analysis suggested that male-driven gene flow in the Hokkaido population has been affected by Isolation by distance (IBD) and geographic barriers.

Room 6-01

A diversity of selection modes has resulted from changing of target haplotypes in a gastric cancer-associated genomic region

Risa L. Iwasaki, Yoko Satta
The Graduate University for Advanced Studies (SOKENDAI), Hayama, Kanagawa, Japan

Abstract

We investigated signatures of ongoing positive selection in a gastric cancer-associated genomic region, which is classified into three subhaplotypes (composed of three SNPs): TCA, CCA, and CAG. CCA diverged from the ancestor TCA by T to C substitution around 1.0-2.4 mya. Subsequently, a new subhaplotype, CAG, diverged from CCA ~240 kya. Using 1KGP, we unveiled that distinct selection has been operating on each subhaplotype in genetically close populations (five East Asian subpopulations): we detected signatures of selection on standing variation in two subhaplotypes, CCA and CAG, in Han Chinese, while selective sweeps were detected only on CCA in Japanese, Chinese Dai, and Vietnamese. Onsets of selection on both CCA/CAG were estimated to be earlier than the split of the relevant populations. We examined whether such rapid change of selection mode occurred in Africa (AFR), South Asia (SAS), America (AMR), and Europe (EUR). While we detected no signatures of selective sweep in EUR and AMR, selection on a single subhaplotype has been detected on CCA and CAG in subpopulations of SAS and AFR, respectively. All onsets of selection were estimated to be > 8-30 kya, suggesting that independent selection occurred in each subpopulation after population splits. We conclude that the three subhaplotypes have been maintained in each population, but the selection target haplotypes are different among individual populations. Although we do not know the biological functional difference between CCA and CAG, this observation suggests that ‘target sites of selection’ change frequently within individual populations.

Room 6-02

When the gene regulatory network (GRN) is weakly but broadly perturbed – GRN stabilization by microRNAs

Guang-An Lu, Yixin Zhao, Qinjian Chen, Zhixiong Tang, Chung-I Wu
State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guang Zhou, Guang Dong, China

Abstract

Recent studies have increasingly pointed to microRNAs (miRNAs) as the agent of GRN (gene regulatory network) stabilization and, hence, developmental canalization against constant but small environmental perturbations. In this study, we construct a Dicer-1 knockdown line (*dcr*-KD) in Drosophila that shows modest reductions in all miRNAs. The hypothesis is that flies under modest miRNA reductions will gradually deviate from the developmental norm, often resulting in late-stage failures such as the shortened longevity. In the optimal culture condition, the survival to adulthood is indeed normal in the *dcr*-KD line but, importantly, the adult longevity is reduced by ~ 90%. Under suboptimal conditions of high temperature, the lethality occurs earlier in late pupation and, as the perturbations are shifted earlier, the affected stages are shifted correspondingly. We further hypothesize that the developmental failure is associated with GRN aberration even before phenotypic aberrations become observable. The analysis of the L3 transcriptomes shows *dcr*-KD to have little influence on GRN under normal culture conditions. However, under temperature perturbation, *dcr*-KD exacerbates the stress of temperature shift on the transcriptome. These experiments support the view that miRNAs stabilize the GRN. As the development progresses (or the perturbation strengthens), GRN would be increasingly dependent on miRNAs for stabilization. In conclusion, miRNAs appear to be the genome’s solution to weak but pervasive environmental perturbations.

Room 6-03

Taxonomic composition and seasonal dynamics of the air microbiome in West Siberia

Elena S. Gusareva[1,2], Nicolas P.E. Gaultier[2], Balakrishnan N.V. Premkrishnan[2], Carmon Kee[2], Serene Boon Yuean Lim[2], Cassie E. Heinle[2], Rikky W. Purbojati[2], Ang Poh Nee[2], Sachin R. Lohar[2], Vladimir N. Kharkov[3], Daniela I. Drautz-Moses[2], Vadim A. Stepanov[3], Stephan C. Schuster[2]

[1]Asian School of Environment, Nanyang Technological University, Singapore, Singapore, Singapore. [2]Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, Singapore, Singapore. [3]Institute of Medical Genetics, Tomsk National Medical Research Centre, Russian Academy of Sciences, Tomsk, Russian Federation

Abstract

A recent series of technological and analytical advancements (such as high-volumetric air samplers, an ultra-low biomass processing pipeline, low-input DNA sequencing libraries, and high-throughput sequencing technologies) have enabled comprehensive and meaningful characterization of the airborne microbial organismal dynamics found in the atmosphere. Here, we provide a metagenomic airborne community analysis of three time-series surveys of the near-surface atmosphere in two seasonal settings in West Siberia to study microbial diversity and dynamics. A total of 78 airborne biomass samples from 39 time intervals were analysed, within a temperature range of 48^{o}C (26^{o}C to -22^{o}C). We observed a 5-170-fold decrease in DNA yield extracted from the airborne biomass in winter compared to summer, nevertheless yielding sufficient material for metagenomics analysis. The airborne microbial communities included Actinobacteria and Proteobacteria, Ascomycota and Basidiomycota fungi as major components, as well as some Streptophyta plants. In summer, bacterial and fungal plant pathogens, and wood-rotting saprophytes were predominant. In winter, Ascomycota moulds and cold-related or stress environment bacterial species were enriched, while the fraction of wood-rotting and mushroom-forming Basidiomycota fungi was largely reduced. As recently reported for the tropical climate, the airborne microbial communities performed a diel cycle in summer, however, in winter diel dynamics were not marked. In conclusion, we describe taxonomical composition, as well as seasonal and diel dynamics of airborne microbial communities in West Siberia for the first time. In contrast to earlier studies of bioaerosols in this area, our approach is cultivation free and not based on nucleic acid amplification.

Room 6-04

Ancestral polymorphisms shape the adaptive radiation of Metrosideros across the Hawaiian Islands

Jae Young Choi[1], Xiaoguang Dai[2], Ornob Alam[1], Julie Z Peng[3], Priyesh Rughani[2], Scott Hickey[2], Eoghan Harrington[2], Sissel Juul[2], Julien F Ayroles[3], Michael D Purugganan[1], Elizabeth Stacy[4]
[1]New York University, New York, USA. [2]Oxford Nanopore Technologies, New York, USA. [3]Princeton University, Princeton, USA. [4]University of Nevada Las Vegas, Las Vegas, USA

Abstract

Some of the most spectacular adaptive radiations begin with founder populations on remote islands. How genetically limited founder populations give rise to the striking phenotypic and ecological diversity characteristic of adaptive radiations is a paradox of evolutionary biology. We conducted an evolutionary genomics analysis of genus *Metrosideros*, a landscape-dominant, incipient adaptive radiation of woody plants that spans a striking range of phenotypes and environments across the Hawaiian Islands. Using nanopore-sequencing, we created a chromosome-level genome assembly for *M. polymorpha* var. *incana* and analyzed whole-genome sequences of 131 individuals from 11 taxa sampled across the islands. Demographic modeling and population genomics analyses suggested that Hawaiian *Metrosideros* originated from a single colonization event and subsequently spread across the archipelago following the formation of new islands. The evolutionary history of Hawaiian *Metrosideros* shows evidence of extensive reticulation associated with significant sharing of ancestral variation between taxa and secondarily with admixture. Taking advantage of the highly contiguous genome assembly, we investigated the genomic architecture underlying the adaptive radiation and discovered that divergent selection drove the formation of differentiation outliers in paired taxa representing early stages of speciation/divergence. Evolutionary analysis of the origins of the outlier SNPs showed enrichment for ancestral variations under divergent selection, and a distribution of allele states consistent with differential sorting of incompatibility alleles. In conclusion, we propose an ancient hybridization model of origin for the Hawaiian *Metrosideros* complex, wherein reassortment of ancestral variation generated from hybridization of genetically distinct predecessors has fueled the island adaptive radiation.

Room 6-05

Genomic evidence of parasexual reproduction in parasite *Leishmania tropica*

Hedvig Glans[1,2], Hideo Imamura[3], Tim Downing[4]
[1]Karolinska University Hospital, Stockholm, Sweden. [2]Karolinska Institutet, Stockholm, Sweden.
[3]Universitair Ziekenhuis Brussel, Brussels, Belgium. [4]Dublin City University, Dublin, Ireland

Abstract

Haploid meiotic gametes have not been observed in single-celled parasite *Leishmania tropica*, resulting in questions about potential parasexual reproduction mechanisms. This tropical kinetoplastid infects humans, causing a wide range of cutaneous and often life-threatening disease pathologies. Its digenic life cycle includes meiosis in the vector sandfly and mitotic cell division in the mammalian host. Patchy recombination patterns result from this predominance of mitotic over meiotic cycles coupled with low coinfection rates of genetically distinct parasites in the sandfly. Moreover, chromosome instability during cell divisions yields aneuploidy with some chromosome copy-numbers of 3 or more. During parasexual mating, the fusion of diploid cells creates a tetraploid intermediate that undergoes chromosomal recombination and gradual chromosomal loss. Here, genome-wide SNP diversity from 22 *L. tropica* isolates showed extensive aneuploidy and chromosome-specific runs of homozygosity and heterozygosity. We found ancestries on the same chromosome (36) similar to the reference genome with virtually no derived alleles, then clusters of heterozygous SNPs, and then homozygous SNPs for the derived alleles with specific recombination breakpoints at the inferred origin of replication. Other chromosomes had similar changes at strand-switch regions separating polycistronic transcriptional units and marked changes in heterozygosity. One isolate had evidence of a chromosome 10 with a monosomic 5’ end and tetrasomic 3’ end with a breakpoint at a strand-switch region. These patterns suggest chromosome loss from an initial tetrasomic state may cause these unusual ancestry patterns. This illustrates the need for genomic surveillance of tropical parasites to detect emerging hybrids that could spread more widely.

Room 6-06

Defective satellite DNA clustering into chromocenters underlies hybrid incompatibility in *Drosophila*

Madhav Jagannathan[1], Yukiko Yamashita[2,3]
[1]ETH Zürich, Zürich, Switzerland. [2]Whitehead Institute, Cambridge, MA, USA. [3]MIT, Cambridge, MA, USA

Abstract

It has been often postulated that the rapid divergence of satellite DNA repeats between species may underlie hybrid incompatibility (HI), although the underlying cellular mechanisms have remained elusive. Recently, we demonstrated that sequence-specific satellite DNA-binding proteins cluster pericentromeric satellite DNA repeats from multiple chromosomes into nuclear foci known as chromocenters. Chromosome de-clustering due to loss of satellite DNA-binding proteins led to micronuclei formation and cell death, highlighting an important role for satellite DNA in genome encapsulation. Here we show that the sterile gonads and atrophied somatic tissues in incompatible *Drosophila melanogaster – Drosophila simulans* hybrids exhibit chromocenter disruption phenotypes such as chromosome de-clustering, micronuclei formation and cell death. Two of the previously identified HI genes in these species are *D.melanogaster Hmr* and *D.simulans Lhr*, which gain a dominant negative function in the hybrid context. Strikingly, we observed that both chromocenter formation and genome encapsulation are restored when hybrid sterility/lethality is rescued by mutating these HI genes. Moreover, ectopic expression of *D.melanogaster Hmr,* which induces lethality in the normally viable female hybrids, resulted in chromocenter disruption and micronuclei formation, indicating that these phenotypes are the direct consequence of genetic incompatibilities between these species. Finally, sterile male hybrids between the more closely related species, *Drosophila simulans* and *Drosophila mauritiana*, also exhibited chromocenter disruption, micronuclei formation and cell death, suggesting that these cellular phenotypes are a general feature of hybrid incompatibility. Therefore, we propose that rapid divergence of satellite DNA repeats between closely related species can cause reproductive isolation through chromocenter disruption.

Room 6-07

Transposable elements mediate gene duplication in metazoans

Shengjun Tan, Yong E. Zhang
Institute of Zoology, CAS, Beijing, China

Abstract

Transposable elements (TEs) are mobile genetic units ubiquitous in various organisms and now accepted as catalysts of evolution. According to their transposition mechanisms, TEs are classified into two groups: 1) retrotransposons, including LTRs and non-LTRs; and 2) DNA transposons, including TIR TEs and *Helitrons*. Studies have already shown that L1s (non-LTR TEs) and *Helitrons* are capable of mediating gene duplications in animals. Complementarily, we identified a certain number of gene duplications mediated by both LTR retrotransposons and TIR TEs in metazoans. The duplication rates are variable between species to species, which are seemingly associated with the diverse activity of TEs among taxa. Although the underling mechanisms are quite different, the duplicated copies mediated by LTRs and TIR TEs have a similar chimeric structure: the internal duplication is flanked by TEs. At the fusion points we identified shared short similar sequences, suggesting the involvement of microhomology-dependent template switches at RNA or DNA level. Considering the transposition mechanisms of LTRs and TIR TEs, we proposed two new models to explain the formation of gene duplications by these two types of TEs. Moreover, duplicates flanked by TEs could act as pseudo TEs and amplify via further transpositions. All of these features confer TEs with a strong capability of shuffling genetic materials and endorse TEs as a vibrant force in shaping gene content evolution in animals.

Room 6-08

Characterizing the accessibility of DNA binding functions in the sequence space of an ancient transcription factor

Santiago Herrera-Álvarez, Jaeda E. J. Patton, Joseph W. Thornton
University of Chicago, Chicago, Illinois, USA

Abstract

Differences in the mutational accessibility of new phenotypes can bias evolutionary outcomes, but the extent to which this has influenced historical evolution is unclear. In the steroid receptor (SR) family of chordate transcription factors, three historical substitutions caused a switch in DNA binding specificity from an ancestral estrogen response element (ERE)-binding phenotype to a derived steroid response element (SRE)-binding phenotype. ERE and SRE differ from each other by two nucleotides, resulting in a space of 4^2=16 theoretically possible response elements (REs); it is unknown why out of this space of possibilities, evolution historically produced specificity for SRE rather than for some other RE. To understand whether this was due to higher mutational accessibility of SRE specificity phenotypes from the ancestral SR genotype, we are performing intermolecular deep mutational scanning on resurrected ancestral SRs and their REs. We have engineered 16 yeast strains each containing a separate RE variant driving expression of a GFP reporter gene. Into each strain, we will transform a library of ancestral SR DNA binding domains containing all possible combinations of mutations at four amino acid sites that contact the variable RE bases. After sorting and sequencing yeast by GFP expression, we will construct genotype-phenotype maps of DNA binding for all possible combinations of SR and RE genotypes. This will allow us to quantify the mutational accessibility of new binding phenotypes in ancestral sequence space and evaluate the extent to which this could have influenced the outcomes of SR evolution.

Room 6-09

The evolutionary consequences of a morphological shift following polyploidization

Jill K Olofsson
University of Copenhagen, Copenhagen, Denmark

Abstract

Polyploidization has been prevalent during the evolution of flowering plants, both in ancient and more recent times. The effects of polyplodizations range from instant speciations to more settled shifts in phenotype but often an ecological context is lacking. Cultivated polyploids that following garden escapes have become naturalized are ideal models to build a bridge linking genetic and phenotypic findings to ecological consequences of polyploidization. The genus *Mentha* (mints; Lamieceae) is exceptional in that many species are inter-fertile, forming both sterile and fertile diploids and/or polyploids. The most famous example, is the extensively cultivated and naturalized *M. spicata* that is an allopolyploid hybrid between the two wild *M. longifolia* and *M. suaveolens*. Here we present whole-genome sequencing and morphology of over 100 individuals. We show that despite extensive genomic overlaps between *M. spicata* and *M. longifloia* there has been a morphological shift following polyploidization. Furthemore, the fertile *M. spicata* retains ability to cross with with *M. longifolia* and thus it acts as a genomic highway transporting genes from *M. suaveolens* into *M. logifolia* increasing the genetic diversity and environmental adaptability. The success of the allopolyploid *M. spicata* in the wild is most likely coupled to its increased gene-set which allows a large morphological, and likely physiological, space and hence creates possibilities to expand its ecological niche. Hybridization and polyplyploidization is therefore hallmarks of the evolutionary and biogeographic histories in the genus *Mentha* and are likely to have contribute to the success and adapatability of species following introductions to new environments.

Room 6-10

A retrotransposon presumed to have been endogenized or transposed recently in a marsupial, the red-necked wallaby

Sakura Hayashi, Akihiko Koga
Kyoto university, Kyoto, Kyoto, Japan

Abstract

Long terminal repeat (LTR) retrotransposons include endogenous retroviruses, sequences which endogenous retroviruses originate from, and derived sequences. Mammalian genomes contain several LTR retrotransposon copies resulting from past events of endogenization (infection of the germline by a retrovirus) or transposition (copy-and-paste amplification in a cell). However, the vast majority of their extant copies are no longer transposable because of natural decay or the development of transposition suppression mechanisms by the host organisms. With the eventual goal of detecting an active LTR retrotransposon, we attempted to find an element which is presumed to have been transposed or endogenized recently. Our strategy was to examine albino mutant genes for insertion sequences. Albino phenotypes generally have survival disadvantages; therefore, causative mutant genes tend to be excluded from the gene pool. Thus, when an albino gene is found, its origin is considered relatively recent. Using this strategy, we identified a novel LTR retrotransposon inserted in the *tyrosinase* gene of an albino mutant wallaby. This retrotransposon, which we named *walb*, is 7137 bp in length, carries 389-bp-long LTRs of perfect sequence matching, contains open reading frames with amino acid sequence similarities to the *gag*, *pro*, *pol*, and *env* proteins of retroviruses, and is flanked by a 6-bp-long target site duplication. This intact structure also suggests that *walb* retains transposition activity in the genome of this marsupial.

Room 6-11

Lack of genetic diversity in Hawksbill Turtles nesting in Singapore

Regine Tiong[1], Hie Lim Kim[1,2]

[1]Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, Singapore. [2]Asian School of the Environment, Nanyang Technological University, Singapore, Singapore

Abstract

Critically endangered hawksbill turtles come onto the shores of Singapore to nest, and aiding in their recovery would help to protect the coral reef ecosystems and the diversity of species we see in Singaporean waters. Maintaining genetic diversity in their population is advantageous for adapting to changing environments by acting as a safety net against extinction. To maximise conservation efforts for this hawksbill population nesting in Singapore, we aim to estimate the extent of genetic diversity and population structure of the population. We collected dead hatchlings and eggs from nests laid in 2019 and 2020 and extracted genomic DNA from the eggs, liver tissues of hatchlings, or partially developed foetuses in the eggs. We sequenced mitochondrial (mt) DNA control region of 47 samples collected from 38 nests. The sequence dataset was combined with mtDNA datasets retrieved from previous studies to uncover their genetic relationships with other nesting and foraging populations in the Indo-Pacific. Here we report, for the first time, the mt haplotypes identified from hawksbill turtles nesting in Singapore. Three different mt haplotypes were identified and have only 1 to 3 nucleotides differences each other. These haplotypes are closely related to those found in Malaysian nesting samples and show large divergence from some other haplotypes found in Indo-Pacific. This study revealed small genetic diversity of hawksbill nesting populations in Singapore, their close genetic relationship with nesting populations in neighbouring regions, as well as clear population structure within the Indo-Pacific nesting populations.

Room 6-12

ashermoshe@mail.tau.ac.il

Asher Moshe[1], Einat Hazkani-Covo[2], Omer Israeli[1], Oren Avram[1], Itsik Pe'er[3], Tal Pupko[1]

[1]Tel Aviv University, Tel Aviv, Israel. [2]Open University of Israel, Ra'anana, Israel. [3]Columbia University, New York city, New York, USA

Abstract

Genome rearrangement events play a major role in molecular evolution. The inference of genome rearrangement events has been extensively studied. However, probabilistic evolutionary models that explicitly imitate the evolutionary dynamics of such events, as well as methods to infer model parameters are yet to be fully utilized. Here, we develop a first probabilistic approach to infer genome rearrangement rate parameters using an Approximate Bayesian Computation (ABC) algorithm. We developed two genome rearrangement models, a basic model, which accounts for order changes in the genome, and a more sophisticated one which also accounts for chromosome number changes. We characterize the ABC inference accuracy using simulations. Finally, we test our models on empirical bacterial and yeast datasets.

Room 6-13

Population genomics of Australian burrowing frogs *Neobatrachus* reveal adaptation to polyploidy

Jozefien D. Van de Velde[1], Polina Yu. Novikova[1], Steve Donnellan[2]
[1]Max Planck Institute for Plant Breeding Research, Cologne, Germany. [2]South Australian Museum, Adelaide, Australia

Abstract

Polyploidy or whole-genome duplications (WGDs) are characteristic for plants, but they are recognized as important hallmarks in evolutionary history across the whole tree of life, including vertebrates. Polyploids are often associated with extreme environments, which suggests that WGDs can be either adaptive to environmental stress and/or triggered by such. Indeed stress may increase unreduced gamete formation leading to polyploidy, however, the establishment of a polyploid species is restricted by a number of factors such as minority cytotype exclusion, severe population bottleneck, and cell cycle machinery failure. Namely, polyploidy may be adaptive, but you need to adapt to being a polyploid. In order to understand such complex adaptation patterns to external and internal challenges, we study the diploid-tetraploid species complex of Australian burrowing frogs *Neobatrachus*, amphibians living in a desert. Most polyploid vertebrates reproduce asexually, however these Australian burrowing frogs form an interesting exception, with multiple independently originated autotetraploid sexual species. Chromosome-level genome assembly of *N. pictus*, resequencing of 88 individuals across the genus and genome-wide selection scans reveal the first insight into the adaptation to autopolyploidy shown in vertebrates, namely that the synaptonemal complex plays a major role in this process.

Room 6-14

New Retrogene Formation as a Potential Driver of Local Adaptation Between Island Congeners of Drosophila on the Island of São Tomé

James E. Titus-McQuillan, Rebekah L. Rogers
University of North Carolina at Charlotte, Charlotte, NC, USA

Abstract

The genetic basis of phenotypic differences between species is among the most longstanding questions in evolutionary biology. How new genes form and the processes that selection acts on said novelty to produce differences across species is fundamental to understand how species can persist in an ever-changing environment. Adaptation and genetic innovation arise in the genome by a variety of sources, of particular interest are retrogenes. Understanding the influence of retrogenes during local adaptation may illuminate species diversification. We have identified new retrogene formation on the island of São Tomé between two species of *Drosophila*, *D. yakuba* and *D. santomea*. These two *Drosophila* species both inhabit the island but occupy different range distributions by elevation. Our results show that many retrogenes show differentiation between mainland *D. yakuba* in both island lineages. We find high F_{ST} values between putative retrogenes of the mainland population of *D. yakuba* and its island population. Our analyses show, as well, that *D. santomea* and the mainland *D. yakuba* have high F_{ST} values between the species. Our preliminary research leads us to believe that newly formed retrogenes potentially have a role in local adaptation. Further research is planned to empirically test candidate retrogenes as drivers of adaptation. By understanding the mechanisms and processes that promote novel genes and how these genes interact within the genome are crucially important to understand evolutionary machinery.

Room 6-15

Order-wide classification of homeobox genes across Lepidoptera using reference quality genomes

Peter Mulhair[1], Amber Harper[1], Liam Crowley[1], Douglas H. Boyes[2,3,4], Owen T. Lewis[1], Peter W.H. Holland[1]

[1]University of Oxford, Oxford, United Kingdom. [2]UK Centre for Ecology & Hydrology, Wallingford, United Kingdom. [3]Newcastle University, Newcastle, United Kingdom. [4]Butterfly Conservation, Wareham, United Kingdom

Abstract

Lepidoptera represent one of the most diverse orders in the Animal Tree of Life with over 160,000 described species. The diversity in form and behaviour within this clade can be attributed to genetic adaptations. One group of genes that may have played a role in this diversity of phenotypes are the homeobox genes. Many homeobox genes, such as the well characterised Hox genes, encode transcription factors that are essential during early embryonic development. Changes in these genes have been found to drive the emergence of novel phenotypes in animals. Here we avail of new chromosome level genome assemblies for 53 lepidopteran species to characterise the copy number, synteny and order, and molecular evolution of homeobox genes, with a particular focus on HOX and NK genes of the Antennapedia class. We find that, as across all insects, the gene content is relatively stable across Lepidoptera. Surprisingly, we discover large tandem duplications within the Hox cluster occurring multiple times independently in 6 moth species. These large duplications occur within the special homeobox (*Shx*) genes, which are unique to Ditrysia and show higher rates of sequence evolution relative to other Hox genes. The NK genes show more consistent rates of copy number, but the structure and order of the cluster has undergone significant rearrangements within individual butterfly families. Overall, we provide the largest order-wide characterisation of homeobox genes in Lepidoptera, and show significant variation in copy number, order, and rates of evolution that may have contributed to adaptation within this group.

Room 6-16

A History of Duplications: Tracing Immune-Related Gene Family Expansions Across Molluscan Evolutionary History

Jacopo Martelossi, Aurelia Falegname, Mariangela Iannello, Fabrizio Ghiselli
University of Bologna; Department of Biological Geological and Environmental Science, Bologna, Italy

Abstract

Immuno-related gene families represent one of the most emblematic cases of fast evolving genes across the entire tree of life: from mammals, to plants, to insects. Among the genetic mechanisms increasing the diversity and specificity of the immune response, extensive and lineage-specific gene duplications and rearrangements often play a central role. These phenomena can lead to a broad set of different molecules that reflects not only the different evolutionary histories but also the ecological niches and behaviors of species. Innate immune system is getting increasing attention in molluscs and particularly in bivalves due to their great economic importance, and gene copy number variation already seems to play a key role in the evolutionary dynamics of immunity-related genes. However, most efforts are focused on a few species representing a very limited number of families and an overall picture in a comparative genomics, evolutionary, and phylogenetic framework is still missing. Here, we want to trace main evolutionary shifts in molluscan immune system gene families, with a particular focus on bivalves. For this purpose, we are combining automatic and manual annotation for more than 100 immune-related gene families in 26 molluscan genomes. To help us in this process and to investigate evolutionary changes also at the domain level, we are currently developing a set of Python scripts able to easy build up domain co-occurrence networks directly from Interproscan results.

Room 6-17

Amplicon analysis of animal-associated foraminifera across built and open environments

Rabindra Thakur[1,2], Adena Collens[2], Elinor Sterner[2], Laura A. Katz[1,2], Mattia Greco[2]
[1]University of Massachusetts, Amherst, MA, USA. [2]Smith College, Department of Biological Sciences, Northampton, MA, USA

Abstract

Evidence from early foraminifera biodiversity studies has reported that benthic species are commonly found on the surface of living organisms including giant Xenophyophores and marine invertebrate shells such as clam, snail, crab, etc. as a bycatch. The association of foraminifera with other animals could have various ecological roles such as nutrient cycling and food webs. However, few studies exist that exclusively focus on animal-associated foraminifera, and these have solely relied on the morphological description. To assess a more comprehensive animal-associated foraminifera diversity, we performed amplicon analysis of samples collected from the shells of clams, snails, and crabs from three environments (two ‘built’ environments, fish tanks) and one open salt marsh environment). We characterize both DNA/RNA communities using foraminifera-specific PCR primers. We identified 687 OTUs, many of which appear to be specific to their environments based on distance matrices comparing their abundance and evolutionary relatedness. Based on our preliminary results, several of our most abundant OTUs are sister to the morphospecies Cibicides sp., foraminifera that builds its hard shell out of calcium carbonate. By comparing our data to published literature on animal-specific lineages, we anticipate resolving the taxonomy and have identified that their communities fluctuate over time, and are specific by habitat.

Room 6-18

Speciation processes in sympatric species of land snails from genus *Trochulus* (Gastropoda, Hygromiidae) inferred from morphological and molecular variation

Małgorzata Proćków[1], Elżbieta Kuźnik-Kowalska[2], Joanna R. Pieńkowska[3], Aleksandra Żeromska[4], Paweł Mackiewicz[4]

[1]Museum of Natural History, University of Wrocław, Wrocław, Poland. [2]Department of Invertebrate Systematics and Ecology, Wrocław University of Environmental and Life Sciences, Wrocław, Poland. [3]Department of Cell Biology, Adam Mickiewicz University, Poznań, Poland. [4]Department of Bioinformatics and Genomics, University of Wrocław, Wrocław, Poland

Abstract

The identification of land snail species from the genus *Trochulus* on the basis of shell measurements and morphology is problematic due to their great variation. Genetic studies have also proved inconclusive because they revealed much variation within populations. To cope with these issues, we applied morphometric and molecular approaches to analyse the morphologically similar species *T. coelomphala*, *T. hispidus* and *T. striolatus*, which co-occur in the Alpenvorland of Germany. Although these species can be quite clearly recognised based on shell and reproductive system morphology, we found forms that appeared intermediate in shell characteristics between *T. coelomphala* and *T. hispidus* but showed genital features similar to *T. coelomphala*. However, phylogenetic analyses showed that these forms grouped neither with *T. coelomphala* nor *T. hispidus* but were sister to *T. striolatus* living in the same region. It suggests that the latter and the intermediate forms evolved by sympatric speciation. These analyses showed also that *T. coelomphala* diverged within *T. hispidus* and crossing experiments indicated that their hybrids were interfertile. The inclusion of all available *Trochulus* sequences enabled us to infer evolutionary relationships between them and showed that *T. hispidus* is polyphyletic. Some *Trochulus* samples of one nominal species were grouped within others. The combination of phenotypic plasticity and possible mitochondrial DNA introgression illustrates the complex nature of evolutionary processes and the need for caution in the application of traditional taxonomic practice. This work was supported by the National Science Centre, Poland (Narodowe Centrum Nauki, Polska) under Grant number 2016/21/B/NZ8/03022.

Room 6-19

The Cyclically Seasonal Drosophila subobscura Inversion O_7 Originated From Fragile Genomic Sites and Relocated Immunity and Metabolic Genes

Charikleia Karageorgiou, Rosa Tarrío, Francisco Rodríguez-Trelles
Universitat Autonòma de Barcelona, Barcelona, Spain

Abstract

Chromosome inversions are important contributors to standing genetic variation in *Drosophila subobscura*. Inversion O_7 has long been associated with global warming as it shows a regular seasonal cycle that peaks in summer and rose with a heatwave. However, the fact that it does not respond to thermal variation across geography, suggests that the frequencies of O_7 are not driven by temperature alone. Here we generated a PacBio long read-based chromosome-scale genome assembly, from a highly homozygous isogenic line for an O_{3+4+7} chromosome. The complete continuous sequence of O_7 was isolated by conserved synteny analysis with the available reference genome. Main findings include the following: the assembled O_7 inversion stretches 9.936 Mb, containing > 1,000 annotated genes; O_7 had a complex origin, involving multiple breaks associated with non-B DNA-forming motifs, formation of a microinversion, and ectopic repair in *trans* with the two homologous chromosomes; the O_7 breakpoints carry a pre-inversion record of fragility, including a sequence insertion, and transposition with later inverted duplication of an *Attacin* immunity gene; and the O_7 inversion relocated the major insulin signaling *foxo* gene in tight linkage with its antagonistic regulatory partner *Akt1* and disrupted concerted evolution of the two inverted *Attacin* duplicates, reattaching them to dFOXO metabolic enhancers. Our findings suggest that O_7 exerts antagonistic pleiotropic effects on reproduction and immunity, setting a framework to understand its relationship with climate change. Furthermore, they are relevant for fragility in genome rearrangement evolution and for current views on the contribution of breakage versus repair in shaping inversion-breakpoint junctions.

Room 6-20

Effects of Nucleosome Organization on Mutation Rates in Different Nucleotide Contexts

Jingyi Guan
Sun Yat-sen University, Guangzhou, China

Abstract

Germline mutations are primary sources of genetic variations and essential for evolution of biodiversity. Local mutation rates have been suggested to be affected by nucleosome organization. However, the effects of nucleosome organization on mutation rates in different nucleotide contexts remain incompletely understood. Due to the differences in data sources and methodology, previous studies reported conflicting results.

Here, we identify nucleosomes with stable translational positioning or rotational positioning in human, Arabidopsis and Drosophila based on public sequencing data using the same methodology, and then investigate effects of nucleosomes on mutations in different nucleotide contexts.

The impact of nucleosome positioning (translational or rotational) on mutation rates are different for nucleotides (A/T, C/G and C/G in CpG sites). For example, the mutation rates of C/G in CpG sites and C/G in nonCpG sites show opposite patterns. 10bp and ~200bp periodicity of mutation rates can be observed, which are compatible with periodicity of rotational positioning and translational positioning, respectively. Interestingly, mutation rate at the junction between nucleosome DNA and linker DNA changes dramatically in all surveyed species, but the underlying molecular mechanism is unclear and warrants further study. We also compared our results with others and discussed potential reasons for inconsistent conclusions.

Our work helps understand molecular determinants of mutation rate variation in the genome and has important implications for genome evolution.

Room 7-01

Investigating the role of the host-transposase fusion gene *THAP7* in vertebrate neurodevelopment and human intellectual disability

Rachel L Cosby[1], Catrina Rateb[1], Steven Gay[1], Alban Ziegler[2], Fowzan S Alkuraya[3], Todd Macfarlan[1]
[1]The Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, Maryland, USA. [2]Centre Hospitalier Universitaire d'Angers, Algiers, France. [3]KFSHRC, Riyadh, Saudi Arabia

Abstract

Gene regulatory networks are key contributors to phenotypic novelty. Recent work showed that transcription factors (TFs) and their associated networks can evolve via fusion of host domains and DNA transposases. While this process often generates lineage specific TFs, several host-transposase fusion (HTF) genes are deeply conserved. The biological function of most conserved HTFs is unknown. THAP proteins are a class of vertebrate HTFs that combine a *P*-element like DNA-binding domain (THAP) with various host derived domains. THAP TFs are associated with human disease, including heart disease, cancer, and neurological disorders, making them ideal to study the role of HTFs during vertebrate evolution. We identified missense variants in the *THAP7* gene that segregate with intellectual disability and epilepsy in three families, suggesting *THAP7* plays a role in neurodevelopment. Most variants occur in THAP7's THAP domain, suggesting its TF activity is important. To initially characterize THAP7's function, we used CRISPR to generate mouse embryonic stem cell lines where THAP7 is lost, endogenously tagged, or mutated to model the human patient variants. To identify THAP7's transcriptional network, we determined its genomic binding sites and profiled gene expression via RNA-sequencing in these cell lines. These experiments will identify putative THAP7 target genes and guide future *in-vivo* studies to test the role of THAP7 during development. Our research will reveal the function of THAP7 and illuminate how THAP TFs broadly contribute to vertebrate neurodevelopment and human intellectual disability. It will also provide guidance on why some HTFs are retained over long evolutionary timescales.

Room 7-02

A bird’s eye view of pigment patterning: genetic links between regulation of plumage and eye color in the domestic pigeon

Emily T Maclary, Bridget Phillips, Ryan Wauer, Michael D Shapiro
School of Biological Sciences, University of Utah, Salt Lake City, UT, USA

Abstract

Variation in pigment patterns within and among species of vertebrates reflects underlying changes in gene regulation and cellular function that impact reproduction and survival. Thus, identifying molecular changes that underlie variation in pigment patterning elucidates both the evolutionary origins of diversity and the etiology of pigment-related developmental disorders. The domestic pigeon, *Columba livia*, is an exceptional model for understanding the genetic architecture of pigment patterns, as selective breeding has given rise to hundreds of breeds with extensive variation. We mapped the genetic architecture of pigmentation phenotypes known collectively as piebalding in the domestic pigeon. Piebalding is characterized by patches of pigmented and non-pigmented feathers, and these plumage patterns are often breed-specific and stable across generations. Using quantitative trait locus mapping in F_2 crosses, we identified 3 loci associated with piebalding, including one locus that is associated with piebalding in multiple breeds. This shared locus harbors a candidate gene, *Edrnb2*, that is a known regulator of pigment cell migration and regeneration. While piebalding is associated with the *Ednrb2* locus in all of our crosses, genomic sequence comparisons indicate there may be regulatory heterogeneity among breeds at this locus. We also identified a genetic link between piebald plumage and a dark eye color known as “bull”, which could illuminate shared developmental pathways regulating epidermal and iris pigmentation. Identifying the genetic factors that control variation in pigment patterning in domestic pigeons will enrich our understanding of how pigment patterns are established, and how these pathways can contribute to both diversity and disease.

Room 7-03

Massively parallel discovery of human-specific substitutions that alter enhancer activity

Severin Uebbing[1], Jake Gockley[1,2], James P. Noonan[1]
[1]Yale University, New Haven, CT, USA. [2]Sage Bionetworks, Seattle, WA, USA

Abstract

Human-specific adaptations, such as the expansion of the neocortex, are in part due to modifications of gene regulatory elements. However, finding the exact genetic changes underlying human-specific biology has been challenging. Here we screened over 32,000 human-specific substitutions in putative enhancer elements for their regulatory effects. We investigated two classes of elements that show evidence of novel activity in humans: Human Accelerated Regions (HARs), highly conserved elements that show a significant excess of human-specific sequence changes; and Human Gain Enhancers (HGEs), which show evidence of increased activity during human brain development compared to other species. We used two rounds of Massively Parallel Reported Assays to 1) find HARs and HGEs that show differences in activity between humans and chimpanzees and 2) identify the exact base substitutions responsible for the observed activity change. Overall, over 30% of active candidate enhancers showed differential regulatory activity between humans and chimpanzees. Within those, we found more than 400 human-specific base substitutions that showed a significant effect on regulatory activity in our assay. We found that substitutions acted in both additive and interactive ways to modify enhancer function. These findings provide new insight into how the activity of gene regulatory elements is modified by single base substitutions and provides a set of strong candidate enhancers implicated in human evolution supported by multiple sources of evidence.

Room 7-04

Leveraging the awesome power of *Saccharomyces* to probe the evolution of genetic networks.

Taylor Wang, Abigail Keller, Maitreya J Dunham
University of Washington, Seattle, WA, USA

Abstract

Causal loci for phenotypes of interest do not exist in isolation, but rather in complex, densely connected groups termed genetic interaction networks. Given the high degree of functional conservation shared with humans, *S. cerevisiae* represents an ideal model organism for understanding the connection of genetic networks to complex traits. However, questions remain about the conservation of genetic networks among various, distantly related species. To address this, I study how functional rewiring has evolved in S. cerevisiae and one of its most distantly related *Saccharomyces* relative, *S. uvarum*. I use differential gene essentiality as my phenotype of interest, which was previously observed to occur for 12% of comparable orthologues.

My work leverages the combined power of genetic mapping and transposon mutagenesis strategies to map the molecular basis of *S. cerevisiae* species-specific essential genes. While hybrid infertility poses a barrier to QTL mapping approaches, I use meiotic null alleles of MMR factors and a helicase to decrease aneuploidy and increase viability in *S. cerevisiae* x *S. uvarum* hybrid offspring. In concert, I have modified existing transposon mutagenesis workflows in preparation for the generation of hemizygote transposition libraries for synthetic lethality screens in remarkable S. cerevisiae species specific essential *yfgΔ/yfgΔ* hybrids. Armed with these powerful tools, I have selected 39 candidate genes that are essential in S. cerevisiae, but dispensable in *S. uvarum*. In mapping the genetic network changes among 20 mya diverged *Saccharomyces* yeasts, I aim to provide new insight into the level of conservation across the Eukaryotic tree of life.

Room 7-05

Identifying regulatory mechanisms affected by hydrogen sulfide in an extremophile poeciliid fish

Kerry L McGowan[1], Enrique Jimenez Schwarzkopf[1], Sascha Duttke[2], Courtney N. Passow[3], Ryan Greenway[3], Lenin Arias-Rodriguez[4], Michael Tobler[3], Joanna L Kelley[1]

[1]Washington State University, Pullman, WA, USA. [2]University of California, Davis, CA, USA. [3]Kansas State University, Manhattan, KS, USA. [4]Universidad Juarez Autonoma de Tabasco, Villahermosa, Tabasco, Mexico

Abstract

The fish *Poecilia mexicana* (Atlantic molly) has colonized springs in southern Mexico containing hydrogen sulfide (H_2S), a toxic gas that lethally inhibits ATP production in most metazoans. In sulfidic populations of *P. mexicana*, there are convergent gene expression changes across drainages related to H_2S detoxification and aerobic respiration compared to non-sulfidic populations. However, the regulatory mechanisms that underlie these changes are unknown. We hypothesized that the expression of transcription factor genes and their targets that affect the sulfidic phenotype would vary in *P. mexicana* gill tissues in the presence/absence of H_2S. We used network analyses to identify genes that correlated with H_2S and a literature search to classify those that affected H_2S detoxification. We also hypothesized that newly transcribed RNAs captured using capped small RNA-sequencing would reveal expression differences between sulfidic and non-sulfidic populations using adult gill tissues. We found that significantly different transcription initiation clustered by ecotype (sulfidic versus non-sulfidic), indicating that the presence/absence of H_2S had the largest impact on the location of RNA polymerase II initiation sites and the level of initiation. In sulfidic fish, promoter regions upstream from active transcription start sites were enriched for motifs that bind transcription factors known to inhibit the production of endogenous H_2S. Finally, we are using whole genome data and simulations to identify promoter regions under selection upstream from H_2S-related genes using a null model that accounts for demography and background selection.

Room 7-06

Co-evolution of a long non-coding RNA and its protein partners across placental mammals with different implantation strategies

Ioannis Tsagakis[1,2], Adrian Whitehouse[1], Mary J. O'Connell[3], Niamh Forde[2], Julie L. Aspden[1]
[1]School of Cellular and Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, United Kingdom. [2]Discovery and Translational Sciences Department, Leeds Institute of Cardiovascular and Metabolic Medicine, Faculty of Medicine and Health Sciences, University of Leeds, Leeds, United Kingdom. [3]School of Life Sciences, Faculty of Medicine & Health Sciences, University of Nottingham, Nottingham, United Kingdom

Abstract

Long non-coding RNAs (lncRNAs) are transcripts >200 nucleotides that do not encode proteins, but can interact with proteins and regulate gene expression. Whilst full-length sequence conservation is rare for lncRNAs, short regions of higher conservation can exist across species.

The *XIST* lncRNA mediates dosage compensation via X chromosome inactivation (XCI) of a single X chromosome in females. Maintaining gene dosage across sexes is vital for placental mammal prenatal development, failure of which is embryonic lethal. Despite its presence throughout placental mammals, most studies have focused on mouse *Xist*. It is yet to be determined whether mouse Xist-protein interactions are shared across placental mammals where the timing and nature of XCI differ substantially. Here, we aimed to dissect *XIST*'s interactors in placental mammals with different implantation strategies.

Spen, Hnrnpk, Ciz1, Rbm15 and Wtap proteins were previously identified as mouse *Xist* functional interactors. Their average amino acid identity is >70% across human, mouse, cow and pig. RT-qPCR and western blotting revealed coordinate expression of *XIST* and putative protein partners in endometrial tissues/cells of those species. RNA immunoprecipitations showed SPEN, hnRNPK and CIZ1 proteins bind human *XIST* whereas due to antibody limitations we could only observe CIZ1 binding to cow *XIST*. Differential binding of candidate proteins to *XIST* could not be attributed to positive selection based on selective pressure variation analyses using codon-based models of evolution. Overall, we find a shared *XIST*-CIZ1 interaction across human and cow and plan to orthogonally validate it using *in vitro* transcription pulldowns in human and cow.

Room 7-07

Temperature-dependent small RNA expression depends on wild genetic backgrounds of *Caenorhabditis briggsae*

Daniel D. Fusca, Julie M. Claycomb, Asher D. Cutter
University of Toronto, Toronto, Ontario, Canada

Abstract

Geographically distinct populations can adapt to the temperature conditions of their local environment, leading to temperature-dependent fitness differences between populations. *Caenorhabditis briggsae* nematodes that inhabit different geographic latitudes are genetically distinct and distinct in fitness responses to temperature consistent with local adaptation. The genetic mechanisms underlying local adaptation, however, remain unresolved. To investigate the potential role of small noncoding RNAs in genotype-specific responses to temperature, we quantified small RNA expression using high-throughput sequencing of *C. briggsae* worms from tropical and temperate strain genotypes reared under three temperature conditions. Strains representing both tropical and temperate regions showed significantly lower expression of PIWI-interacting RNAs (piRNAs) at high temperatures, primarily mapping to a large piRNA cluster ~7 Mb long on chromosome IV. We also documented depressed expression of 22G-RNAs antisense to protein-coding genes at high rearing temperatures for the thermally-intolerant temperate strain genotype, but not for the tropical strain. Reduced 22G-RNA expression was widespread along the lengths of multiple chromosomes, indicative of a genome-wide response. Targets of the EGO-1/CSR-1 22G-RNA pathway were most strongly impacted compared to other 22G-RNA pathways, implicating the CSR-1 Argonaute and its RNA-dependent RNA polymerase EGO-1 in the wild genotype-dependent modulation of *C. briggsae* 22G-RNAs under chronic thermal stress.

Room 7-08

Single-cell co-expression analysis reveals commonality and specificity of fly brain cells

Ming Yang[1], Benjamin Harrison[1], Daniel Promislow[1,2]

[1]Department of Lab Medicine and Pathology, University of Washington, Seattle, Washington, USA.

[2]Department of Biology, University of Washington, Seattle, Washington, USA

Abstract

A complex organ is characterized by diverse cells, each with distinct functions. Single-cell RNA-sequencing measures gene expression for individual cells, enabling researchers to discover the specific genes that contribute to diverse cellular functions that have evolved within and between tissues. Current analyses focus primarily on identifying differentially expressed genes across cells. However, patterns of co-expression between genes are probably more indicative of biological processes. Here, we constructed cell type-specific gene co-expression networks using the fly brain transcriptome atlas to identify the commonality and specificity of gene co-expression patterns across various cell types. We found that co-expressed gene pairs are highly cell type-dependent and the co-expression network similarities between cell types echo neuroanatomical regions. We identified a set of genes whose expression is highly coordinated across multiple cell types, and are mainly involved in macromolecular machinery and neuron-specific biological processes, suggesting these cellular functions are fundamental to all brain cells. To examine cell type heterogeneity, we designed a pathway coherence index to quantify cell type-specific pathway activities. The result showed both broadly coherent pathways and highly cell type-specific ones. In particular, the spliceosome pathway is significantly coherent in multiple cell types. However, unlike the conserved sub-networks described earlier, the spliceosome appears to have distinct sets of covarying genes in each cell type, suggesting that spliceosomal composition and regulation may drive neuronal diversity. Together, our study demonstrates that studying single cell data through a network approach can provide novel insights into understanding the evolution of cell-specific functional diversity in a complex organ.

Room 7-09

Pooling-based phylogenetic analysis elucidates transcription factor binding sites under weakly accelerated evolution in great apes

Xinru Zhang, Yifei Huang
Pennsylvania State University, University Park, PA, USA

Abstract

Many non-coding genomic elements show elevated substitution rates in the human lineage compared to other species. While numerous phylogenetic methods have been developed to identify these accelerated regions, there is currently no method designed to explore weakly accelerated evolution driven by weak positive selection or relaxed purifying selection. To address this problem, we introduce two novel phylogenetic methods to pool data from thousands of genomic elements with similar functions. One is a group-level likelihood ratio test (LRT) for concatenated genomic elements, while the other combines p-values from element-level LRTs. With a larger sample size per test, our new approaches have a significantly higher power to detect weakly accelerated evolution than previous methods.

Using the pooling-based approaches, we conduct a systematic investigation of transcription factor binding sites (TFBSs) under accelerated evolution in primates. Different from previous methods focusing on individual elements, our pooling-based analysis shows that the evolution of TFBSs for 7 transcription factors is accelerated at the group level in great apes. Compared with previously reported acceleration of conserved elements, accelerated evolution of TFBSs is statistically significant but has a smaller effect size. Furthermore, using a phylogenetic-based mixture model, we estimate that there are more accelerated TFBSs than previously reported accelerated regions. Finally, we show that both positive selection and relaxed purifying selection may drive accelerated evolution of TFBSs. Taken together, our analysis shows widespread, weak acceleration of TFBS evolution in great apes, which cumulatively may have a substantial contribution to the evolution of gene regulation.

Room 7-10

Sexual antagonism, temporally fluctuating selection, and variable dominance shape allele frequency dynamics of a regulatory polymorphism in *Drosophila melanogaster*

Amanda Glaser-Schmitt[1], Meike J Wittmann[2], Timothy J S Ramnarine[1], John Parsch[1]
[1]Ludwig-Maximilians-University Munich, Planegg-Martinsried, Germany. [2]Bielefeld University, Bielefeld, Germany

Abstract

Understanding how genetic variation is maintained within species is a major goal of evolutionary genetics that can shed light on the preservation of biodiversity. Here, we examined the maintenance of a regulatory single-nucleotide polymorphism (SNP) of the X-linked *Drosophila melanogaster* gene *fezzik*. The derived variant at this site is at intermediate frequency in many worldwide populations, but absent in populations from the ancestral species range in sub-Saharan Africa. We collected and genotyped wild-caught individuals from a single European population biannually over a period of five years, which revealed a general difference in allele frequency between the sexes and a consistent change in allele frequency across seasons in females but not in males. Modelling based on the observed allele and genotype frequencies suggested that both sexually antagonistic and temporally fluctuating selection help to maintain variation at this site. The derived variant is predicted to be female-beneficial, and may be mostly recessive; however, based on uncertainty surrounding our dominance estimates, we cannot rule out other forms of dominance. By examining gene expression and size-related phenotypes, we found that phenotypic dominance was variable and dependent upon developmental stage, genetic background, and trait considered, suggesting variable dominance is important in shaping this SNPs dynamics. We further determined that *fezzik* expression and genotype are associated with starvation resistance in a sex-dependent manner, suggesting a potential phenotypic target of selection.
By characterizing the mechanisms of selection acting on this regulatory SNP, our results improve our understanding of how selection maintains genetic and phenotypic variation in natural populations.

Room 7-11

The impact of gene network topology on the evolution of gene-specific expression noise

Nataša Puzović, Julien Dutheil
Max Planck Institute for Evolutionary Biology, Plön, Schleswig-Holstein, Germany

Abstract

Expression noise is the variability of the amount of gene product among isogenic cells grown in identical conditions. Expression noise originates from the inherent stochasticity of diffusion and binding of the molecular players involved in transcription and translation. It was shown that expression noise is an evolvable trait and that central genes in gene networks exhibit less noise. A possible explanation for this pattern is an increased selection pressure on central genes to reduce intrinsic expression noise since they propagate expression noise downstream. We test this hypothesis by simulating the evolution of gene-specific expression noise in a population of model gene regulatory networks under selective and non-selective conditions. Stabilizing selection was imposed on the expression level of all genes in the network and rounds of mutation, selection, replication and recombination were performed for 10k generations. A differential selective pressure is observed among genes of different centrality. Namely, the reduction of gene-specific expression noise as a response to stabilizing selection is higher in genes with a higher centrality metrics than genes with lower centrality metrics. In conclusion, the position of the gene in the gene regulatory network can shape its expression noise level, and the effect of network characteristics on the evolvability of gene-specific expression noise is currently being explored.

Room 7-12

Differences in Drosophila eye size are associated with temporal variation in orthodenticle expression.

Montserrat Torres-Oliva[1,2], Elisa Buchberger[1], Alexandra D. Buffry[3], Maike Kittelmann[3], Lauren Sumner-Rooney[4], Pedro Gaspar[3,5], Georg Bullinger[1], Genoveva Guerrero[6], Fernando Casares[6], Saad Arif[3], Nico Posnien[1], Maria D. S. Nunes[3], Alistair P. McGregor[3], Isabel Almudi[6,3]
[1]University Goettingen, Goettingen, Germany. [2]Christian-Albrechts-University of Kiel, Kiel, Germany. [3]Oxford Brookes University, Oxford, United Kingdom. [4]Oxford University Museum of Natural History, Oxford, United Kingdom. [5]MRC London Institute of Medical Sciences, London, United Kingdom. [6]CABD, Seville, Spain

Abstract

The compound eyes of insects exhibit extensive variation in ommatidia number and size, which affects how they see and adapt to different environments and lifestyles. However, very little is known about the genetic and developmental bases underlying differences in compound eye size. We found that Drosophila mauritiana has larger eyes compared to D. simulans due to differences in ommatidia size. We identified an X-linked chromosomal region in D. mauritiana that results in larger eyes when introgressed into D. simulans. Using a combination of fine-scale mapping and gene expression analysis we further investigated positional candidate genes and found that orthodenticle (otd) is expressed earlier in D. mauritiana than in D. simulans during ommatidial maturation. We confirmed that this gene is required for the correct organisation and size of ommatidia in D. melanogaster. Using ATAC-seq, we have identified several candidate eye enhancers of otd as well as potential direct targets of this transcription factor that are differentially expressed between D. mauritiana and D. simulans. Taken together, our results suggest that differential timing of otd expression contributes to natural variation in ommatidia size between D. mauritiana and D. simulans, which provides new insights into the mechanisms underlying the regulation and evolution of compound eye size in insects.

Room 7-13

Selection and thermostability at G-quadruplexes—novel regulatory elements of the genome

Wilfried Guiblet[1,2], Michael DeGiorgio[3], Xiaoheng Cheng[1], Francesca Chiaromonte[1], Kristin Eckert[1], Yi-Fei Huang[1], Kateryna Makova[1]

[1]Penn State University, University Park, PA, USA. [2]NCI, Bethesda, MD, USA. [3]Florida Atlantic University, Boca Raton, FL, USA

Abstract

Approximately 1% of the human genome has the ability to fold into G-quadruplexes (G4s)—non-canonical strand-specific DNA structures forming at G-rich motifs. G4s regulate several key cellular processes (e.g., transcription), and have been hypothesized to participate in others (e.g., firing of replication origins). Moreover, G4s differ in their thermostability and this may affect their function. Yet, G4s may also hinder replication, transcription, and translation, and increase genome instability and mutation rates. Therefore, depending on their genomic location, thermostability, and functionality, G4 loci might evolve under different selective pressures, which has never been investigated. Here we conducted the first genome-wide analysis of G4 distribution, thermostability, and selection. We found an overrepresentation, high thermostability, and purifying selection for G4s within genic elements in which they are expected to be functional—promoters, CpG islands, and 5’ and 3’ UTRs. A similar pattern was observed for G4s within replication origins, enhancers, eQTLs, and TAD boundary regions, strongly suggesting their functionality. In contrast, G4s on the non-transcribed strand of exons were underrepresented, unstable, and evolved neutrally. In general, G4s on the non-transcribed strand of genic elements had lower density and were less stable than those on the transcribed strand, suggesting that the former are avoided at the RNA level. Across the genome, purifying selection was stronger at stable G4s. Our results suggest that purifying selection preserves the sequences of functional G4s, whereas non-functional G4s are too costly to be tolerated in the genome. Thus, G4s are emerging as fundamental, functional genomic elements.

Room 7-14

Testing the genomic shock hypothesis using transposable element expression in yeast hybrids

Marika Drouin, Mathieu Hénault, Johan Hallin, Christian R. Landry
Université Laval, Québec, Québec, Canada

Abstract

Transposable elements (TEs) insertion mutations are the source of structural variants and can cause genetic instability and gene expression changes. A host genome can limit the spread of TEs with various repression mechanisms. However, the ability to regulate TEs propagation could be disrupted in hybrid genomes. Indeed, the combination of two divergent genomes following interspecific hybridization brings together two transposable elements (TEs) populations. According to the genomic shock hypothesis proposed by McClintock in 1984, stress and regulatory interference triggered by hybridization could lead to TE proliferation. Recent studies in yeast found no transposition rate increase in hybrids but have not ruled out the possibility that TEs are still derepressed transcriptionally following hybridization. Thus, whether total expression of a TE family is higher in hybrids than in their parental species remains to be examined. We leveraged publically available RNA-seq data on yeast hybrids of the Saccharomyces genus and performed differential expression analysis of their LTR retrotransposons (Tys). Our analyses show that Ty elements are generally not differentially expressed in hybrids, even when the hybrids are exposed to a stressful condition as low temperature. Overall, our results do not support the hypothesis that hybridization could act as a systematic trigger of TE expression in yeast. The particularities of the post-translational copy number control repression mechanism in yeast could explain the absence of transcriptional genomic shock.

Room 7-15

Trans regulatory evolution of a young microRNA during *Drosophila* spermatogenesis

Yumei Huang, Tian Tang
State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou, Guangdong, China

Abstract

Cis and *trans* regulatory evolution underlies phenotypic innovation and diversification. Relatively less is understood about regulatory evolution mediated by adaptive *trans* regulatory divergence. Here, we compare the testis transcriptomes at single-cyst level between knockout lines of miRNA-983, an adaptively evolving young miRNA, in *D. melanogaster* and *D. simulans*, and the replacement line of dme-miR-983 with dsi-miR-983 in *D. melanogaster*. Overall, expression divergence of about two fifths of expressed genes in testis is affected by miR-983-mediated regulatory changes, predominately by effects of transcriptomic context only (21.7%), followed by effects of miR-983 ortholog only (10.1%) and effects of both context and miR-983 ortholog (9.3%). MiR-983 orthologs exert stronger regulatory effects in the conspecific transcriptomic context, particularly in meiotic and post-meiotic cells during spermatogenesis. While the majority of genes with both context and miR-983 ortholog effects are subjected to compensating interactions and stabilizing selection, genes with enhancing interactions strikingly increase at meiotic and post-meiotic phases under directional selection. Moreover, knocking out dme-miR-983 in *D. melanogaster* affects sperm length and sperm competitive ability, which can partially be rescued by dsi-miR-983. Taken together, this work reveals natural selection acts on *trans*-acting elements and results in gene expression divergence during spermatogenesis.

Room 7-16

Dissecting the evolutionary and functional architecture of human enhancer sequences with multiple ancestral origins

Sarah L Fong[1], John A Capra[2]
[1]Vanderbilt University, Nashville, TN, USA. [2]University of California, San Francisco, San Francisco, CA, USA

Abstract

Motivation: Functional divergence of cis-regulatory enhancer sequences is a major driver of vertebrate speciation. However, the relationship between the evolutionary histories of enhancer sequences and their functional constraint is unclear. To address this question, we traced the evolutionary origins of transcribed human enhancer sequences active across diverse cellular contexts.

Results: While most human transcribed enhancer DNA can be traced to a single evolutionary origin, we estimate that 40% of enhancers are composed of DNA from multiple ancestral origins (i.e. of multiple ages). These “multi-origin” enhancers have evolutionary architectures consisting of older “core” sequences flanked by younger “derived” sequences. Within multi-origin enhancers, we find that both the derived and core sequences show evidence of independent biochemical enhancer activity. However, specific transcription factors (TFs) have stable preferences for binding core and derived regions that span sequence origins. Despite the evidence for activity and TF binding in both core and derived sequences, derived regions are under lower purifying selection pressures than adjacent cores. As a result, derived regions tolerate more common genetic variation and are enriched for eQTL associated with gene expression variability in human populations.

Conclusions: We propose that the integration of younger, derived sequences with conserved core sequences generates regulatory substrates with robust enhancer activity across both core and derived regions and the potential for functional variation enriched in younger derived regions. Our analyses demonstrate that considering enhancer evolutionary architectures can aid interpretation of evolutionary forces acting on enhancer sequences and functional variation across human populations.

Room 7-17

Comparative transcriptomics reveal candidate genes for adaptation to Antarctic environment in the Emperor penguin

Federica Pirri[1], Samuele Greco[2], Marco Gerdol[2], Alberto Pallavicini[2], Lorenzo Zane[1], Marine Benoiste[3], Clément Cornec[4], Célin Le Bohec[5], Emiliano Trucchi[6]

[1]Department of Biology, University of Padova, Padova, Italy. [2]Department of Life Sciences, University of Trieste, Trieste, Italy. [3]Institut Pluridisciplinaire Hubert Curien, Strasbourg, France. [4]CNRS, Strasbourg, France. [5]University of Strasbourg, Strasbourg, France. [6]Department of Life and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy

Abstract

The Emperor penguin (*Aptenodytes forsteri*) is the only warm-blooded vertebrate, thriving and breeding in the harshest Antarctic winter conditions. Its closest relative, the King penguin (*A. patagonicus*), breeds exclusively in year-round ice-free sub-Antarctic islands lacking similar extreme cold adaptations.

Their clear ecological divergence contrasted by a recent phylogenetic divergence, makes this system an ideal model to investigate the role of changes in gene expression for adaptation to different environmental conditions.

Unique transcriptomic data from natural populations of each species, from 5 different tissues (*i.e.,* brain, liver, kidney, skin and muscle) of 20 fasting chicks, were employed *i)* to assemble *de novo* the reference transcriptomes of each species using total mRNA sequencing and *ii)* to compare the transcriptional profiles across tissues and between species using 3’-end RNA sequencing.

The transcriptomes of the Emperor and the King penguin contain 106,060 and 80,605 transcripts, respectively, encompassing more than 84% of genes in the Aves orthologs database.

Differential expression analyses identified tissue and species-specific expression patterns, revealing both the functional characterization of each tissue and the gene expression changes between the two species.

Our comparative transcriptomic study, the first realized between these sister species, revealed a number of differentially expressed genes which are the candidates underlying the relevant adaptations to the Antarctic lifestyle in Emperor penguins as *e.g.*, response to cold, UV radiation resistance, adipogenesis, response to vitamine D, fatty-acid metabolism, stress response, regulation of circadian rhythm and protection against oxidative stress.

Room 7-18

Highly Dynamic Gene Family Evolution Suggests Changing Roles for PON Genes Within Metazoa

Sarah A Lucas, Allie M Graham, Nathan L Clark
University of Utah, Salt Lake City, Utah, USA

Abstract

One way species adapt to changes in selective pressures is by changing gene family size. This includes gene duplication to increase dosage or diversify enzymatic substrates and gene deletion due to relaxed selection. We recently found that the Paraoxonase 1 (PON1) gene was lost repeatedly in different aquatic mammal lineages. Using 25 genomes across metazoans, we explored the evolutionary history of the PON family to determine how dynamic this gene family is. Previous findings showed fish, birds, and amphibians have one ancestral PON while therian mammals have three. Ultimately, we determined the mammalian expansion occurred in the ancestor of all mammals after the divergence of sauropsida but before the divergence of monotremes from therians. We also found that there have been multiple independent expansion/contraction events throughout metazoan history. The maintenance of these expanded copies in mammals, bivalves, and echinoderms is a sign of gene family diversification in these lineages. The loss of two PON genes in monotremes and PON1 in marine mammals suggest that either the selective pressure to retain them has decreased or its loss provided a selective advantage. The complete lack of PON in ctenophores, crustaceans, and insects indicates this gene is not required for their survival. In the face of repeated expansions and deletions in the context of changing environments, we suggest this promiscuous enzyme’s gene family is likely responding to a range of selective pressures including pathogen infection and mitigation of oxidative damage based on their substrates.

Room 7-19

Emergence of regulatory incompatibilities during early speciation of Daphnia pulex and Daphnia pulicaria

Trung V Huynh, Sen Xu
University of Texas at Arlington, Arlington, Texas, USA

Abstract

Regulatory incompatibilities are predicted to be a mechanism for the establishment of post-zygotic barriers such as hybrid inviability and sterility. While the Bateson-Dobzhansky-Muller model provides a theoretical explanation for the emergence of such regulatory incompatibilities through the accumulation of adaptive or neutral substitutions that are deleterious on a hybrid genetic background, empirical data of how such regulatory incompatibilities emerge during early speciation is still scarce. Here, we investigate regulatory divergence between *Daphnia pulex* and *Daphnia pulicaria* and hybrid misexpression in their F_1 hybrid to understand the spectrum of regulatory incompatibilities during early speciation. The *D. pulex* and *D. pulicaria* system is believed to be early in speciation, and the main reproductive barriers are believed to be ecological pre-zygotic barriers. Since genetic post-zygotic barriers have yet to be established in this system, we were able to freely generate hybrids between several clones of both species and measure the prevalence of hybrid misexpression and the effects of *cis*- and *trans*- regulatory divergence. Furthermore, functional analysis revealed whether the regulatory incompatibilities were enriched in areas of adaptive evolution since *D. pulex* are known to inhabit ephemeral ponds which have different selection pressures than the lakes that *D. pulicaria* inhabit.

Room 7-20

Using a genetic algorithm to simulate gene regulatory network evolution and explore how the role of a gene affects the evolution of its sequence

Carlos J Rivera-Rivera[1], Djordje Grbic[2], Afa N Juste[3], Davide Pisani[4]

[1]University of Bristol, Bristol, BR, United Kingdom. [2]IT University of Copenhagen, Copenhagen, Capital, Denmark. [3]University of Puerto Rico, San Juan, PR, Puerto Rico. [4]University of Bristol, Bristol, Bristol, United Kingdom

Abstract

Genomes are subject to highly complex selective forces that act at vastly different spatial and time scales. While molecular phylogenetics focuses on the traces these selective forces leave on the genomic sequence, little attention has been given to the marks that gene regulatory networks (GRNs) leave on the genomic sequence. GRNs are sets of regulatory genes that coordinate the function and development of most organisms by interacting with each other and with regions in the genome. Here we explore how GRN architecture influences gene sequence evolution, by creating a simulator of genome sequence evolution that takes into account selective forces that may result from the role of a gene in its GRN. This simulator emulates a genetic algorithm (GA) that mutates sequences, and these mutations are reflected onto a regulatory matrix that defines how each gene interacts with itself and all other genes. The fitness of each individual is calculated by recursively multiplying a seed vector with the regulatory matrix, simulating progressive discrete time-steps of gene regulation. Here we show our results for a test run in which four populations of 100 individuals were simulated, and evolved over 20,000 generations, and the average gene expression profiles show that the architecture of the GRN may be reflected in the gene phylogenies.

Room 8-01

The genetic architecture of a female mimicry trait in swordtail fish

Tristram O Dodge, Molly Schumer
Stanford University, Stanford, CA, USA

Abstract

Alternative reproductive strategies are common across the tree of life and often involve female mimicry. Despite their importance as an evolutionary strategy, little is known about the genetic basis of these phenotypes. In swordtail (*Xiphophorus*) fish, the false gravid spot (FGS) is a female mimicry trait that evolved over ~3 million years ago and segregates at intermediate frequency within multiple species. Using a genome-wide association study, we show that the FGS phenotype in *X. birchmanni* is almost entirely explained by variation in a 50 kb region upstream of *Kit ligand* (*Kitlg*), a well-described pigmentation gene. By quantifying allele specific expression in hybrids, we find *cis*-acting regulatory changes are responsible for a 3-fold increase in *Kitlg* expression in FGS individuals. This regulatory region is structurally complex, containing deletions, inversions, and repetitive sequences. Phylogenetic evidence suggests that this *Kitlg* regulatory region has also introgressed between species, potentially explaining the FGS's unusual distribution. The FGS is polymorphic in nearly all *Xiphophorus* species in which it is found, and simulations implicate balancing selection as a likely cause of its maintenance. Disentangling the evolutionary tradeoffs that result in the maintenance of this female mimicry trait is an exciting direction for future work.

Room 8-02

Gene regulatory divergence in post zygotic reproductive isolation

Athmaja Viswanath, Asher D Cutter
University of Toronto, Toronto, Ontario, Canada

Abstract

The world’s astounding biodiversity results from speciation, the process of formation of new species. Diverging populations accumulate independent mutations leading to reproductive isolation over time. The hybrids of these populations often have reduced fitness due to negative epistatic interactions between diverged genes. Such negative interactions can manifest as misexpression of genes due to divergence in gene regulatory mechanisms caused by mutations in *cis*-regulatory elements and *trans*-acting factors that ultimately lead to incompatibility, in the form of hybrid sterility and inviability. I aim to understand the contributions of regulatory divergence to reproductive isolation using the interspecies hybrids of two pairs of *Caenorhabditis* nematodes: *C. briggsae-C. nigoni* & *C. remanei-C. latens*. For the hybrids of *C. briggsae-C. nigoni*, I analysed publicly-available transcriptome profiles of two introgression lines of sterile hybrid males, both demonstrating similar defects in spermatogenesis induced by non-overlapping fragments from the *C. briggsae* X-chromosome in a *C. nigoni* genomic background (Li et al. 2016). My analysis identified genes showing distinct classes of non-additive inheritance and regulatory divergence. I am extending this approach to hybrids of *C. remanei-C. latens* by measuring tissue-specific gene expression data for both sexes and cross directions to compare allele-specific expression. My preliminary results using hybrids of *C. briggsae-C. nigoni* suggest that non-overlapping introgressions affect many of the same male reproductive genes in the same way, causing a preponderance of underdominant expression due to compensatory and combined *cis-trans* divergence. The similar transcriptomic effects of non-overlapping regions on X-chromosome implicates multiway incompatibilities drive hybrid male sterility in this system, comprised of interactions between male reproductive genes.

Room 8-03

Rapid diversification and duplication of protamine genes in *Drosophila*

Ching-Ho Chang[1], Harmit S Malik[1,2]

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Many animals have independently acquired and deployed short, positively-charged proteins, called protamines, to tightly package sperm genomic DNA. Unlike highly-conserved histones, protamines evolve rapidly and show recurrent duplications in various lineages. However, the causes and consequences of this diversification remain poorly understood. The prevailing hypothesis posits that sexual selection, particularly sperm competition, drives protamine diversification. However, this model lacks experimental support and cannot explain the rapid evolution of protamines in species lacking sperm competition, including humans. We phylogenetically cataloged diversity and innovation of protamine-like sperm nuclear basic proteins (or SNBPs) across *Drosophila* species to understand the biological forces driving protamine evolution. Protamines in *Drosophila* are not homologous to mammalian protamines; they independently arose from proteins containing a high mobility group (HMG) DNA-binding domain commonly found in transcription factors. We discovered that protamine genes independently duplicated to and amplified on sex chromosomes seven times across *Drosophila* species. These sex chromosome-linked protamine duplicates might involve genetic conflicts between sex chromosomes by either inducing or suppressing a meiotic drive. We found that 11 of 13 testis-specific genes with HMG domains have higher protein evolution rates (ω) than 95% of *Drosophila* genes. Using polarized McDonald–Kreitman tests, we showed that 4 of them have evolved under positive selection in the *D. melanogaster* lineage. Several protamine genes essential for fertility in *D. melanogaster* have been lost in other species. Conversely, two recently evolved protamine genes are critical for fertility in *D. melanogaster*. Our analyses demonstrate recurrent selection and evolutionary turnover of protamine genes in *Drosophila*.

Room 8-04

The effects of introgression across thousands of quantitative traits revealed by gene expression in wild tomatoes

Mark S Hibbins, Matthew W Hahn
Indiana University, Bloomington, IN, USA

Abstract

It is now understood that introgression can serve as powerful evolutionary force, providing a source of genetic variation that can shape the course of trait evolution. It also introduces shared evolutionary history that is not captured by the species phylogeny, potentially complicating evolutionary inferences on traits. Such inferences are often carried out on gene expression data, where the measurement of thousands of trait values allows for powerful generalized inferences while controlling for shared phylogeny. Here we present a Brownian motion model for quantitative trait evolution under the multispecies network coalescent framework, which shows that introgression can generate apparently convergent patterns of evolution when averaged across thousands of quantitative traits. We test these predictions by leveraging whole-transcriptome expression data from ovules in the wild tomato genus *Solanum*. Examining two rooted triplets that both have evidence for post-speciation introgression but that differ substantially in its magnitude, we find that patterns of ovule expression evolution are consistent with inferred histories of introgression in both sign and magnitude. Additionally, in the triplet with a higher rate of introgression, we observe a correlation between local gene tree topology and expression similarity, implicating a role for introgressed cis-regulatory variation in generating these broad-scale patterns. Our results have important implications for the study of gene expression evolution —and quantitative trait evolution more broadly—in systems with evidence of introgression.

Room 8-05

Frequent co-domestication of PIF-like transposable element proteins in insects

Dragomira N. Markova[1], Fatema Begum Ruma[1], Claudio Casola[2], Esther Betran[1]
[1]Department of Biology, University of Texas at Arlington, Arlington, TX, USA. [2]Texas A&M University, College Station, TX, USA

Abstract

Transposable elements (TEs) are genetic units that move and amplify within a host genome. An increasing number of studies have shown that one of the most direct contributions of TEs to their host is through the process of 'molecular domestication' whereby the genes normally encoded by and serving the replication of a TE are co-opted by the host genome to create new gene(s) with cellular function. We have been studying four domesticated transposases from the *PIF/Harbinger* DNA family of TEs in *Drosophila melanogaster*, named ***D****rosophila* ***P****IF* ***L****ike* ***G****enes* (*DPLGs*). All *PIF* TEs known in plants and animals have two independent transcription units: one encodes a protein representing the catalytic transposase, while the other encodes a MADF domain. We hypothesize that there should often be co-domestications of transposase and MADF proteins because the transposase translocates to the nucleus by the MADF protein. This is true for *HARBI1* and NAIF1 in humans, *DPLG7* and *DPM7* in *Drosophila* and two co-domestication in *Arabidopsis*. To provide further support to this model, we investigated numerous insect genomes for additional evidence of *PIF* TE domestication events and explore the co-domestication of the MADF protein form the same TE insertion. We present evidence of four protein domestication events in insects: a co-domestication of both transposase and MADF in *Anopheles* (Diptera) and *Nasonia* (Hymenoptera), as well as one transposase only domestication event in butterflies and moths (Lepidoptera), and in cockroaches (Blattodeae). Thus, our results show that domestication of *PIF* transposases is frequently accompanied by the co-domestication of a cognate MADF protein.

Room 8-06

Functional analyses of recurrently domesticated PIF-transposasesin Drosophila melanogaster

Fatema Begum Ruma[1], Dragomira N. Markova[1], Diwash Jangam[2], Cedric Feschotte[3], Esther Betran[1]
[1]Department of Biology, University of Texas at Arlington, Arlington, TX, USA. [2]Stanford University, Palo Alto, CA, USA. [3]Cornell University, Ithaca, New York, USA

Abstract

Transposable elements (TEs) are segments of DNA that encode for a protein and thus move and amplify within a host genome. An increasing number of findings revealed fascinating insight that TE encoded proteins can be co-opted by their host for its own benefit and become an important contributor to the origination of new host proteins. Here, we study the function of four domesticated transposases from the *PIF/Harbinger*DNA family of TEs in *Drosophila melanogaster*, named ***D****rosophila* ***P****IF* ***L****ike* ***G****enes* (*DPLGs*). All four *DPLGs* in *D. melanogaster* are highly diverged, under purifying selection, and most likely arose through independent domestication events. *DPLGs* co-express with transcription factors (like MADF transcription factor) across development. HA-tagged *DPLG2-4* co-localize with DNA in ovary nuclei suggesting that *DPLGs* are potentially domesticated as regulatory proteins. Protein localization and RNA-Seq studies show similar and overlapping patterns of localization for *DPLG2-4* in the ovaries supporting their functional relatedness. A subset of the *DPLGs* are also involved in neuronal and gonadal functions, and affect the viability*,* fertility and survival of *D. melanogaster*. Currently, we are further elucidating the genetic interactions of these four genes based on loss-of-function mutants and gene expression which could give us more insights into the regulatory pathway they are part of. Moreover, given the functional relatedness of these independently domesticated *PIF* transposase, we provide a model in which the domestication of a transposase might promote the domestication of related transposases from same TE family. This model can shed new light on evolutionary dynamics and new gene origination that have shape the genomes.

Room 8-07

Assessing the influence of regulatory landscapes on gene expression across tissues

Mary Lauren Benton[1], Douglas M Ruderfer[2], John A Capra[3]
[1]Baylor University, Waco, TX, USA. [2]Vanderbilt University Medical Center, Nashville, TN, USA. [3]University of California San Francisco, San Francisco, CA, USA

Abstract

The regulatory genome is involved in all essential cellular and evolutionary processes, and many non-coding genetic variants contribute to disease risk by disrupting *cis*-regulatory elements (CREs). Despite evidence that groups of CREs associated with genes (CRE landscapes) influence the phenotypic effects of genetic variation, current strategies for interpreting CRE variation consider CREs in isolation.

We fill this gap by integrating three-dimensional chromatin conformation data with functional genomic and evolutionary characterization of human CREs to study the relationship between CRE landscape attributes and variation in gene expression. We hypothesized that the attributes of the CRE landscape of a gene influences its robustness to genetic variation. Leveraging CRE landscapes defined for genes expressed in ten human tissues, we found that landscapes with more CREs are associated with stability of gene expression across individuals, supporting the hypothesis that CRE landscape complexity mediates stability of expression. Furthermore, CRE landscapes with an increased proportion of tissue-specific elements are associated with tissue-specific gene expression. We also found that housekeeping genes and genes required for cell proliferation have fewer CREs compared to other loss-of-function intolerant genes. However, evolutionary constraint on individual CRE sequences is not associated with CRE landscape size. This work summarizes the CRE landscape composition across tissues to provide insight into how CRE landscapes regulate gene expression. Further study of the effects of CRE alteration in the context of the regulatory landscape will facilitate better interpretation of the relevance of perturbations to the gene regulatory architecture to evolutionary divergence and disease risk.

Room 8-08

Evolution of gene regulatory networks and understanding cis-trans interactions underlying complex traits

Erik DIAZ VALENZUELA[1], Ruairidh Sawers[2], Angélica Cibrian Jaramillo[1]
[1]LANGEBIO, CINVESTAV, Irapuato, Guanajuato, Mexico. [2]Penn State, Pennsylvania, Pennsylvania, USA

Abstract

The process of domestication requires the rapid transformation of the wild morphology into the cultivated forms that humans select for. This process often takes place through changes in the regulation of genes, yet there is no definite pattern on the role of cis- and trans-acting regulatory variation in the domestication of the fruit among crops. Using allele- specific expression and network analyses we characterized the regulatory patterns and the inheritance of gene expression in wild and cultivated accessions of chili pepper, a crop with remarkable fruit morphological variation. We propose that gene expression differences associated to the cultivated form are best explained by cis-regulatory hubs acting through trans-regulatory cascades. We show that in cultivated chili, the expression of genes associated with fruit morphology is partially recessive to those in the wild relative, consistent with the hybrid fruit phenotype. Decreased expression of fruit maturation and growth genes in cultivated chili suggest that selection for loss-of-function took place in its domestication. Trans-regulatory changes underlie the majority of the genes showing regulatory divergence and had larger effect sizes on gene expression than cis-regulatory variants. Network analysis of selected cis- regulated genes, including ARP9 and MED25, indicated their interaction with many transcription factors involved in organ growth and fruit ripening. Differentially expressed genes linked to cis-regulatory variants and their interactions with downstream trans-acting genes have the potential to drive the morphological differences observed between wild and cultivated fruits and provide an attractive mechanism of morphological transformation during the domestication of the chili pepper.

Room 8-09

Reconstructing the 3D chromatin organization of archaic hominins reveals that genome folding shaped human-Neanderthal phenotypic divergence and introgression

Evonne McArthur[1], David C. Rinker[1], Erin Gilbertson[2], Geoff Fudenberg[3], Maureen Pittman[4,2], Kathleen Keough[4], Katherine S. Pollard[4,2,5], John A. Capra[2,1]

[1]Vanderbilt University, Nashville, TN, USA. [2]University of California, San Francisco, San Francisco, CA, USA. [3]University of Southern California Department of Computational and Quantitative Biology, Los Angeles, CA, USA. [4]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA. [5]Chan Zuckerberg Biohub, San Francisco, CA, USA

Abstract

Modification of gene regulation was a driving force in the divergence of modern humans and archaic hominins. While previous investigation has focused on changes in cis-regulatory elements, the three-dimensional (3D) organization of the genome plays a critical role in regulating gene expression by facilitating and insulating enhancer-promoter interactions. However, the role of 3D genome organization changes in recent human evolution has not been explored because the degradation of ancient samples does not permit experimental interrogation of archaic hominin 3D genome folding.

We address this gap by applying novel deep learning methods for inferring 3D genome organization from DNA sequence patterns to Neanderthal, Denisovan, and diverse modern human genomes. Using the resulting genome-wide 3D genome folding maps, we highlight archaic-specific and modern human-specific 3D patterns (*e.g.* chromatin loops). We find that tolerance to 3D genome variation constrained Neanderthal introgression: regions more tolerant of 3D variation in modern Africans are enriched for introgression in modern Eurasians. We also evaluate the legacy of introgression on the 3D genome organization of humans and identify examples where introgression imparted divergent 3D genome folding to Eurasians. For example, using *in silico* mutagenesis, we identify substantial changes in 3D folding patterns for introgressed variants associated with traits relevant to human-Neanderthal differences, including height, fat distribution, and blood pressure. In summary, our application of deep learning to predict archaic 3D genome folding provides a window into previously unobservable molecular mechanisms underlying how genetic differences lead to phenotypic divergence between modern and archaic hominins.

Room 8-10

Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection

Jian-Rong Yang, Feng Chen, Peng Wu, Shuyun Deng, Heng Zhang, Yutong Hou, Zheng Hu, Xiaoshu Chen
Sun Yat-sen University, Guangzhou, China

Abstract

Eighteen of the twenty amino acids are each encoded by more than one synonymous codon. Due to the differential tRNA supply in the cell, synonymous codons are not used with equal frequencies, a phenomenon termed “codon usage bias” (CUB). Previous studies have demonstrated that CUB of endogenous genes trans-regulates the translational efficiency of other genes. We hypothesized similar effects for CUB of exogenous genes on host translation and test it in the case of viral infection, a common form of naturally occurring exogenous gene translation. We analyzed public Ribo-Seq datasets from virus-infected yeast and human cells and showed that virus CUB trans-regulated tRNA availability and therefore the relative decoding time of codons. Manipulative experiments in yeast using 37 synonymous fluorescent proteins confirmed that an exogenous gene with CUB more similar to that of the host would possess decreased translational load on host per unit of expression, whereas the expression of the exogenous gene was elevated. The combination of these two effects was that exogenous genes with CUB overly similar to that of the host would severely impede host translation. Finally, using a manually curated list of viruses, natural hosts and symptomatic hosts, we found that virus CUB tended to be more similar to that of symptomatic hosts than that of natural hosts, supporting a general deleterious effect of excessive CUB similarity between viruses and hosts. Our work revealed repulsion between virus and host CUBs when they are overly similar, a phenomenon potentially useful in forecasting virulence for given virus-host pairs.

Room 8-11

Inteins in the Terminase of Actinophages – Parasites all the Way Down?

Sean Gosselin, Shelly Evia, Johann Peter Gogarten
University of Connecticut, Mansfield-Storrs, CT, USA

Abstract

Phages often contain inteins (aka protein introns) in their terminase. About half of the terminases in Actinophage cluster E and about 10% of the terminases in cluster A1 were found to be invaded by inteins.

The evolutionary histories of terminases and their inteins were reconstructed from their nucleotide and amino acid sequences. Phylogenetic trees, calculated for inteins and exteins separately, clearly revealed the frequent transfer of the terminase inteins between divergent phages within the A1 and E clusters. Phages from the same geographical location (e.g., South Africa) whose terminases intermingled with those from other geographical regions, contained inteins that formed a well-supported clade, suggesting a recent local spread of the intein.

In the case of five phages isolated in the Auckland region, all five phages contain a terminase intein. At first sight, this could be interpreted as an intein invasion of the ancestral phage followed by vertical inheritance of the intein; however, while the extein sequences are divergent (31 polymorphic nucleotide sites), the intein sequences are identical, revealing that the terminase genes were only very recently invaded by the intein.

Are inteins selfish genetic elements (molecular parasites) that are selected for their ability to invade previously uninvaded genes, or have these elements been housebroken and now (also) fulfill a function beneficial to the host? Our analyses show that inteins undergo frequent horizontal gene transfer and that a possible benefit to the phage from possessing the terminase intein has not resulted in a predominance of vertical inheritance.

Room 8-12

Phylogenetic signal and correlation in host susceptibility to different viruses

Camila Souza Beraldo[1], André Coppe Pimentel[2], Marcos Martins[2], Ben Longdon[3], Rodrigo Cogni[2]
[1]University of Helsinki, Helsinki, Uusimaa, Finland. [2]University of São Paulo, São Paulo, São Paulo, Brazil. [3]University of Exeter, Penryn, Cornwall, United Kingdom

Abstract

Host shifts — when a pathogen jumps from one host species to another — have been described as one of the main factors leading to emerging infectious diseases. The harm that a pathogen causes to a host (virulence) varies following a host shift. Differences in susceptibilities among host species mean that pathogens may be more likely to switch between certain groups of hosts. Factors that determine the variation in host susceptibility are still unknown, but one possible predictor is the host evolutionary history. In this study, we examined how phylogenetically related hosts vary in susceptibility when dealing with infections of two viruses differing in pathogenicity. We infected 41 species of Drosophilidae with *Drosophila* A virus (DAV), a virus initially described as avirulent, and we measured host mortality (virulence) and virus replication (viral load). Then, we compared our results to previously collected data from the virulent *Drosophila* C virus (DCV). We found large variation in DAV virulence and viral load, with benign infections in some cases and high mortality in others. There was phylogenetic correlation in viral load, with species presenting similar viral load clustering together in the phylogeny. However, we did not find correlation for virulence, indicating that DAV virulence was not predictable based on viral load. In addition, we did not find correlation between DAV and DCV results, suggesting that variation in host susceptibility cannot be anticipated by other pathogens infections.

Room 8-13

Exploring the combined fitness landscape of co-evolving human and viral proteins

Michael J Chambers[1,2], Thomas E Dever[3], Meru J Sadhu[4]
[1]NHGRI, Bethesa, MD, USA. [2]Georgetown University, Washington, DC, USA. [3]NICHD, Bethesa, MD, USA. [4]NHGRI, Betheda, MD, USA

Abstract

Here we propose a novel approach to explore an evolutionary arms race between the human innate immune system and poxviruses. Protein kinase R (PKR) is a component of the innate immune system that detects double-stranded RNA (dsRNA), a general byproduct of viral infection. Once activated, PKR halts translation within the cell and prevents viral replication. Poxviruses encode K3L, which serves as a counter defense against PKR. K3L directly antagonizes PKR and prevents a halt in translation, which enables the virus to replicate. Both PKR and K3L are under diversifying selection pressure, meaning that specific residues in each of these proteins are evolving at a rapid rate. It is this selection pressure that drives an evolutionary arms race in which superior variants of PKR and K3L are continually pursued. The focus of our proposal is to prospectively generate and characterize missense variants that are available to PKR and K3L, which will highlight critical residues and provide some insight as to how PKR combats a rapidly evolving viral antagonist. Our approach is to systematically generate hundreds of single-residue missense variants of both PKR and K3L, then combine all variants together into an all-pairs library. To characterize this all-pairs library we have modernized a yeast growth assay, which allows us to characterize hundreds of thousands of unique variant pairs in a single culture. This approach allows us to scan a vast combinatorial space in a single experiment, highlighting critical residues and providing a glimpse into the evolutionary fitness landscape of PKR and K3L.

Room 8-14

Evolution of the HIV-1 protease folding stability

David Ferreiro[1,2], Miguel Arenas[1,2,3]

[1]Biomedical Research Center (CINBIO), University of Vigo, Vigo, Spain. [2]Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain. [3]Galicia Sur Health Research Institute (IIS Galicia Sur), Vigo, Spain

Abstract

The emergence of variants resistant to host immune systems and antiretroviral therapies is fundamental for viruses. However, resistance mutations can also result in some costs to the viral protein, such as decreasing the normal protein activity and folding stability. Concerning the latter, here we explored the influence of relevant resistance mutations on the folding stability of the HIV-1 protease (PR) and the overall evolution of folding stability of this protein. In particular, we analyzed the protein folding stability in a variety of HIV-1 PR variants with and without common resistance mutations to therapies and we found that these mutations decrease (in average) the folding stability, although some of them can also maintain or increase it. We also found that the folding stability of the HIV-1 PR fluctuated widely over time, without showing a clear trend. We conclude that resistance mutations can decrease the HIV-1 PR folding stability but this pattern cannot be extended to all resistance mutations and, that the evolution of the HIV-1 PR folding stability is complex with multiple fluctuations over time.

Room 8-15

Evolutionary analysis and lineage 1 designation of SARS-CoV-2 genomes

Xiaolu Tang[1], Ruochen Ying[1], Xinmin Yao[1], Guanghao Li[2], Changcheng Wu[1], Yiyuli Tang[2], Zhida Li[3], Bishan Kuang[3], Feng Wu[3], Changsheng Chi[3], Xiaoman Du[3], Yi Qin[3], Shenghan Gao[4], Songnian Hu[4], Juncai Ma[4], Tiangang Liu[5], Xinghuo Pang[6], Jianwei Wang[7], Guoping Zhao[8], Wenjie Tan[9], Yaping Zhang[2], Xuemei Lu[2], Jian Lu[1]

[1]Peking University, Beijing, China. [2]Kunming Institute of Zoology, Kunming, China. [3]Yuxi Rongjian Information Technology Co., Ltd., Yuxi, China. [4]The Institute of Microbiology, Beijing, China. [5]Wuhan University, Wuhan, China. [6]Beijing Center for Disease Prevention and Control, Beijing, China. [7]Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China. [8]Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [9]Chinese Center for Disease Control and Prevention, Beijing, China

Abstract

The pandemic due to the SARS-CoV-2 virus, the etiological agent of Coronavirus Disease 2019 (COVID-19), has lasted for more than one year. And the SARS-CoV-2 genome has accumulated a large number of genomic variants. In the early outbreak of this pandemic, based on two tightly linked SNPs, we previously divided SARS-CoV-2 into two major lineages: the ancestral “S lineage” (U8,782 and C28,144) and the derived “L lineage” (C8,782 and U28,144). To improve the tracing of the viral genomes’ evolution during the development of the pandemic, we further divided the L lineage into two major sublineages (L1 and L2) using SNVs at sites 3037, 14408, and 23403. Subsequently, we categorized them into 130 sublineages (37 in S, 35 in L1, and 58 in L2) based on marker SNVs at 201 additional genomic sites. This lineage/sublineage designation system has a hierarchical structure and reflects the relatedness among the subclades of the major lineages. The sublineages exhibited substantial differences spatially and temporally, which provides a basis for us to track the evolution and transmission dynamics of SARS-CoV-2.

Room 8-16

Contrasting biotic and abiotic drivers of adaptive evolution in a host-pathogen conflict

Michelle Hays
Stanford University, Stanford, CA, USA

Abstract

To be successful, organisms must adapt to both abiotic (e.g. environmental pressures) as well as biotic (e.g. parasites) selection pressures. Although both of these pressures can drive evolutionary innovation, theory predicts that antagonistic relationships may drive recurrent episodes of adaptation. Dissecting these selective pressures has implications for understanding many biological processes, such as treatment of infectious disease and cancers, where cells are adapting to both host immunity (biotic) and environmental (abiotic) therapeutics. Some RNA viruses of yeasts encode a “killer” toxin-antitoxin system, which protects virus-bearing killer cells but kills virus-lacking sensitive cells. As a result, these RNA viruses can be maintained in host populations despite imposing a metabolic cost to their host. Killer requires multiple viral genomes for toxin production as well as host cellular components. With sensitive cells in the environment, this system is a four-party genetic conflict, with competing fitness tradeoffs.

Together with Arjan de Visser’s lab we are identifying beneficial mutations in populations of coevolved killer and sensitive cells over 500 generations. We are identifying the genomic basis of toxin-resistant lineages, other beneficial mutations, and looking for evidence of killer cell counter-adaptation in the face of resistant competitors. We expect this work will give insight into how killer phenotypes drive evolution in heterogeneous competing populations and will begin to tease apart the complex and often competing selective pressures at play in this system. Understanding how antagonistic coevolution and environmental selective pressures both drive adaptation, is fundamental for understanding how genomes evolution and biological diversity.

Room 8-17

Analyses of *Culicidae* genomes reveal ancient domestication of a *synctitin* – like viral envelope gene in mosquitoes

Tamanash Bhattacharya[1,2], Kathryn E Kistler[1], Harmit S Malik[1,2]

[1]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Howard Hughes Medical Institute, Seattle, WA, USA

Abstract

Domesticated viral elements often serve important functions in the host. Consequently, they are subjected to strong selective pressure that allow their retention over time. This is especially true for one particular class of viral genes, the *envelope (env)*. While they allow viruses to infect and propagate in host cells, domesticated *env* genes perform novel functions in their adopted hosts. The *synctitin* family of *env* genes, for example, play a critical role in placental morphogenesis in primates and rodents via their membrane fusion activity. Domesticated *env* genes are also found in genomes of invertebrates, like the *Drosophila Iris* gene. Notably, *Iris* homologs are present in other Dipteran insects, like mosquitoes. While past unavailability of reliable sequence information allowed only limited analyses of these *env* genes, recent release of several mosquito genome sequences allowed us to revisit and expand on these past results. Here, we show that *Iris* homologs *(Iris-like)* occur in all extant members of the *Culicidae* family, following a independent domestication event 140-200 million years ago. While *Drosophila Iris* genes lack membrane fusion features present in canonical envelope proteins, such features are conserved in all *Iris-like* genes. Branch site analysis of mosquito *Iris-like* suggest that signatures of positive selection are absent in the vast majority of *Iris-like* lineages, which lies in contrast to *Drosophila Iris,* which is under strong positive selection. Collectively, *Iris-like* represent the single-oldest domestication event for any known endogenous viral element in animals and likely perform functionally distinct roles in mosquitoes compared to the fruit fly *Iris* genes.

Room 8-18

Dynamics and Metagenomics of Oral Disease in Scandinavian Brown Bear (*Ursus arctos*)

Adrian E Forsythe[1], Jaelle Brealey[2,3], Maya Gadhvi[1], Henrique Leitão[1], Thijs Hofstede[1], Katerina Guschanski[1,4]

[1]Department of Animal Zoology, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden. [2]Uppsala, Uppsala, Sweden. [3]Department of Natural History, Norwegian University of Science and Technology, Trondheim, Norway. [4]Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh, United Kingdom

Abstract

Anthropogenic disruptions to the planet's ecosystems impact ecological communities at the macro- and microscale, including mammalian species and their associated microbiomes. Host-associated microbial communities are generally regulated by external environmental and internal host-specific factors, including the host's immunity. The stable state of these communities can become disrupted by changes to any of these factors, resulting in dysbiosis and disease. For example, the loss of genetic diversity in a wild population can disrupt regulation of host-associated microbial communities by the host. Using extensive museum records, we have quantified the prevalence of oral disease in Swedish brown bears over the last 200 years. During this period, the population narrowly avoided extinction and has since recovered to ca. 3000 individuals. The prevalence of oral disease is inversely proportional to bear population size: Highest during and shortly after the bottleneck and decreasing as the population recovered. Using calcified dental plaque, we have characterised a distinct oral microbial community within individuals with dental caries compared to those that are healthy, and is associated with the abundance of several bacterial taxa, some of which are (close relatives of) human opportunistic oral pathogens. Furthermore, disease-associated taxa are functionally enriched in the molecular pathways associated with carbohydrate fermentation and acid production. This work demonstrates interaction of host-associated microbial communities and disease prevalence in Scandinavian brown bears, corresponding with human-driven population decline. Understanding how host-associated microbial communities respond to host and environmental disturbances may be an important predictor of their impact on host condition.

Room 8-19

Deep phylogenetic splits in a single-species genus, *Agraulis vanillae.*

Jules Claeys, Stephen Montgomery, Francesco Cicconardi
University of Bristol, Bristol, United Kingdom

Abstract

A thorough knowledge of taxonomy and of its basic unit, the species, is key to numerous scientific fields, including evolutionary biology, ecology, conservation and biodiversity. Discerning two close species is sometimes complex due to uncertainties in the definition of species, and the possible for cryptic dissimilarities, or more pronounced diversity, within morphological species. A possible example of such taxonomic problems is in *Agraulis vanillae*. This butterfly belongs to the Heliconiini clade, has at times been included as part of the *Dione* genus, or as possessing multiple subspecies, both based on morphological characters. However, molecular data has confirmed *Agraulis* as its own monophyletic clade, currently with a single species, *A. vanillae*. Recent work, however, provided hints of deep splits between North and South American populations. Here, we extend the geographic spread of the sampling to test this hypothesis. We compile cDNA and mtDNA sequences across 9 different loci for *Agraulis*, its sister genus *Dione*, and several *Heliconius* species. Maximum likelihood gene trees were built to compare the divergence within *A. vanillae* to divergence between other Heliconiini species in the phylogenetic tree. We also provide date estimations for internal nodes in the tree to compare split estimates within *A. vanillae* to those of recognised species pairs and races in *Heliconius*. These data permit a better understanding of the phylogeny of Heliconiini and discern whether the disparity between North and South American *Agraulis* populations represent two distinct species.

Room 8-20

The evolution of molecular toolkits for biosilicification in demosponges (Porifera)

Maria Eleonora Rossi[1,2], Nathan J. Kenny[3], Bruna Plese[2], Sergi Taboada[4,2], Vasiliki Koutsoveli[2], Davide Pisani[1], Ana Riesgo[2,5]

[1]School of Earth Sciences, University of Bristol, Bristol, United Kingdom. [2]Life Sciences Department, The Natural History Museum, London, United Kingdom. [3]Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, United Kingdom. [4]Facultad de Ciencias, Universidad Complutense de Madrid, Madrid, Spain. [5]Department of Biodiversity and Evolutionary Biology National Museum of Natural Sciences (CSIC), Madrid, Spain

Abstract

Among the four classes of Porifera, three of them construct siliceous skeletons but through divergent enzymatic pathways. In demosponges, the silicification occurs by polycondensation of silica using silicases around an axial filament formed by a protein called silicatein. Most demosponges can produce more than one spicule type, in some occasions with convoluted ornamentations. Some demosponges have more than one silicatein gene (with highly divergent evolutionary pathways). In calcareous sponges, the many alpha-carbonic anhydrases present in the group are linked to the production of several spicule types. Whether the diversity of siliceous spicules is linked to larger molecular complexity in demosponges, is completely unknown. Here we used complete transcriptomes of 71 sponges (19 newly assembled) to create a sound phylogenomic framework to explore the evolution of biosilicification within demosponges and reconstruct the phylogeny of the phylum Porifera. The enzymes required to produce siliceous spicules including silicases, silicateins, and silicon transporters, were screened within our transcriptomic and other datasets available. Character reconstructions were performed in our datasets to understand the evolution of the spicules. Finally, we used BAMM tools to detect and quantify heterogeneity in evolutionary rates across sponges with diverging silicification levels and molecular machineries.

Room 9-01

A union between long-lost cousins: whole genome sequencing reveals introgression between desert warthogs and common warthogs in eastern Africa

Christian H.F. Jørgensen[1], Genís Garcia-Erill[1], Vincent Muwanika[2], Xi Wang[1], Malthe S. Rasmussen[1], Yvonne A. de Jong[3], Thomas M. Butynski[3], Laura D. Bertola[1], Hans R. Siegismund[1], Anders Albrechtsen[1], Rasmus Heller[1]

[1]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [2]Department of Environmental Management, Makerere University, Kampala, Uganda. [3]Eastern Africa Primate Diversity and Conservation Program & Lolldaiga Hills Research Programme, Nanyuki, Kenya

Abstract

African pigs have a particularly contentious evolutionary history. Until recently, desert warthogs (Phacochoerus aethiopicus) and common warthogs (P. africanus) were considered a single species (P. aethiopicus), but molecular evidence suggests they diverged at least 4.4 million years ago. We sequenced the first whole-genomes of four desert warthogs and 35 common warthogs from six countries that cover a large part of their range. From the autosomal data we infer that the divergence between the two species occurred much later, around 1.8 million years ago at the earliest, in line with the paleontological record. Hence, we solve one of the key controversies surrounding warthog evolution and divergence. In contrast to previous studies, we found clear evidence of an evolutionary origin of the common warthog in western Africa and subsequent movement eastwards and later southwards. Despite their considerable genetic differentiation, we found evidence of desert warthog introgression into an ancestral common warthog population, presumably as the eastwards movement of common warthog brought the two into contact around 800,000 years ago. As a consequence, all eastern and southern African common warthog populations contain up to 3% of desert warthog ancestry. It is possible that this introgression from native desert warthogs was adaptive for the common warthog as in eastern and southern Africa it encountered distinct habitats and diseases, including the deadly African Swine Fever. In summary, we provide new insights into the evolution of an important African suid lineage, revealing a complex evolutionary history involving movement, introgression, and adaptation to novel habitats.

Room 9-02

Whole-genome analysis of multiple wood ant population pairs supports similar speciation histories, but different degrees of gene flow, across their European range

Beatriz Portinha[1,2], Vítor C Sousa[1], Jonna Kulmuni[2,3], Pierre Nouhaud[2]

[1]cE3c, Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal. [2]Organismal & Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland. [3]Tvärminne Zoological Station, University of Helsinki, Hanko, Finland

Abstract

Reconstructing divergence histories using genetic data has become gold standard in speciation genomics. It is usually performed by analysing a single population from each species, assuming that the inferred divergence history represents the history between the species. However, this assumption is rarely explicitly tested, and it may not be met when species diverge with spatially heterogeneous gene flow. Here, we tested the extent to which the divergence history inferred from two heterospecific populations varies depending on their geographic locations with the wood ant species Formica polyctena and F. aquilonia, which have contrasting distributions in Europe. With whole-genome polymorphism data obtained from 20 individuals sampled in multiple populations across both species ranges, we reconstructed the histories of distinct heterospecific population pairs using a coalescent approach based on the site frequency spectrum. Analysis of the different pairs consistently supports a scenario of divergence with gene flow. Results suggest that divergence started in the Pleistocene (ca. 500 kya) and occurred with continuous asymmetrical gene flow from F. aquilonia to F. polyctena until a recent time, when migration stopped (2-19 kya, depending on the population pair). Interestingly, we found support for contemporary reciprocal gene flow between the pair of sympatric populations in Finland, where the species hybridise, but no signature of recent bidirectional gene flow in other localities. Overall, our results suggest that histories reconstructed from single population pairs may be reliable and applicable at the species level. However, the context of the populations may still affect inferences of the recent past.

Room 9-03

Oral microbiome evolution in closely related species

Markella Moraitou[1], Jaelle Brealey[1,2], James Fellows Yates[3], Henrique Leitão[1], Thijs Hofstede[1], Christina Warinner[3,4], Katerina Guschanski[1,5]

[1]Department of Ecology and Genetics/Animal Ecology, Uppsala University, Uppsala, Sweden. [2]Department of Natural History, NTNU University Museum, Trondheim, Norway. [3]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany. [4]Department of Anthropology, Harvard University, Cambridge, MA 02138, USA. [5]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Host-associated microbiomes play an important role in many biological processes in the host, ranging from digestion to health to reproduction. Similarly, host diet, ecology, health state and evolutionary past influence the composition of microbial communities. How these different factors structure host-associated microbiomes has been little studied outside of the gut microbiome. Here, we use a metagenomic approach to evaluate the contribution of host evolutionary history and ecology to oral microbiome composition and function in a group of closely related allopatric gorilla taxa. We consider three gorilla subspecies that have diverged from each other 250,000-10,000 years ago, and differ in ecology, diet and social structure. Gorillas are therefore well suited to study the drivers of microbiome diversity on a short evolutionary scale. Using metagenomic analyses of dental calculus collected from 45 museum specimens, we show that gorilla subspecies differ from each other in the taxonomic and functional composition of their oral microbiome. Qualitative assessment of eukaryotic reads retained in dental calculus provides insight into gorilla diet and broadly supports taxon-specific dietary preferences. Ecology, rather than evolutionary relationships, appears to correlate with the observed differences. Finally, we were able to *de novo* assemble several bacterial draft genomes, some of which represent novel taxa.

Room 9-04

Custom-made or universal? The use of target capture to bridge macro- and microevolutionary processes in tillandsioid bromeliads (Tillandsia spp), a Neotropical rapid radiation

Gil Yardeni[1], Marylaure de la Harpe[2], Michael Barfuss[1], Walter Till[1], Clara Groot Crego[1], Norma Rivera[3], Juan Viruel[4], Margot Paris[5], Jaqueline Hess[6], Thorsten Kroemer[7], Christian Lexer[1], Ovidiu Paun[1], Thibault Leroy[1]

[1]University of Vienna, Vienna, Austria. [2]University of Zürich, Zürich, Switzerland. [3]Max Planck Institute for Evolutionary Biology, Plön, Germany. [4]Royal Botanic Gardens, Kew, London, United Kingdom. [5]University of Fribourg, Fribourg, Switzerland. [6]Helmholtz Centre for Environmental Research, Halle, Germany. [7]Universidad Veracruzana, Xalapa, Mexico

Abstract

The study of rapid radiations calls for integrative approaches, bridging phylogenetics and population genetics to uncover the genomic substrate of diversification at different time scales. Target capture approaches emerged as an important tool to study evolutionary radiations in non-model taxa, enabling researchers to retrieve large data sets with few genomic resources. While developing taxon-specific target capture kits requires sustained efforts, universal kits are readily available but may offer comparatively shallower insights, especially in short evolutionary timescales. The species-rich and ecologically diverse Bromeliaceae family provides an excellent system for studying the drivers and constraints of rapid, adaptive radiations. We present a taxon-specific target sequence capture set for bromeliads, designed to address a wide range of evolutionary hypotheses by targeting 1,776 coding regions, including neutral regions and genes putatively involved in several key traits. We compare our bait set to the ‘universal’ Angiosperms353 probe set, examining their power to resolve phylogenomic relationships using concatenation and species tree methods, estimate population genetic statistics and infer admixture with a focus on *Tillandsia* subgenus *Tillandisa,* a particularly young and diverse subgenus (~6 Mya). The taxon-specific set results in high enrichment success across the entire family, yet both kits offer abundant information. The taxon-specific set outperforms the universal set with respect to gene tree concordance and inferred population structure and provides reliable data for inference of admixture and population structure. Importantly, the performance of both kits is comparable for phylogenomic analysis, highlighting the vast potential of universal kits in research at different evolutionary scales.

Room 9-05

Gene tree-species tree reconciliation provides a model of genome duplication history in Chelicerates.

Alan J.S. Beavan, Davide Pisani, Philip C.J. Donoghue
University of Bristol, Bristol, United Kingdom

Abstract

Genome duplication is believed to have occurred several times during the evolution of chelicerates. Current hypotheses suggest genome duplication events in the ancestor of horseshoe crabs, as well as at some point during the evolution of Arachnopulmonata (a clade including spiders and scorpions). Precisely where genome duplications occurred on the tree, as well as how many and when, has important implications for understanding the evolution of chelicerates and the effects of genome duplication on macroevolution. Here, gene tree-species tree reconciliation methods were applied to distinguish between the different genome duplication histories possible in chelicerates. Importantly this method considers proteome-wide patterns of duplication, rather than focussing on individual gene families. Gene duplication events we infer provide strong evidence for genome duplication in horseshoe crabs. Further evidence suggests genome duplication in Araneae (spiders), rather than in stem Arachnopulmonata, as others have suggested. A distribution of synonymous mutations accrued in paralogous gene pairs that duplicated in stem horseshoe crabs was inferred, which supports three rounds of genome duplication, in accordance with previous findings. In spiders, Bayesian gene tree dating methods suggest two rounds of genome duplication. In addition to this, further Bayesian dating analyses were performed to provide estimates for the timing of genome duplication events in these lineages, helping understand how genome duplications may have impacted the macroevolution of these lineages.

Room 9-06

Investigating the cryptic potential of intergenic regions to form transmembrane domains and its link to de novo gene birth in 332 yeast genomes

Emilios Tassios[1,2], Nikos Vakirlis[1], Christoforos Nikolaou[1]

[1]Biomedical Sciences Research Center 'Alexander Flemming', Athens, Greece. [2]University of Crete, Heraklion, Greece

Abstract

De novo gene birth from a previously non-genic region is a mechanism met across a wide variety of organisms. The properties of the non-genic region may affect the functional potential and evolutionary future of the nascent gene in ways that are yet to be clearly defined. Recently, it has been shown that the intergenic regions in budding yeast genome harbor pervasive cryptic potential to, if translated, form transmembrane domains (TM domains). This potential to form TM domains is greater than expected by chance given the nucleotide composition of intergenic sequences, and early evidence shows that any peptides that might emerge from such regions could be more likely to be beneficial to the cells. Here, we investigated this newly discovered genomic tendency across the entire Saccharomycotina subphylum, using data from 332 recently published genomes. We consistently found a robust, statistically significant enrichment in the potential to form TM domains in the intergenic regions of most species. This enrichment exhibits variation from species to species which cannot be readily explained by intrinsic genomic and intergenic properties, by the orientation of neighboring genes or even by the theoretical hydrophobic content of the regions. However, species belonging to the same genus appear to have similar levels of enrichment. Our preliminary findings hint towards yet unknown sequence motifs that might account for this genomic property, and hence could be essential to the process of de novo gene birth in yeasts.

Room 9-07

Ongoing and ancient introgression in *Oreochromis* tilapias

Adam Ciezarek[1], Antonia Ford[2], Tarang Mehta[3], Will Nash[3], Benjamin Ngatunga[4], Asilatu Shechonge[4], Rashid Tamatamah[4], Nasser Kasozi[5], di Palma Federica[6], Wilfried Haerty[3], Martin Genner[7], George Turner[8]

[1]Earlham institute, Norwich, United Kingdom. [2]Roehampton University, London, United Kingdom. [3]Earlham Institute, Norwich, United Kingdom. [4]Tanzania Fisheries Research Institute, Dar es Saleem, Tanzania, United Republic of. [5]National Agricultural Research Organisation, Arua, Uganda. [6]University of East Anglia, Norwich, United Kingdom. [7]Bristol University, Bristol, United Kingdom. [8]Bangor University, Bangor, United Kingdom

Abstract

The *Oreochromis* tilapias are an economically important group of fish for aquaculture, whose production has expanded dramatically in the last two decades. Linked to this success has been the introduction of exotic species, including into Tanzania, a hotspot for *Oreochromis* diversity. Introductions have had significant negative ecological impacts on indigenous *Oreochromis* species, including species displacement and loss of population structure through hybridisation, with potential implications for local adaptations. This history of introgression has made untangling the population history of *Oreochromis* difficult. We address this issue using genome-wide sequencing of 575 individuals from 31 *Oreochromis* species across Tanzania and east Africa, producing a first phylogenomic tree of the genus using coalescent and concatenation methods. We identify strong signatures of ancestral introgression between species, the extent of which is determined by habitat type, shared drainage basin occupation and phylogenetic distance between species. Furthermore, the introduction of populations for aquaculture has led to modern hybridisation with native populations in several different water bodies across Tanzania. We further aim to identify whether the same genomic regions to have repeatedly undergone introgression as a result of these introductions. We anticipate that our results will have important implications when managing the translocation of species for food production.

Room 9-08

A Transcription-Based Investigation on Germline-Specific Functional Signatures in Metazoa

Giovanni Piccinini, Liliana Milani
University of Bologna, Department of Biological, Geological, and Environmental Sciences, Bologna, Italy

Abstract

Sexual reproduction in pluricellular organisms, like most Metazoa, usually involves the establishment of a specific cellular lineage designated for genetic inheritance across generations: the germline. Recently, a set of genes observed in animals were proposed as determinants for both somatic pluripotent and germ cells, suggesting the presence of a metazoan evolutionary shared molecular signature involved in the characterization of such cellular lineages (the Germline Multipotency Program: GMP).

In the present study, we investigated germline-related molecular patterns in a metazoan-wide transcription-based analysis. From online databases we selected RNA-Seq experiments including both somatic and female germline-related samples (from whole gonads to isolated germ cells) for a total of 9 species of Metazoa, covering 7 phyla. We identified upregulated transcripts in germline-related samples within each species, and tried to find similar cross-specific signatures. We could not find a clear metazoan-wide orthology signal, with only two genes whose orthologues were upregulated in all but one species. Not even GMP determinants could be retrieved as ubiquitously upregulated in all germline-related samples. However, a broader signal was present for the functional/domain annotation, with 16 InterProScan codes (almost exclusively relative to DNA replication/nuclear import) biased toward germline-related transcripts in almost all species (confirmed by comparison between specific GO enrichment analyses). The difficulty in finding clear sequence homology between upregulated transcripts of different species might be due to asynchrony between samples. Nonetheless, the retrieval of shared molecular functions might suggest that usage of similar mechanisms reflects the evolution of animal germline-specific genes, rather than unequivocal ubiquitous gene homology.

Room 9-09

The Global Ant Genomics Alliance (GAGA): towards a phylogenomic understanding of ant social evolution and diversification

Joel Vizueta[1], Lukas Schrader[2], Zijun Xiong[3], Jacobus J. Boomsma[1], Guojie Zhang[1,3,4]

[1]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [2]Institute for Evolution and Biodiversity, Universität Münster, Münster, Germany. [3]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [4]China National Genebank, BGI-Shenzhen, Shenzhen, China

Abstract

The ants have evolved a stunning global diversity with more than 15,000 extant species belonging to over 330 genera. Their ecological success is rooted in division of labour comparable to, but organizationally beyond, what somatic cells in a metazoan body achieve. The ants are pinnacles of social evolution and dominant members of ecological communities as they interact with other species and have important roles in nutrient cycling and seed dispersal. GAGA was launched in 2017 to generate and study high-quality genomic, transcriptomic, microbial and ecological data for >200 species, covering most if not all major ant taxa. During the first phase of the project, samples of more than 250 species from 130 genera have been collected, involving 64 collectors in 30 countries. Stringent DNA-quality requirements and limited biomass availability were challenging but mostly overcome and we have now reached our first milestone of >100 genomes. We have generated high-quality and contiguous assemblies using PacBio long-reads combined with short-read stLFR for 124 species. Ten assemblies were improved to chromosome-level resolution using Hi-C libraries, while some species for which we had little biomass were assembled using stLFR data. Although most of the genomes are from Myrmicine and Formicine subfamilies, reflecting the high genus/species diversity of these clades, 12 of the world’s 17 ant subfamilies are represented in our dataset. We will provide an overview of GAGA’s progress so far, and of the major questions that ongoing analyses are addressing to better understand the astonishing adaptive radiation of the ants.

Room 9-10

Evolution of Regulatory Complexity through Network Expansion and Genome Organization

Sam H. A. von der Dunk, Paulien Hogeweg, Berend Snel
Utrecht University, Utrecht, Utrecht, Netherlands

Abstract

The evolution of cellular complexity is a major question in evolutionary biology. Increase of regulatory complexity results from an interplay between adaptive and non-adaptive forces at multiple organizational levels. Using a computational model, we investigate how complexity emerges in the *Caulobacter crescentus* cell-cycle regulatory network. In 10 replicate *in silico* evolution experiments we study how cells tasked with executing a successful cell-cycle adapt to a gradient of limited nutrient conditions.

We find adaptive complexification of the gene regulatory network, which enhances cell-cycle behaviour and overcomes the energetic cost of genome expansion. Several distinct trajectories are traversed by evolution in the different replicates. In four replicates, cells evolve a generalist strategy to cope with a variety of nutrient levels, and in two other replicates, different specialist cells evolve in specific nutrient levels. The generalist and specialist strategies are contingent on the regulatory mechanisms that arise early in evolution, but they are not directly linked to network expansion and overall fitness.

Our simple model shows that gene regulation cannot be understood from the gene regulatory network alone. Genomic location of critical genes is used to time the cell-cycle, forming a *de* novo cell-cycle checkpoint. In real organisms, many more organizational levels beside the regulatory network and genome organization are available, such as the protein-protein interaction network and chromatin modifiers, so there are more opportunities for integration giving rise to complex regulation.

Room 9-11

New faster algorithms to find combinatorially complete datasets in high-throughput mutagenesis experimental data

Aleksandr Poslavsky, Dmitry N Ivankov
Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Abstract

Epistasis, the dependence of the mutational effect on the genetic background, is the principal obstacle for predicting phenotype and fitness from genotype. The most straightforward approach to identify epistasis is to analyze the so-called combinatorially complete datasets, consisting in the simplest case of a reference genotype, two different single mutants, and a double mutant with both of the single mutations.

A combinatorially complete dataset built from K single mutations at K positions contains 2^K genotypes and represents a K-dimensional hypercube in the genetic space. Using random mutagenesis, researchers measured phenotypes of thousands-to-millions genotypes, which provides an invaluable source of the information about epistasis. However, identification of all hypercubes from such data is a non-trivial task since the measured genotypes populate genetic space randomly.

Previously, we presented the first algorithm of that kind called “HypercubeME”. The central concept in the algorithm is the diagonal of a hypercube: if two hypercubes have the same diagonal, they are parallel to each other; if, in addition, they are located at the distance of one mutation, they form a next-dimension hypercube. Now, we present three more algorithms, which use the same concept of hypercube diagonal but are much more efficient in hypercube identification. Two of them have complexity $O(N \times L \times A)$ while the third one has complexity $O(N \times L \times [\log L + A])$, where L is the length of the sequence and A is the size of the alphabet, which outperforms the original algorithm having $O(N^2 \times K)$ complexity.

Room 9-12

Using 2b-RAD enzymes wisely: From loci number to functional structure.

Carles Galià-Camps[1], Cinta Pegueroles[1], Carlos Carreras[1], Xavier Turon[2], Marta Pascual[1]
[1]Departament de Genètica, Microbiologia i Estadística - IRBio, Universitat de Barcelona, Barcelona, Catalunya, Spain. [2]Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Blanes, Catalunya, Spain

Abstract

RAD-seq methods allow to reduce the fraction of genome analysed in population genomic studies, solving efficiently the economical and informational trade-off. Among them, 2b-RAD sequencing has the interesting characteristic to allow the performance of secondary genomic reductions using base-selective adaptors, optimizing sequencing resources. However, the efficiency of the different enzymes, the number of yielded loci, their genic-intergenic structure and how base-selection affects these estimates remain unknown. Here, we tested the efficiency of the 2b-enzymes AlfI, BaeI and CspCI on 80 chromosome-level eukaryote genomes (from 150Mb to 3Gb). Briefly, we first simulated "*in-silico*" digestions on the genomes. Following, we simulated the base-selective approach by computationally selecting the loci ending with A-T (W) or C-G (S). Finally, we mapped back the loci from all datasets (total loci, W-loci and S-loci) to their reference genomes to infer the proportion of genic and intergenic regions. Our results demonstrate that the number of loci significantly correlates with genome size, being CspCI the enzyme with the best correlation. AlfI provided the highest number of loci while BaeI provided the lowest, yet the latter is the enzyme with the highest probability to obtain paralogous loci. We evaluated whether the number and functional structure of selected loci is proportional to the non-selected dataset. This work will be key when deciding what enzyme and base-selective adapters use according to researchers' interests and their economic availability in future population genomic studies.

Room 9-13

Testing the effects of 2b-RAD loci selection on the tunicate species Styela plicata.

Carles Galià-Camps[1], Carlos Carreras[1], Xavier Turon[2], Marta Pascual[1]

[1]Departament de Genètica, Microbiologia i Estadística - IRBio, Universitat de Barcelona, Barcelona, Catalunya, Spain. [2]Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Blanes, Catalunya, Spain

Abstract

2b-enzymes have the potential to perform a secondary reduction of the yielded genome fragments by applying base-selective adaptors. However, it has never been proved that the loci generated by both selected and non-selected strategies provide the same genetic information. Here, we tested if the same genotypes and genetic distances among individuals are obtained when using non-selective (N) and base-selective (W) adaptors for the invasive ascidian *Styela plicata*. The small genome size of this species (~400 Mb) makes it a perfect model organism to test different levels of genotyping without high sequencing costs. We used AlfI and CspCI to digest the genome of four *Styela plicata* individuals followed by library preparation with and without base-selective adaptors. Our data could demonstrate for AlfI that regardless using N or W selection, we recover the same genotype for the 92% of the shared loci. The major incongruence factor is the shallow mean depth per locus in the N dataset, and only 1.5% of the mismatches is produced by methodological errors during library preparation and sequencing. We also estimated the number of reads necessary for this study organism in order to avoid genotype mismatches due to low sequence mean depth per locus in the future. From these results, we propose a pipeline to guide laboratory procedures, based on a first trial of the enzymes, optimization of library preparation to avoid sequencing ghost bands, and a bioinformatic estimation of the number of sequences needed to allow a correct genotyping with an adequate sequencing effort.

Room 9-14

Penguin Toll-like receptor 15 has undergone cryptic pseudogenization and may be involved in the immune response to *Aspergillus* spp.

Steven R Fiddaman[1], Michal Vinkler[2], Simon G Spiro[3], Hila Levy[1], Christopher A Emerling[4], Amy C Boyd[5], Evangelos A Dimopoulos[1], Juliana A Vianna[6], Theresa L Cole[7], Hailin Pan[8], Miaoquan Fang[8], Guojie Zhang[7,8,9,10], Tom Hart[1], Laurent AF Frantz[11,12], Adrian L Smith[1]

[1]University of Oxford, Oxford, United Kingdom. [2]Charles University, Prague, Czech Republic. [3]Zoological Society of London, London, United Kingdom. [4]Reedley College, Reedley, CA, USA. [5]Jenner Institute, Oxford, United Kingdom. [6]Pontificia Universidad Católica de Chile, Macul, Santiago, Chile. [7]University of Copenhagen, Copenhagen, Denmark. [8]BGI-Shenzhen, Shenzhen, China. [9]State Key Laboratory of Genetic Resources and Evolution, Kunming, China. [10]Center for Excellence in Animal Evolution and Genetics, Kunming, China. [11]School of Biological and Chemical Sciences, London, United Kingdom. [12]Ludwig Maximilian University of Munich, Munich, Germany

Abstract

Aspergillosis, caused by the fungal pathogen *Aspergillus* spp., is a significant cause of mortality in wild and captive birds. We investigated whether Toll-like receptors could be a genetic determinant of aspergillosis susceptibility in penguins. Toll-like receptors are at the font-line of immune defence and respond to conserved pathogen-associated molecular patterns (PAMPs). We found that *TLR15*, which recognizes fungal agonists, has been pseudogenized at least eight times in the *Eudyptes* (crested) penguins. However, in an analysis of >100 wild *Eudyptes* penguins, we also found that the full-length haplotype persists at low frequencies (<20%). Functionally testing the full-length *Eudyptes TLR15* haplotype, we determined that it is non-functional, despite clearly being expressed. This contrasts with related bird species (Emperor penguin, Northern fulmar and chicken), in which we established the receptor was functional. Given the several independent pseudogenizations of TLR15 and the presence of a full-length, but non-functional, extant haplotype, we propose that the ancestral *Eudyptes TLR15* was a cryptic pseudogene that has since been overtly disrupted. Our work provides insight into the process of gene function erosion and could form the basis for conservation interventions in an iconic, yet vulnerable group of birds.

Room 9-15

Extensive lineage-specific rediploidisation masks a shared whole genome duplication in the Paddlefish-Sturgeon ancestor

Anthony K Redmond[1,2], Manu Kumar Gundappa[3], Daniel J Macqueen[3], Aoife McLysaght[1]

[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [2]Department of Science and Health, Institute of Technology Carlow, Carlow, Ireland. [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Whole genome duplication (WGD) events are common in eukaryotic evolution, and provide raw genetic material to enhance species diversification and phenotypic novelty, though such outcomes may not arise for millions of years. Evidence from salmonid fish, which experienced an ancestral WGD by autotetraploidization, suggests that returning from a tetraploid to a diploid state may be asynchronous, with different sections of the genome rediploidising millions of years apart, helping to explain the time-lag between WGD and its consequences. Under this model, speciation can occur during the rediploidisation process, allowing genes that were tetraploid at speciation to later resolve into diploid ohnolog pairs independently in each lineage. This has important evolutionary implications, with all functional divergence of such ohnologs being lineage-specific. Importantly, there currently remains little evidence for lineage-specific rediploidization outside salmonids. Here, we reassess WGD history in paddlefish and sturgeons, which are sister lineages believed to have independent WGD events. Contrary to this, over one third of our ohnolog gene trees support a shared ancestral WGD, which cannot be explained by phylogenetic error. Ohnolog trees with the same topology cluster along the genome, forming segments of common rediploidisation history, with large contiguous regions supporting a single ancestral WGD, followed by either ancestral or lineage-specific rediploidization. We conclude that the paddlefish-sturgeon WGD was a shared autotetraploidization, and that most of the genome was still tetraploid when these lineages diverged. These findings indicate that lineage-specific rediploidisation, and its implications for ohnolog functional evolution, may be a general feature after many WGD events.

Room 9-16

Performance of AIC and BIC to select correct models of evolution

Heiko A Schmidt[1], Stephen M Crotty[2], Sebastian Burgstaller-Muehlbacher[1]
[1]CIBIV, University of Vienna & Medical University of Vienna, Vienna, Austria. [2]School of Mathematical Sciences, University of Adelaide, Adelaide, SA, Australia

Abstract

Phylogenetic tree reconstructions are frequently performed in various fields of biological and epidemiological research. Before the actual reconstruction typically a suitable model of evolution has to be determined. This model selection step usually employs information criteria like the Akaike (AIC) or the Bayesian Information Criterion (BIC). These information criteria compensate for different numbers of parameters contributing to the likelihood computation. This allows for the comparison of likelihoods for different models even if they are not nested because or use different parameterizations.

Model selection is typically performed using the whole multiple sequence alignment that is used as input for the tree reconstruction, or on separate partitions in case that combined multi-gene data is used.

Here we study the performance of the AIC and BIC. To that end we performed a large simulation study. We use a number of well-known evolutionary models. Furthermore, we use wide ranges of parameterizations and sequence lengths. We show the different accuracies of the information criteria for estimating the different evolutionary models. Furthermore, we discuss the impact of parameters like amount of rate heterogeneity and sequence length on the model selection performance.

Room 9-17

PopHumanVar: An Interactive App for the Functional Characterization and Prioritization of Genomic Variants

Aina Colomer-Vilaplana, Jesús Murga-Moreno, Aleix Canalda-Baltrons, Clara Inserte, Daniel Soto, Antonio Barbadilla, Sònia Casillas
Universitat Autònoma de Barcelona, Bellaterra, Catalunya, Spain

Abstract

Since the divergence with chimpanzees, and especially when migrating across the globe, our species has faced frequent environmental and social challenges. In response to these adaptive pressures, natural selection has shaped our genomes, leaving signatures that can be revealed from present-day patterns of genetic variation. PopHumanScan (PHS) is a collaborative database amassing 2859 putative human genomic regions underlying natural selection. This catalog encompasses the 22 non-admixed human populations of the 1000 Genomes Project phase 3 (1000GP) and pinpoints signatures of presumed selective processes at different historical ages based on the combination of eight different population genetics metrics.

A further step for interpreting these candidate regions is pinpointing the allele ultimately responsible for this selection. However, few methods are capable of doing so. Here, we present PopHumanVar (PHV), an interactive application that graphically represents different layers of information, including natural selection statistics, as well as functional and age information, for the biallelic positions in 1000GP. It amasses SNP-based information from GEVA, VEP, GWAS Catalogue, ClinVar, RegulomeDB and DisGeNET, as well as accurate estimations of the integrative haplotype score (iHS).

Complementary to PHS, PHV is designed to facilitate the exploration and thorough analysis of candidate genomic regions, generating useful summary reports of prioritized variants that are putatively causal of recent selective sweeps. PHV was built using the open-source programming language R and the package Shiny. It is open and freely available at https://pophumanvar.uab.cat.

Room 9-18

Exploration of tRNA fitness landscape reveals that the wild type allele is sub-optimal and mutationally robust

Tzahi Gabzi[1], Yitzhak Pilpel[1], Tamar Friedlander[2]
[1]Weizmann Institute, Rehovot, Israel. [2]Hebrew University, Rehovot, Israel

Abstract

Fitness landscape mapping and the prediction of evolutionary trajectories on these landscapes are major tasks of evolutionary biology. Evolutionary dynamics is tightly linked to the landscape topography, but this relation is not straightforward. Models predict different evolutionary outcomes depending on mutation rates: high fitness genotypes should evolve under low mutation rates and lower-fitness mutationally robust (flat) genotypes - at higher mutation rates.
Yet, so far flat genotypes have been demonstrated in very few cases. The quantitative conditions for their emergence were studied only in simplified single-locus two-peaks landscapes. In particular, it is unclear whether a combination of fit and flat alleles can co-exist.
Here we analyze a previously measured fitness landscapes of a yeast tRNA gene. We find that the wild type allele is sub-optimal and located in a flat mutational neighborhood.
To link this finding to theory, we construct a generalized multi-locus fitness landscape model, and analyze the conditions for existence of flat alleles. If all loci have equal properties, we find an avalanche of transitions from all-alleles-fit to all-alleles-flat within a narrow range of mutation rates. This threshold mutation rate inversely scales with the number of loci and with allele length. If mutation rates or gene length vary across the genome, a mixture of fit and flat alleles can co-exist in a broad range of mutation rates. As tRNA genes are known to be exceptionally short and highly mutable, we propose that they are flat while the majority of the genome is fit.

Room 9-19

Does selection on codon usage bias affect the rate of protein evolution?

Ana Filipa Moutinho, Adam Eyre-Walker
School of Life Sciences, University of Sussex, Brighton, United Kingdom

Abstract

The genetic code is degenerate, as 64 codons translate into 20 different amino acids. Each amino acid is, therefore, coded by more than one codon. These codons, however, are not equally used. Some are more frequent than others, a phenomenon described as codon usage bias (CUB). Such biased use of codons has been reported across all kingdoms of life, with a greater bias observed in highly expressed genes. This relationship with gene expression has been associated with selection for greater translational efficiency. As such selective pressures delay the rate at which synonymous sites evolve, we hypothesize that it could also slow the rate of protein evolution if two amino acids have different preferred codons. We test this hypothesis by looking at patterns of protein evolution using polymorphism and substitution data in bacteria. We developed two statistics to measure the extent of selection acting on codon usage: *Y*, which is applied at the polymorphism level; and *Z*, which is applied at the substitution level. Positive values of these statistics mean that more mutations go from low to high relative synonymous codon usage (RSCU) (unpreferred to preferred codons). In both analyses, we observed that both *Y* and *Z* are skewed towards positive values, suggesting that, on average, more mutations go from low to high RSCU values. Moreover, we observed a negative correlation between Z and rates of protein substitution. These findings suggest that selection on CUB influences patterns of protein evolution in bacteria.

Room 9-20

Divergent MHC evolution in Neotropical cichlid radiations

Seraina E Bracamonte, Melinda J Hofmann, Carlos Lozano-Martín, Marta Barluenga
Museo Nacional de Ciencias Naturales, Madrid, Spain

Abstract

Ecological speciation was proposed as a major driver for the generation of new species. Divergent natural selection on populations facing heterogeneous habitats leads to distinct phenotypes among them and convergent phenotypes among populations in similar but geographically isolated habitats. Parasites regularly show spatially variable distributions with some degree of habitat specificity. Due to the strong selective pressures that they exert on the host immune system, they have the potential to select for host population divergence that can ultimately result in diversification. The hypervariable vertebrate immunogenes of the major histocompatibility complex (MHC) provide parasite-specific recognition and they were repeatedly shown to co-vary with parasite communities. We study the extent to which MHC diversity in a Neotropical cichlid species complex, the Midas cichlid (*Amphilophus* spp.), may have contributed to diversification. The Midas cichlid, one of the most convincing examples of sympatric parallel speciation, inhabits the Nicaraguan great lakes and several isolated crater lakes in which it has independently radiated into ecologically and morphologically similar species assemblages following colonization. We find species-specific, lake-specific, and habitat-specific signatures of MHC allelic and functional diversity but no clear pattern of parallelism among ecomorphotypes, suggesting that MHC diversity may have been shaped by both ecological factors and the colonization history.

Room 10-01

Unifying theoretical and molecular approaches to non-genetic inheritance

Irene Adrian-Kalchhauser[1], Sonia E Sultan[2], Lisa Shama[3], Helen Spence-Jones[4], Stefano Tiso[5], Claudia Isabelle Keller Valsecchi[6], Franz J. Weissing[5]

[1]University of Bern, Bern, Switzerland. [2]Wesleyan University, Middletown, Conneticut, USA. [3]AWI Sylt, List, Germany. [4]University of Saint Andrews, St Andrews, United Kingdom. [5]University of Groningen, Groningen, Netherlands. [6]IMB Mainz, Mainz, Germany

Abstract

Biologists are currently engaged in a lively conversation about whether it is necessary to expand our view of biological inheritance to include 'non-genetic' factors. In particular, molecular epigenetic mechanisms (such as DNA methylation, histone modifications, and small noncoding RNAs) have been interpreted as additional 'streams' of information distinct from DNA sequence transmission.

However, understanding the evolutionary and ecological roles of 'non-genetic' inheritance is daunting due to the complexity and diversity of the molecular mechanisms involved. Also, despite a shared interest in transgenerational effects in scientific disciplines ranging biomedicine to evolutionary theory, the usage of current terminology (e.g., ‘epigenetics’) differs considerably among fields.

Reviewing the molecular aspects of non-genetic inheritance, we conclude that molecular ‘epigenetic’ mechanisms are complex, functionally diverse, and variable across organisms but at the same time can be summarized and simplified into three general features. These features have implications for empirical and theoretical research. We also conclude that the simplified, yet mechanistically accurate concept of ‘inherited gene regulation’ (IGR) provides an adequate functional description of the vast majority of molecular nongenetic inheritance systems and could help solve terminological cross-field disparities.

Room 10-02

Germline de novo mutation rates on exons versus introns in humans

Miguel Rodriguez Galindo[1], Sònia Casillas[2,3], Donate Weghorn[4,5], Antonio Barbadilla[2,3]
[1]Center for Genomic Regulation, Barcelona, Spain. [2]Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain. [3]Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain. [4]Center for Genomic Regulation, Bellaterra (Cerdanyola del Vallès), Spain. [5]Universitat Pompeu Fabra, Barcelona, Spain

Abstract

A main assumption of molecular population genetics is that genomic mutation rate does not depend on sequence function. Challenging this assumption, a study has found a reduction in the mutation rate in exons compared to introns in somatic cells. This reduction was ascribed to an enhanced exonic mismatch repair system activity. If this reduction happens also in the germline, it can compromise studies of population genomics, including the detection of the footprint of selection when using introns as proxies of neutrality.

We compile and analyze published germline de novo mutation (DNM) data to test if the exonic mutation rate is also reduced in germ cells. We reveal that exonic and intronic DNM densities do not differ between exons and introns, after accounting for trinucleotide sequence composition and an excess of nonsynonymous exonic variation arising from ascertainment bias. We further explore factors that can impact differently DNM densities at exons and adjacent introns, namely, extended sequence context dependency and several chromatin features.

After controlling for several factors, we find no reduction in the mutation rate in exons compared to introns in the germline genome, in contrast to what has been previously described in somatic cells. Therefore, there is no evidence of an enhanced mismatch repair system activity in exons with respect to adjacent introns in germline cells.

Room 10-03

Transcriptional changes analysis unlock the secret of drought resistance in Mediterranean earthworm during aestivation

Natasha Tilikj, Marta Novo
Complutense University of Madrid, Madrid, Spain

Abstract

Aestivation is a form of dormancy employed by the earthworm species Carpetania matritensis when confronted with prolonged periods of drought in the Iberian Peninsula. Since there is limited information on aestivating earthworms, we performed a global transcriptomic comparison of aestivating worms exposed to 5% humidity and control worms kept at conditions of 20% humidity. Our results yielded a total of 6,352 differentially expressed transcripts in the aestivating group, with a total of 65% being downregulated. Results obtained from GO and KEGG pathway enrichment analysis, as well as REVIGO summarization demonstrate a marked suppression of energy requiring biosynthetic processes most notably translation, coupled with an overall reduction of protein turnover. Downregulation of genes involved in the digestive system appears to lead to carbon metabolism suppression. Our findings seem to indicate an increased activity of genes required for DNA repair, even though the expression of apoptotic genes as well as genes coding for antioxidant enzymes is present in varying degrees. All things considered our results suggest a strong physiological shift in C. matritensis when exposed to 5% humidity. Earthworms play an important role in soil quality maintenance, so it is of great importance to understand the mechanism of drought-induced aestivation as a future tool to mitigate the effects of climate change.

Room 10-04

phastSim: efficient simulation of sequence evolution for pandemic-scale datasets

Nicola De Maio[1], Lukas Weilguny[1], Conor Walker[1], Yatish Turakhia[2,3], Russell Corbett-Detig[2,3], Nick Goldman[1], William Boulton[1]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridgeshire, United Kingdom. [2]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California, USA. [3]Genomics Institute, University of California Santa Cruz, Santa Cruz, California, USA

Abstract

Sequence simulators are fundamental tools in bioinformatics, as they allow us to test data processing and inference tools, as well as being part of some inference methods. The ongoing surge in available sequence data is however testing the limits of our bioinformatics software. One example is the large number of SARS-CoV-2 genomes available, which are beyond the processing power of many methods, and simulating such large datasets is also proving difficult. We present a new algorithm and software for efficiently simulating sequence evolution along extremely large trees (e.g. > 100,000 tips) when the branches of the tree are short, as is typical in genomic epidemiology. In these scenarios, our approach is considerably faster than existing simulators (e.g. able to simulate 100 times as many sequences as Indelible, and 10,000 times as many sequences as pyvolve, in the same time).

Our method is based on the Gillespie algorithm, and implements an efficient multi-layered search-tree structure that provides high computational efficiency by taking advantage of the fact that only a small proportion of the genome is likely to mutate in each branch of a phylogeny of closely related organisms. A further advantage of this approach is that indels can be incorporated into the same data structure without sacrificing algorithmic efficiency, by growing and pruning branches on the multi-layered tree structure.

Our open source software allows easy integration with other Python packages as well as a variety of evolutionary models, including new ones that we developed to more realistically model SARS-CoV-2 genome evolution.

Room 10-05

Unraveling the genomic basis of adaptation to life in caves in subterranean beetles

Pau Balart-Garcia[1], Alexandra Cieslak[1], Paula Escuer[2], Julio Rozas[2], Ignacio Ribera[1], Rosa Fernández[1]
[1]Institute of Evolutionary Biology (CSIC - Universitat Pompeu Fabra), Barcelona, Catalonia, Spain.
[2]Universitat de Barcelona, Barcelona, Catalonia, Spain

Abstract

Despite the evolutionary relevance of adaptation to life in caves, little is known about the genetic basis underlying the remarkable phenotypes of animals that conquered subterranean habitats. An ideal system for exploring the evolution of troglobitic characters is the Leptodirini tribe (Leiodidae, Coleoptera), a speciose clade of terrestrial beetles that colonized subterranean environments ca. 30 Mya. To interrogate the genomic underpinnings of adaptation to life in caves, we first generated transcriptomes for antennae and body samples of the cave-dwelling beetle Speonomus longicornis to investigate the chemosensory gene repertoire across Coleoptera through a phylogenomic approach. Our results suggested a diminished diversity of odorant and gustatory gene repertoires compared to polyphagous beetles from surface habitats. Moreover, S. longicornis showed a large diversity of odorant-binding proteins, and no homologous genes to sugar receptors were detected. Furthermore, in order to explore patterns of evolutionary convergence in the process of subterranean specialization, we generated transcriptomes for thirteen surface and cave-dwelling species showing gradual subterranean phenotypes and representing independent underground colonizations. Our findings suggest that the gene repertoire of these cave beetles may result from adaptation to their highly specific ecological niche, and that gene duplication and loss played an important role in the evolution of gene families involved in chemosensation. Altogether, our results shed light on the genomic basis of chemoreception in cave-dwelling beetles and pave the way towards understanding the genomic underpinnings of adaptation to the subterranean lifestyle at a deeper level.

Room 10-06

Aquaporin evolution in the context of arthropod terrestrialization

Gemma I Martínez-Redondo, Pau Balart-García, Rosa Fernández
IBE, Institute of Evolutionary Biology (Universitat Pompeu Fabra -CSIC), Barcelona, Barcelona, Spain

Abstract

The transition from sea to land (also known as terrestrialization) is one of the most remarkable evolutionary events that shaped life as we know it today. Terrestrialization involves many physiological changes in different biological processes, such as breathing or reproduction, to overcome the environmental barriers and adapt to life on land. One of the most important challenges that species must face is water loss, which alters their osmotic and hydric homeostasis. Aquaporins are a superfamily of membrane intrinsic proteins that create channels through which water and other small molecules can traverse the cell membrane, and are heavily involved in osmoregulatory processes. Previous studies have linked the expansion of aquaporins to terrestrial adaptation in tetrapods and other vertebrates. However, the diversity and evolutionary dynamics of aquaporins in most animal groups is still unknown. Arthropods represent an ideal system to explore the potential role of aquaporins as facilitators of terrestrialization, since they colonized the land multiple times independently and at different times in Earth history, with terrestrial arthropods massively outnumbering the aquatic ones. In this study, we interrogate aquaporin repertoire evolution in the main lineages of terrestrial and aquatic arthropods. We first predicted aquaporin sequences from 37 terrestrial and 31 aquatic arthropod genomes and transcriptomes from all main lineages (Myriapoda, Chelicerata and Pancrustacea). Then, we inferred the evolutionary relationships of these proteins to test whether terrestrial and aquatic arthropod aquaporins have different evolutionary histories. We discuss our findings in the context of arthropod terrestrialization.

Room 10-07

Dynamics of dN/dS at short time scales

Evgenia A Belousova[1], Anastasia V Stolyarova[2], Alexey S Kondrashov[1,3], Georgii A Bazykin[2,4]
[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russian Federation. [2]Skolkovo Institute of Science and Technology, Skolkovo, Russian Federation. [3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, USA. [4]Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences, Moscow, Russian Federation

Abstract

The dN/dS ratio is a widely used measure of the strength and direction of selection pressure acting on the evolving sequences. It was originally developed to infer evidence of selection between distantly diverged species, but it is also commonly used for closely related populations. A number of works show that the relationship between the strength of selection, the time passed since the divergence event and dN/dS is complex. Here, we study the dynamics of dN, dS, and dN/dS between two diverging populations under positive selection. Since the rate of fixation of adaptive nonsynonymous mutations is much higher than that of the selectively neutral synonymous mutations, we expect the nonsynonymous differences between populations to be accumulated faster than the synonymous ones over the short time scales after the divergence, leading to the increased dN/dS ratio. Over the long time scales, the fixation of synonymous mutations might decrease dN/dS, converging it to the value expected under the given amount of positive selection. The proposed effect is not expected when comparing pairwise differences between genotypes sampled from the diverged populations instead of the fixed substitutions. In this work, we use simulations and analytical modeling to describe the dynamics of accumulation of adaptive and neutral differences over short and long evolutionary distances.

Room 10-08

Molecular parallelisms between pigmentation in the avian iris and the integument of ectothermic vertebrates

Małgorzata A Gazda[1,2], Pedro Andrade[1], Pedro Araújo[1], Sandra Afonso[1], Jacob A Rasmussen[3], Cristiana I Marques[1], Ricardo J Lopes[1], Thomas Gilbert[3], Miguel Carneiro[1]

[1]CIBIO, Vairao, Portugal. [2]IBENS, Paris, France. [3]Center for Evolutionary Genomics, Copenhagen, Denmark

Abstract

Birds exhibit striking variation in eye color that arises from interactions between specialized pigment cells named chromatophores. The types of chromatophores present in the avian iris are lacking from the integument of birds or mammals, but are remarkably similar to those found in the skin of ectothermic vertebrates. To investigate molecular mechanisms associated with eye coloration in birds, we took advantage of a Mendelian mutation found in domestic pigeons that alters the deposition of yellow pterin pigments in the iris. Using a combination of genome-wide association analysis and linkage information in pedigrees, we mapped variation in eye coloration in pigeons to a small genomic region of ~8.5kb. This interval contained a single gene, SLC2A11B, which has been previously implicated in skin pigmentation and chromatophore differentiation in fish. Loss of yellow pigmentation is likely caused by a point mutation that introduces a premature STOP codon and leads to lower expression of SLC2A11B through nonsense-mediated mRNA decay. There were no substantial changes in overall gene expression profiles between both iris types as well as in genes directly associated with pterin metabolism and/or chromatophore differentiation. Our findings demonstrate that SLC2A11B is required for the expression of pterin-based pigmentation in the avian iris. They further highlight common molecular mechanisms underlying the production of coloration in the iris of birds and skin of ectothermic vertebrates.

Room 10-09

Wild animal oral microbiomes reflect the history of human antibiotics use

Jaelle C Brealey[1,2], Henrique G Leitão[1], Thijs Hofstede[1], Daniela C Kalthoff[3], Katerina Guschanski[1,4]
[1]Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden. [2]NTNU University Museum, Trondheim, Norway. [3]Swedish Museum of Natural History, Stockholm, Sweden. [4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Following the advent of industrial-scale antibiotics production in the 1940s, antimicrobial resistance (AMR) has been on the rise and now poses a major global health threat. Because AMR can be exchanged between humans, livestock and wildlife, evaluating the potential of wild animals to act as AMR reservoirs is essential. However, AMR is a normal function of natural environments, including host-associated microbiomes, which makes it challenging to distinguish between anthropogenic and natural sources. Overcoming this difficulty requires historical samples that span from before the mass production of antibiotics to today. We used shotgun metagenomics sequencing of dental calculus, the calcified form of the oral microbial biofilm, to determine the abundance and repertoire of AMR genes in the oral microbiome of wild Swedish brown bears from museum specimens collected over the last 200 years. Our temporal metagenomics approach allowed us to establish a baseline of natural AMR in the pre-antibiotics era and to quantify a significant increase in total AMR load and diversity of AMR genes that is correlated with the known history of human antibiotics use in Sweden. We also demonstrated that Swedish public health policies were effective in reducing human-associated AMR contamination in wildlife.

Room 10-10

Why are X chromosomes enriched in male-expressed microRNAs?

Antonio Marco
University of Essex, Colchester, United Kingdom

Abstract

Genes are often differentially expressed between males and females. In *Drosophila melanogaster*, the analysis of sex-biased microRNAs (short non-coding regulatory molecules) has revealed striking differences with protein-coding genes. Mainly, the X chromosome is enriched in male-biased microRNA genes, whilst it is depleted of male-biased protein-coding genes. We suggest that this is due to high-rates of de novo emergence of microRNAs in the X-chromosome, and that novel microRNAs tend to be expressed in testis. To test this hypothesis we analysed the expression profile of microRNAs in males, females and various tissues in *D. pseudoobscura*, in which an autosome translocated into the X chromosome, effectively becoming part of a sex chromosome (neo-X). We found that the pattern of sex-biased expression is generally conserved between *D. melanogaster* and *D. pseudoobscura* (except for one locus), also in the neo-X. *D. pseudoobscura*-specifc microRNAs tend to be male-biased and the rate of microRNA emergence in the neo-X chromosome is higher than in the autosomes, supporting our hypothesis. In summary, the apparent paradox resulting from male-biased protein-coding gene depleted in the X chromosome and an enrichment in male-biased microRNAs is a consequence of different, yet well characterized, evolutionary dynamics.

Room 10-11

The Impact of Selection at The Amino Acid Level on the Synonymous Codon Usage in Alternative Genetic Codes

Konrad Wiesław Pawlak, Paweł Błażej, Paweł Mackiewicz
Department of Bioinformatics and Genomics, Faculty of Biotechnology, University of Wrocław, Wroclaw, Poland

Abstract

The aim of the research is to broaden knowledge in the field of genomics by determining the influence of selection at the amino acid level on synonymous codons usage and determining how important this factor is in shaping this phenomenon and linking two forces shaping codon usage in the genome, i.e. mutational and selection pressure. This factor is poorly studied and usually overlooked in genomic research, and may be important in understanding the biological processes that shape genomic codon usability. The pattern of codon usage is significantly influenced by selection against changes at the amino acid level and the associated different probabilities of single nucleotide substitution in the codons. alternative genetic codes for this aspect were also tested in the work.

Room 10-12

Molecular epidemiology of HIV in Oryol district, Russia

Ksenia R Safina[1,2], Georgii Bazykin[1,2], Yulia Sidorina[3], Natalya Efendieva[3], Elena Belonosova[3], Dmitry Kireev[4]

[1]Skolkovo Institute of Science and Technology, Moscow, Russian Federation. [2]The Institute for Information Transmission Problems of Russian Academy of Sciences, Moscow, Russian Federation. [3]Oryol Regional Center for AIDS and Infectious Diseases Control and Prevention, Oryol, Russian Federation. [4]Central Research Institute of Epidemiology, Moscow, Russian Federation

Abstract

Studying AIDS epidemics in Russia is hindered by the lack of genetic data on HIV which is available for less than one percent of the HIV-positive population in the country. We decided to conduct a molecular epidemiology study in one particular geographic region of Russia where a high enough population coverage can be attained. We decided on the Oryol district where ~3,000 HIV-positive people reside, aiming to cover at least a quarter of its HIV-positive population.

Our dataset on the Oryol district currently consists of 725 sequences (the pol region fragment), with 627 samples collected in 2018 or later. 95% and >99% of the samples are accompanied by a patient's suggested transmission route and gender, respectively. Samples were aligned and curated manually; the resulting alignment of 1,244 bp was used to reconstruct phylogeny.

We identified molecular clusters that are expected to be related to transmission clusters and studied how they correspond to various metadata (e.g. gender, transmission route). Our data do not show statistically significant clusterization of male/homosexual samples on the phylogeny. Male/female distribution over HIV subtypes suggests underreporting of homosexual transmission route in our dataset. We also assessed the effectiveness of antiretroviral therapy, identified drug-resistant variants, and tracked their expansion over the population. 6.1% (24/393) of naive patients carry mutations that are predicted to result in drug resistance against NRTI or NNRTI which reflects the population-wide level of drug resistance; for some of them, transmission from a patient receiving therapy could be identified phylogenetically.

Room 10-13

Hidden paralogs and reconstructing the animal tree of life

Charley McCarthy[1], Peter Mulhair[2,1], Karen Siu-Ting[3], Chris Creevey[3], Mary O'Connell[2,1]
[1]University of Nottingham, Nottingham, United Kingdom. [2]University of Leeds, Leeds, United Kingdom. [3]Queen's University Belfast, Belfast, United Kingdom

Abstract

Our understanding of how complex tissues and systems have evolved within animals is contingent upon a clear reconstruction of animal evolutionary history. However, there remain conflicting hypotheses regarding important nodes within the animal tree of life (ATOL), primarily whether *Ctenophora* (comb jellies) or *Porifera* (sponges) constitute the sister group to all other animals. We have investigated the effect of “hidden paralogy” on conflicting ATOL hypotheses, using an approach previously demonstrated to resolve similar conflicts within *Lissamphibia*. This approach assesses whether gene families are solely orthologous or contain paralogs arising from post-speciation gene loss events based on their ability to recapitulate otherwise-uncontentious relationships, e.g. the monophyly of *Bilateria*.

We examined 4 previously-published ATOL datasets which had variously supported a *Ctenophora*-sister or *Porifera*-sister tree. We found that these datasets – regardless of their original tree – predominantly contained gene families which could recapitulate the monophyly of *Ctenophora* but largely failed to do so for other animal groups. Across these datasets, many gene families possessed phylogenetic signal favouring *Ctenophora*-sister but were also incapable of unequivocally rejecting alternative hypotheses. When genes which failed to recapitulate ≥3 major animal groups were removed and datasets were reanalysed under a CAT-GTR model, datasets with greater *Ctenophora* sampling supported a *Porifera*-sister tree. Posterior predictive simulations indicate this approach led to a better fit of the CAT-GTR model to sequence data, regardless of eventual tree topology. Our findings do not definitively root the ATOL, but instead illustrate the importance of careful selection of both orthologs and evolutionary models in animal phylogenomics.

Room 10-14

Inversions and translocations in a highly polymorphic fungus Schizophyllum commune

Asia S. Kamyshnikova[1], Anastasia V. Stolyarova[1], Georgii A. Bazykin[1], Alexey S. Kondrashov[2]
[1]Skolkovo Institute of Science and Technology, Moscow, Russian Federation. [2]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russian Federation

Abstract

Schizophylum commune is a basidiomycete fungus with the highest genetic polymorphism known among all eukaryotic species. Previous studies have focused on single nucleotide polymorphisms and have not covered mutations that affect larger regions of the genome. Such mutations, called genome rearrangements, may play an important role in increasing genetic diversity. In this work, we have focused on studying inversions and translocations.

We analyze whole-genome sequencing data of 55 haploid genomes of S. commune. Since the genomes were assembled to contigs de-novo without regard to the reference genome, we can infer genomic rearrangements by looking for discordant alignments of contigs.

Using the alignment with the reference genome, we have searched for possible rearrangements in the genotypes of two samples, one from the American population and another from Russian.

We have searched for rearrangements within one scaffold (#13) and have found that there are 39 translocations and 81 inversions in the Russian genotype and 13 and 19 in the American (the reference genome belongs to the American population). We have excluded rearrangements with lengths smaller than 100 nucleotides to decrease the number of false-positives.

Future goals are to verify the found rearrangements and to make sure they are not a result of poor assembly or alignment, and to examine the patterns of such events and their possible association with functional sequences or relation to the levels of nucleotide diversity or local recombination rate in a given region. Furthermore, we plan to use Hi-C data to discriminate real rearrangements from possible misalignments.

Room 10-15

Genes involved in damage repair show different evolutionary pressures in long- and short-lived bivalves.

Mariangela Iannello, Giobbe Forni, Giovanni Piccinini, Ran Xu, Jacopo Martelossi, Fabrizio Ghiselli, Liliana Milani
Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

Abstract

Bivalves provide the widest range of lifespans within a metazoan class: while most species live from few years to few decades, others can live over 150 years. The queen of longevity is by far the ocean quahog *Arctica islandica*: with a recorded maximum lifespan of 507 years, this species is the longest-lived non-colonial metazoan known so far. Such extraordinary lifespan disparity makes bivalve molluscs useful models to explore the mechanisms associated with ageing and longevity.

In this work, we analyzed transcriptomic data from 30 bivalves, including both long- and short-lived species. We inferred the ratio of non-synonymous to synonymous substitutions (dN/dS) of ~7,000 orthogroups and investigated those having different selective pressures in the long-lived species compared to short-lived ones. We found that in ~2,000 genes the 4 long-lived species share the same dN/dS, which is different from dN/dS in short-lived species. The functional annotation of genes with such convergent selective pressure in long-lived species shows an enrichment of genes involved in damage repair at DNA, RNA and protein level, response to extra damage accumulation (autophagy and mitophagy) and programmed cell death. In most cases, we found that such genes are under stronger selective constraints in long-lived species, compared to short-lived species. This suggests that a strict control of the damage repair system could represent a shared ground plan for extraordinary longevity in bivalves.

Room 10-16

Ancient Loss of Selenocysteine in a Catalytic Site Spurs Compensatory Mutations and Convergent Adaptation in a Mammalian Selenoprotein

Jasmin Rees[1], Gaurab Sarangi[2], Qing Cheng[3], Elias Arnér[3], Martin Floor[4,5], Jordi Villà-Freixa[4,5], Baldomero Oliva Miguel[6], Sergi Castellano[1,7]

[1]Great Ormond Street Institute of Child Health, University College London, London, United Kingdom. [2]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [3]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [4]Department of Basic Sciences, Universitat Internacional de Catalunya, Barcelona, Spain. [5]Department of Biosciences, Universitat de Vic - Universitat Central de Catalunya, Barcelona, Spain. [6]Department of Health and Experimental Sciences, Universitat Pompeu Fabra, Barcelona, Spain. [7]UCL Genomics, University College London, London, United Kingdom

Abstract

Selenocysteine is the 21st amino acid. It is analogous to cysteine, with cysteine having a sulfur-containing thiol group and selenocysteine having a selenium-containing selenol group. Substitutions between selenocysteine and cysteine in orthologous proteins are rare and under constraint, suggesting a limited level of functional exchangeability between these amino acids. This has been taken as evidence for a unique role of selenium in vertebrate evolution, supported by the reduction of catalytic activity that often follows the replacement of selenocysteine by cysteine in proteins. Despite this, some orthologous proteins of vertebrates have lost selenocysteine in place of cysteine. We hypothesised that compensatory changes may follow this amino acid exchange to maintain catalytic activity. Our research focused on GPx6, a selenoenzyme that has independently exchanged selenocysteine for cysteine in multiple mammalian lineages. We experimentally reconstructed past evolutionary changes of GPx6 over the lineage leading to *Eumuroida* and show that the immediate loss of selenocysteine, and corresponding catalytic activity, is followed by adaptive amino acid changes in the catalytic domain. By computationally estimating the activation free energy of the active sites of the derived and ancestral proteins, we suggest that the amino acid changes play a collective, compensatory role in restoring catalysis. Further, we show that many of these changes are repeated across lineages and propose convergent evolution at the protein level. We demonstrate that selenoenzymes are able to recover their catalytic activity following the loss of selenocysteine but the evolutionary and fitness path of proteins to do so is likely narrow.

Room 10-17

Evolutionary genomic analysis of chemoreceptor gene families in chelicerate genomes: Lessons from the chromosome-level reference assembly of the spider *Dysdera silvatica*

Paula Escuer[1], Vadim A Pisarenco[2], Angel A Fernández-Ruiz[1], Joel Vizueta[3,1], Jose F Sánchez-Herrero[4], Miquel A Arnedo[1], Alejandro Sánchez-Gracia[1], Julio Rozas[1]

[1]Universitat de Barcelona (UB), Barcelona, Spain. [2]Universitat de Barcelona (UB), Barcelona, Catalonia, Spain. [3]University of Copenhagen, Copenhagen, Denmark. [4]Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol, Barcelona, Spain

Abstract

Dysdera silvatica (Arachnida, Araneae) is a nocturnal ground-dwelling spider endemic from the Canary Islands. The genus *Dysdera* has undergone a remarkable diversification mostly associated with shifts in the level of trophic specialization, becoming an excellent model to study the genomic drivers of adaptive radiations. We generated a chromosome-level assembly based on the Hi-C scaffolding technique (1.37 Gb; scaffold N50 of 174.2 Mb). The largest scaffolds (or pseudochromosomes), cover 87% of the total assembly size, and match consistently with the seven chromosomes reported in the karyotype of this species.

We performed a comprehensive genomic analysis of the two main Arthropod chemoreceptor gene families, those encoding gustatory and ionotropic receptors (1). We identified 545 members, with a notable underrepresentation in the X chromosome, and found very recent gene duplication bursts. Furthermore, we found that 44% of these receptors are localized in 83 genome clusters. To estimate the genetic differentiation levels of intra and inter-clustering we defined a new gene clustering index (GCI). We obtained high and highly significant GCI levels across pseudochromosomes in both gene families, ranging from 0.418 to 0.982. Globally our results indicate a very recent origin of many chemoreceptors and points to the unequal crossing-over as the main mechanism of their origin. Our reference assembly represents a new valuable resource to gain insights into the structure, organization, and evolution of chelicerate genomes, including the role that structural variants and large gene families played in the extraordinary biology of spiders.

1. Vizueta *et a*l. (2020). *Mol. Biol. Evol.* 37, 3601-3615.

Room 10-18

Gene family changes associated with the recurrent emergence of the wood-boring habit in distantly related lineages of beetles

Shelby Landa, Jingjia Li, Claudio Casola
Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA

Abstract

Wood-boring beetles (WBBs) spend most of their life-cycle in tunnels dug in lignified plant tissues and play essential ecological roles in forest habitats. The unique combination of selective pressures experienced by WBBs led to the emergence of a constellation of traits related to xylophagy, resistance to host defenses, symbiosis with specific bacteria and yeasts, and larval morphological specializations. Because multiple lineages of Coleoptera have convergently adapted to the wood-boring habit, we sought to test if distantly related WBBs share similar patterns of gene family evolution. To this aim, we reconstructed gene family evolution dynamics in nine beetles with fully sequenced genomes, including 4 WBBs from three distant clades, and assessed if the wood-boring habit in beetles is associated with specific gene duplications and losses. We found that WBBs experienced the highest rate of shared gene family expansions compared to other beetles. In total, 49 gene families showed accelerated expansions in WBBs, including 11 families with no more than 1 gene in other beetles and 3 non-beetle insects. These 11 families are associated with digestive tract morphogenesis, regulation of immune response, negative regulation of response to stimulus and amino acid biosynthesis. We also identified 85 gene families with decreased size in WBBs, including ten families that have been lost in the four wood-boring species and showed functional enrichment for serine-type endopeptidase activity, intracellular cholesterol transport and chitin metabolic process. These results pointed to candidate genes and gene families that might be implicated in the adaptation to a wood-boring habit in beetles.

Room 10-19

Coalescent tree recording with selection for fast forward-in-time simulations.

Remi Matthey-Doret
University of Bern, Bern, Switzerland

Abstract

Forward simulations are increasingly important in evolutionary genetics to simulate selection with realistic demography, mating systems and ecology. To reach the performance needed for genome-wide simulations a number of new simulation techniques have been developed recently. Kelleher et al. (2018) introduced a technique consisting in recording the entire genetic history of the population and placing mutations on the coalescent tree. Such method, implemented in both SLiM and SimBit, can of course only be used for neutral loci. I recently introduced a simulation technique that speed up fitness calculation by assuming that fitness effects among haplotypes are multiplicative (Matthey-Doret, 2021). More precisely, fitness measures are stored for subsets of the genome and, at time of reproduction, if no recombination happen within a given subset, then the fitness for this subset for the offspring haplotype is directly inferred from the parental haplotype. Here, I present a hybrid of the above two techniques. The algorithm records the genetic history of a species, directly places the mutations on the tree and infers fitness of subsets of the genome from parental haplotypes. At recombinant sites, the algorithm explores the tree to reconstruct the genetic data at the recombining segment. I benchmarked this new technique implemented in SimBit and report it outperforms all previous techniques to simulate selection, showing a particularly drastic advantage at low recombination rate. Such developments of new simulation techniques are pushing the horizon of the realism with which we can simulate species molecular evolution.

Room 10-20

Bacterial niche adaptation improves plasmid maintenance

Julia Kloos, João A. Gama, Joachim Hegstad, Ørjan Samuelsen, Pål J. Johnsen
Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Abstract

Plasmid spread is an important driver of multidrug resistance evolution in bacteria, and selection for reduced plasmid-imposed fitness costs represents a key mechanism for their maintenance.

We studied the co-evolution of two clinical multidrug resistant plasmids in a uropathogenic *E. coli* host. After 300 generations, genetic adaptation in the chromosomal CCR and ArcAB regulatory systems increased fitness of the evolvedstrains in their laboratory environment, independent from plasmid presence. In addition, the adapted hosts ameliorated the fitness burden of evolved as well as ancestral plasmids pleiotropically. The use of single-gene knock-out strains revealed that global transcription regulation via CCR but not ArcAB plays a specific role in plasmid-cost reduction. An increase in intracellular cAMP concentration confirmed that CCR-associated mutations impacted cAMP-dependent gene regulation. Finally, transcriptional down-regulation of plasmid genes was identified as the possible mechanism of plasmid cost mitigation.

Taken together, our results represent a new solution to the plasmid paradox, where plasmid hitch-hiking on bacterial niche adaptation likely promotes the long-term relationship of newly arising plasmid-host pairs.

Room 11-01

Exploring a local genetic interaction network using evolutionary replay experiments

Ryan C Vignogna, Sean W Buskirk, Gregory I Lang
Lehigh University, Bethlehem, PA, USA

Abstract

Understanding how genes interact is a central challenge in biology. Experimental evolution provides a useful, but underutilized, tool for identifying genetic interactions, particularly those that involve non-loss-of-function mutations or mutations in essential genes. We previously identified a strong positive genetic interaction between specific mutations in *KEL1* (P344T) and *HSL7* (A695fs) that arose in an experimentally-evolved *Saccharomyces cerevisiae* population. Because this genetic interaction is not phenocopied by gene deletion, it was previously unknown. Using “evolutionary replay” experiments we identified additional mutations that have positive genetic interactions with the *kel1*-P344T mutation. We replayed the evolution of this population 672 times from six timepoints. We identified 30 populations where the *kel1*-P344T mutation reached high frequency. We performed whole-genome sequencing on these populations to identify genes in which mutations arose specifically in the *kel1*-P344T background. We reconstructed mutations in the ancestral and *kel1*-P344T backgrounds to validate positive genetic interactions. We identify several genetic interactors with *KEL1*, we validate these interactions by reconstruction experiments, and we show these interactions are not recapitulated by loss-of-function mutations. Our results demonstrate the power of experimental evolution to identify genetic interactions that are positive, allele specific, and not readily detected by other methods, shedding light on an under-explored region of the yeast genetic interaction network.

Room 11-02

Detecting genetic interactions in a large bacterial dataset

Rohan S Mehta, Robert A Petit III, Timothy D Read, Daniel B Weissman
Emory University, Atlanta, GA, USA

Abstract

Interactions between genes are a major part of evolution, but they are fundamentally difficult to study due to problems of scale. In the context of bacterial evolution, the widespread presence of horizontal gene transfer emphasizes the importance of gene-gene interactions. Detecting gene-gene interactions without performing large numbers of assays requires the development of computational techniques that can handle the necessary volume of genomic data Here we present such a method, focusing on recent evolutionary events. We apply our method to a database of over 40,000 genomes of *S. aureus*. We demonstrate that our method can suggest genes for future study and that our method illuminates the current discourse on bacterial genomics.

Room 11-03

Binding affinity landscapes constrain the acquisition of breadth in anti-influenza antibodies

Angela M Phillips[1], Katherine R Lawrence[1,2], Alief Moulana[1], Thomas Dupic[1], Jeffrey Chang[1], Aleksandra M Walczak[3], Thierry Mora[3], Michael M Desai[1]
[1]Harvard University, Cambridge, MA, USA. [2]MIT, Cambridge, MA, USA. [3]CNRS, Paris, France

Abstract

The adaptive immune system responds to new pathogens by mutating antibody-encoding genes and selecting for antibody variants that bind the pathogen of interest. The rapid evolution of pathogens like influenza makes it difficult to produce antibodies that are broadly protective against diverse strains, though on occasion, broadly neutralizing antibodies (bnAbs) are produced. BnAbs often have many more mutations than strain-specific antibodies. The extent to which these mutations interact non-additively or bestow trade-offs for binding different antigens remains unknown and is critical for understanding how to elicit protective antibodies. Herein, we implement a high-throughput method for measuring bnAb binding affinities to determine *(1)* the impact of mutations on binding to diverse antigens and *(2)* whether interactions between mutations make bnAb evolution path-dependent. To this end, we constructed antibody libraries containing all possible combinations of mutations in two anti-influenza bnAbs of varying breadth. Measurement of the equilibrium binding constants of these 2^{16} and 2^{11} variants revealed that the acquisition of affinity to diverse antigens likely occurred sequentially, rather than simultaneously, as affinity to increasingly divergent antigens requires progressively more somatic mutations. This finding has important implications for the design of vaccination regimens, suggesting that bnAbs may be more frequently elicited in response to successive exposures to distinct antigens.

Room 11-04

A free-living protist with non-canonical DNA replication and segregation systems

Dayana E. Salas-Leiva[1], Eelco C. Tromer[2,3,4,5], Bruce A. Curtis[1], Jon Jerlström-Hultqvist[1], Martin Kolisko[6], Zhenzhen Yi[7], Joan S. Salas-Leiva[8], Lucie Gallot-Lavallée[1], Geert J. P. L. Kops[3,4,5], John M. Archibald[1], Alastair G. B. Simpson[1], Andrew J. Roger[1]

[1]Dalhousie University, Halifax, Nova Scotia, Canada. [2]University of Cambridge, Cambridge, United Kingdom. [3]Oncode Institute, Utrech, Netherlands. [4]Hubrecht Institute–KNAW, Utrech, Netherlands. [5]University Medical Centre Utrecht, Utrech, Netherlands. [6]Institute of Parasitology Biology Centre, České Budějovice, Czech Republic. [7]South China Normal University, Guangzhou, China. [8]CONACyT-Centro de Investigación en Materiales Avanzados, Chihuahua, Mexico

Abstract

All cells must replicate and segregate their DNA with precision. These processes are part of regulated cell-cycle mechanisms that begin at S-phase with the replication of DNA and end after M-phase, when replicated chromosomes are segregated into daughter cells. To gain insights into the diversity of these systems in eukaryotes, we carried out a comparative genomics analysis of the highly divergent eukaryote supergroup Metamonada. There is an over-representation of genomes of parasites in the supergroup, thus, we generated a draft genome for the free-living flagellate *Carpediemonas membranifera*. Our analyses reveal extensive streamlining of the DNA replication and segregation machineries within Metamonada. Surprisingly, Carpediemonas species are further streamlined as they lack all proteins of the origin recognition (ORC) complex and Cdc6, several subunits of GINS and of polymerases delta and epsilon, as well as most structural kinetochore subunits, a microtubule plus-end tracking complex and the Ndc80 complex. In addition, the presence-absence patterns of the orthologs involved in double strand break repair in Metamonada point to the existence of a highly specialized homologous recombination pathway. *Carpediemonas* is the first known eukaryote to have such drastically altered DNA replication and segregation systems. Overall, our analyses indicate the existence of underlying and as-yet undescribed mechanism that can accomplish replication initiation and licensing in eukaryotes. We propose that *Carpediemonas* may employ an origin-independent DNA replication mechanism based on Dmc1-dependent homologous recombination.

Room 11-05

The Rad9-Rad1-Hus1 complex is found in Microsporidia

Anne Caroline Mascarenhas dos Santos, Xuan Fang, Oscar Juárez, Jean-François Pombert
Illinois Institute of Technology, Chicago, IL, USA

Abstract

Microsporidia are genetically diverse intracellular parasites, some of which underwent severe genome reduction throughout their life histories. The most extreme cases are found in the Encephalitozoonidae, with species harboring genomes as small as 2.3 Mbp and proteomes restricted to roughly 2,000 proteins. The sequences left encoded in these tiny genomes are highly divergent, which hinders functional predictions, such that half of the *Encephalitozoon* proteome has no known function. The high levels of sequence divergence observed in microsporidians have been postulated to originate from the accumulation of mutations in their genomes due to the lack of DNA repair proteins, including the Rad9-Rad1-Hus1 DNA repair clamp (9-1-1 complex). Using a new computational approach combining 3D structure prediction with structural homology searches in the tridimensional plane, we identified the 9-1-1 complex in Microsporidia together with other previously-thought missing components from the DNA damage checkpoint signaling pathway. To validate these computational predictions, each subunit of the 9-1-1 complex was cloned and expressed in *E. coli*, identified by Western-Blot, and purified by affinity (Ni-NTA), ion-exchange (DEAE) and size exclusion chromatographies for upcoming crystallization experiments.

Room 11-06

Microbial genomic trait evolution is dominated by frequent and rare pulsed evolution

Yingnan Gao, Martin Wu
University of Virginia, Charlottesville, VA, USA

Abstract

On the macroevolutionary timescale, does trait evolution proceed gradually or by rapid bursts (pulses) separated by long periods of stasis? Although studies have shown pulsed evolution is prevalent in animals, our knowledge about the tempo and mode of evolution across the tree of life is very limited. This long-standing debate calls for a test in bacteria and archaea, the most ancient and diverse forms of life with unique population genetic properties.

Using a likelihood-based framework, we analyzed patterns of microbial genomic trait evolution in 6,668 bacterial and 263 archaeal genomes that represent a broad range of macroevolutionary timescales. Here we show that pulsed evolution is both prevalent and predominant in microbes. For the first time, we detected two distinct types of pulsed evolution that are predicted by the punctuated equilibrium and quantum evolution theories. The small frequent jumps in bacteria (0.06-0.32 jumps per lineage per Myr) closely match the estimated bacterial speciation rate (0.03-0.05 speciation per lineage per Myr). The rare large jumps are correlated with the cladogenesis of major bacterial lineages.

In conclusion, our results suggest that microbial genomic traits evolve in episodes of jumps followed by long periods of stasis and the origination of major bacterial lineages happens in quick bursts instead of through slow divergence of species over time. In terms of the tempo and mode of trait evolution, pulsed evolution is the rule rather than the exception across the tree of life, despite the drastically different population genetic properties of animals, plants and microbes.

Room 11-07

Vertebrate co-opting of viral calciomics

Stefen A Laukien[1], Frank H Laukien[2]
[1]Independent Researcher, Boston, Massachusetts, USA. [2]Bruker Corp., Billerica, Massachusetts, USA

Abstract

As we examine and learn more about the virome, it has become increasingly evident that retroviruses have played a significant role in the evolution of vertebrates. Syncytin is an important cell-cell fusion protein in placental development, which appears to be nearly identical to a viral protein called env, which causes the virus to fuse with its host cell. Retroviruses can alter calcium ion homeostasis and utilize Ca2+ and cellular Ca2+ binding proteins to their benefit. Acting as an intracellular messenger, Ca2+ plays a key role during the various stages of differentiation in stem cells. Calcium, as an attractor in progenitor cells and stem cells, drives the differentiation of specialized neuronal cells, and the concentration of free intracellular calcium has also been shown to initiate osteo-differentiation of mesenchymal stem cells. In Xenopus, the release of calcium ions produced by ryanodine receptors directs differentiation of somite cells, playing a novel developmental role during myofibrillogenesis. Viruses can bond calcium ions to their protein shells to increase rigidity and resilience, while also using calcium ions to modulate capsid mechanics. We propose that in a similar manner to which vertebrates utilize syncytin to bind the placenta to the uterus, vertebrates have also co-opted the binding of calcium ions in combination with the viral ability to alter intercellular calcium concentration to create the first bones. This may better explain the relatively close or near simultaneous evolutionary emergence of the jaw, of teeth composed of both dentin and bone, and of the pelvic girdle.

Room 11-08

Expression and fusogenic function of a CfERV lineage in canines

Abigail S Jarosz[1], Erica Cech[1], Lindsey Davis[1], Amanda L Pendleton[2], Jaime Modiano[3], Julia V Halo[1]
[1]Bowling Green State University, Bowling Green, Ohio, USA. [2]Purdue University, West Lafayette, IN, USA. [3]University of Minnesota, Saint Paul, Minnesota, USA

Abstract

Despite their being putatively challenged, retroviral infection has not been observed in any contemporary canid. Canid ancestors were infected by retroviruses as reflected by the presence of endogenous retroviruses (ERVs) in the domestic dog genome at a substantially lower presence when compared to other mammals, at just 0.15% of their nuclear genome. One lineage, ERV-Fc1(a), belongs to a canid-specific recombinant virus that was spread to canine ancestors by interspecies transmission and includes the youngest ERVs found in canids with many polymorphic integrations. Our analysis of ERV polymorphism and sequence variation indicates multiple circulating viruses infected canid ancestors within the 20 million years culminating in a recent burst of germline invasion in the common ancestors of wolves and dogs. Further phylogenetic analysis confirmed that this lineage contains a gag and pol gene most closely related to HERV-Fc and an env gene to HERV-W. Both lineages have elements that are associated with several human diseases. Although circulating XRVs remain elusive to canids, there have been reports of gammaretrovirus-like particles and enzyme activities in canine leukemias and lymphomas. We have identified a transcriptionally active ERV group in canine tissues. Sequence analysis of expressed env transcripts indicates all cluster with the young CfERV-Fc1(a) group. Remarkably, we also observe the majority of Fc1(a) env genes have open reading frames and contain all the necessary motifs to maintain apparent viral function. We demonstrate some of these polymorphic insertions have retained their fusogenic ability, and were likely involved in ancestral, and putatively ongoing spread.

Room 11-09

The Role of Sensory Receptor Genes in Drosophila Suzukii's Preference for Ripe Fruit

Xin Yu Zhu Jiang[1,2], Alice Gadau[2], Sylvia Durkin[2], Nicolas Svetec[2], Li Zhao[2]
[1]Hunter College, New York, NY, USA. [2]Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY, USA

Abstract

Unlike most *Drosophilids*, *Drosophila suzukii* has evolved a preference to oviposit on ripe fruit instead of rotten fruit. This shift in preference is facilitated by an enlarged ovipositor that can puncture ripe fruit. *D. subpulchrella*, its sister species, also has an enlarged ovipositor, but has an intermediate preference for ripe fruit. *D.biarmipes* also has an intermediate preference for ripe fruit, but without an enlarged ovipositor, it is unable to puncture the skin of ripe fruits. Previously, changes in preference have been attributed to changes in the sensory and central nervous system. Therefore, changes in sensory receptors may be a major factor in *D.suzukii*'s preference for laying eggs on ripe fruit. To determine key receptor gene changes, we created RNA-seq libraries of the ovipositor of *D. suzukii, D. subpulchrella,* and *D. biarmipes*. I aligned these RNA-seq libraries with Hisat2, quantified them with Stringtie, and analyzed them using DESeq2. We found 18 sensory receptors significantly differentially expressed between *D.biarmipes* and *D. subpulchrella*, and 24 significantly receptor genes expressed between *D.suzukii* and *D. subpulchrella*, and 24 significantly receptor genes expressed between *D. suzukii* and *D. biarmipes*. From the four types of receptor (olfactory, ionotropic, mechanosensory, gustatory) genes, mechanosensory receptor genes had the greatest number of significantly differentially expressed genes in the ovipositor, indicating a possibly significant role in *D.suzukii*'s preference for ripe fruit.

Room 11-10

The myth of "the" mutation rate: A lack of correlation across rates for different types of mutations among genotypes within a species challenges most theories for how mutation rates evolve

Sarah Schaack, Eddie K. H. Ho
Reed College, Portland, Oregon, USA

Abstract

We present direct estimates of mutation rates and spectra for an array of mutation types from multiple genotypes from multiple populations within *Daphnia magna*, an aquatic microcrustacean. Following up on our work showing base substitution rates vary among genotypes by 1 and 3 orders of magnitude, respectively, across nine genotypes surveyed, we now show rate estimates for all types of mutation quantified so far: base subs, microsatellites, transposable element gains and losses, indels, and duplications. Our data illustrate the large amount of variance in this key trait, and further show a lack of correlation among mutation rate estimates across types among different genotypes. These results pose a challenge to most prevailing theories aimed at explaining how mutation rates evolve over large time scales.

Room 11-11

Janus-Faced Impact of ATP1B4 gene co-option on evolution of mammals

Nikolai Modyanov, David Kennedy, Steven Haller
University of Toledo College of Medicine and Life Sciences, Toledo, Ohio, USA

Abstract

Orthologous ATP1B4 genes, encoding the Na,K-ATPase BetaM subunit in lower vertebrates, were co-opted for new functions after split between marsupials and placental mammals but before the radiation of placental mammals. Through insertion of two extended Glu-rich clusters in the N-terminal domain, eutherian BetaM completely lost its ancestral role and became the skeletal and atrial cardiac muscle-specific protein of the inner nuclear membrane, which is strongly expressed during perinatal development and is implicated in the regulation of gene expression.
Ablation of the X-chromosome *Atp1b4* gene in mice results in high mortality of knockout neonates. Transcriptome analysis of skeletal muscle from neonatal wild type and knockout littermates revealed broad changes in expression of muscle-specific genes as well as genes regulating lipid metabolism and thermoregulation. These results imply that evolutionarily acquired functions of BetaM are physiologically essential, and even may be necessary for survival of placental mammals in natural conditions providing an evolutionary advantage.
On the other hand, *Atp1b4* disruption rather unexpectedly has caused profound beneficial effects on metabolic parameters of adult mice. Atp1b4 knockout males, which survived to adulthood, have significantly lower body fat, exhibit enhanced metabolic rate and insulin sensitivity, and are resistant to high-fat diet-induced obesity. These data indicate that BetaM has an essential role in development of pathways regulating metabolism of adult placental mammals including their predisposition to obesity. Finally, ablation of the evolutionarily acquired functions of the *Atp1b4* gene is in fact a simulation of the alternative pathway of mammalian evolution.

Room 11-12

Origin of complexity in hemoglobin evolution

Arvind S Pillai[1], Shane A Chandler[2], Anthony Signore[3], Yang Liu[4], Carlos Romero[1], Jay F Storz[3], Arthur C Laganowsky[4], Georg Hochberg[5], Joseph W Thornton[1]

[1]University of Chicago, Chicago, Illinois, USA. [2]Oxford university, Oxford, United Kingdom. [3]University of Nebraska, Lincoln, Nebraska, USA. [4]Texas A&M University, College station, Texas, USA. [5]Max Planck Institute for Terrestrial Microbiology, Marburg, Hesse, Germany

Abstract

Multimeric protein complexes are involved in virtually all cellular processes, but we have no detailed knowledge of how such multimers and their functions arose during historical evolution. Here we use ancestral protein reconstruction and biophysical assays to dissect the evolutionary origins of vertebrate hemoglobin (Hb), a heterotetramer of paralogous α and β subunits, which mediates oxygen transport by binding cooperatively to oxygen. We show that modern hemoglobin evolved from an ancient globin monomer and characterize the historical ‘missing link’ through which the modern tetramer evolved—a non-cooperative dimer with high oxygen affinity - that existed prior to the gene duplication that generated distinct α- and β-subunits. We establish that a single historical substitution at this dimer’s protein surface was sufficient to confer tetramerization. Acquisition of this quaternary association dramatically alters the oxygen-binding function and confers cooperativity, indicating that the ancient active site was already functionally linked to the surface region that later became a protein-protein interface. These observations reveal that evolution can produce new multimeric complexes and yield new functional properties via simple genetic mechanisms that recruit existing biophysical features into higher-level architectures.

Room 11-13

Small-scale population structure of a hatchery-impacted Coho salmon population in British Columbia, Canada

James K Bull[1], Matthew P Josephson[1], Sean M Rogers[1,2]
[1]University of Calgary, Calgary, Alberta, Canada. [2]Bamfield Marine Sciences Centre, Bamfield, British Columbia, Canada

Abstract

Understanding genetic population structure provides valuable insight into patterns of migration and gene flow, particularly in managed or disturbed wild populations. For Pacific salmon species, the spatial scale of genetic differentiation has, additionally, important implications for production, fisheries, and conservation. Although broad-scale patterns of differentiation are well-established in these species, the exact spatial scale over which differentiation occurs remains unclear. The operation of hatcheries for Pacific salmon provides further complication due to potential genetic differences between hatchery-born and wild-born fish even under management paradigms that attempt to minimize such difference. To determine if spatial-genetic structure is possible at small-scales (< 10s of kilometres), and if the operation of a hatchery may contribute to structuring, we sampled 138 returning adult Coho salmon (*Oncorhynchus kisutch*) from across tributaries draining into the Nitinat River on Vancouver Island, Canada. Additionally, we sampled 114 returning adults that swam into an integrated-broodstock hatchery operating on the Nitinat River and who were used as broodstock for hatchery production. Each sample was sequenced using a combination of RADseq and Parental Based Tagging (PBT) approaches and spatial-genetic structure detected using standard approaches.

Room 11-14

The genome-wide rate and spectrum of EMS-induced heritable mutations in the microcrustacean Daphnia: on the prospect of forward genetics

Marelize Snyman, Trung Huynh, Matthew T Smith, Sen Xu
University of Texas at Arlington, Arlington, Texas, USA

Abstract

Forward genetic screening using the alkylating mutagen ethyl methanesulfonate (EMS) is an effective method for identifying phenotypic mutants of interest, which can be further genetically dissected to pinpoint the causal genetic mutations. An accurate estimate of the rate of EMS-induced heritable mutations is fundamental for determining the mutant sample size of a screening experiment that aims to saturate all the genes in a genome with mutations. This study examines the genome-wide EMS-induced heritable base-substitutions in three species of the freshwater microcrustacean *Daphnia* to help guide screening experiments. Our results show that the 10mM EMS treatment induces base substitutions at an average rate of 1.17×10^{-6}/site/generation across the three species, whereas a significantly higher average mutation rate of 1.75×10^{-6} occurs at 25mM. The mutation spectrum of EMS-induced base substitutions at both concentration is dominated by G:C to A:T transitions. Furthermore, we find that female *Daphnia* exposed to EMS (F_0 individuals) can asexually produce unique mutant offspring (F_1) for at least 3 consecutive broods, suggestive of multiple broods as F_1 mutants. Lastly, we estimate that about 750 F_1s are needed for all genes in the *Daphnia* genome to be mutated at least once with a 95% probability. We also recommend 4-5 F_2s should be collected from each F_1 mutant through sibling crossing so that all induced mutations could appear in the homozygous state in the F_2 population at 70-80% probability.

Room 11-15

Uncovering the Genetic Basis of Bull Eye Color in Domestic Pigeons

Bridget Phillips, Emily Maclary, Michael D. Shapiro
University of Utah, Salt Lake City, Utah, USA

Abstract

Variation in coloration and pigment patterning is widespread among vertebrate species. Studies of pigment variation often focus on hair, skin, and feathers, but pigmentation also varies in non-epidermal tissues, including the iris of the eye. Like epidermal coloration, variation in iris color may play important roles in crypsis and communication. Despite the importance of iris coloration, the genetic changes that lead to iris pigmentation diversity in non-mammalian species are still largely unknown. Here, we use the domestic pigeon (*Columbia livia*) as a model to understand the genetic architecture of iris pigmentation.

Pigeons have three main iris colors: orange, pearl (white), and bull (dark brown). Bull is primarily found in birds with white or piebald plumage, suggesting a link between feather color and eye color. To determine the genetic basis of bull eye color, we used quantitative trait locus (QTL) mapping in two F2 laboratory crosses to identify a single locus that is associated with bull eye color. Our locus contains an interesting candidate gene, *Ednrb2*, which is part of a signaling pathway known to effect feather, skin, and hair pigmentation across species. While QTL mapping identified a single locus, comparative genomic approaches across many breeds did not identify any mutations or haplotypes associated with bull eye, which suggests that there may be multiple breed-specific mutations associated with this trait. Through analyzing the genetics of eye color, we can better understand links between plumage and iris color and discover how evolutionarily conserved pathways control pigment production and deposition across species.

Room 11-16

On the origin of frameshift-robustness of the standard genetic code

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

Abstract

The standard genetic code (SGC) has been extensively analyzed for the biological ramifications of its nonrandom structure. For instance, mismatch errors due to point mutation or mistranslation have an overall smaller effect on the amino acid polar requirement under the SGC than under random genetic codes (RGCs). A similar observation was recently made for frameshift errors, prompting the assertion that the SGC has been shaped by natural selection for frameshift-robustness—conservation of certain amino acid properties upon a frameshift mutation or translational frameshift. However, frameshift-robustness confers no benefit because frameshifts usually create premature stop codons that cause nonsense-mediated mRNA decay or production of nonfunctional truncated proteins. We here propose that the frameshift-robustness of the SGC is a byproduct of its mismatch-robustness. Of 564 amino acid properties considered, the SGC exhibits mismatch-robustness in 93–133 properties depending on the mismatch pattern and frameshift-robustness in 55 properties, respectively, and that the latter is largely a subset of the former. For each of the 564 real and 564 randomly constructed fake properties of amino acids, there is a positive correlation between mismatch-robustness and frameshift-robustness across one million RGCs; this correlation arises because most amino acid changes resulting from a frameshift are also achievable by a mismatch error. Importantly, the SGC does not show significantly higher frameshift-robustness in any of the 55 properties than RGCs of comparable mismatch-robustness. These findings support that the frameshift-robustness of the SGC need not originate through direct selection and can instead be a site effect of its mismatch-robustness.

Room 11-17

Unbiased inference of the fitness landscape ruggedness from imprecise fitness estimates

Siliang Song, Jianzhi Zhang
University of Michigan, Ann Arbor, MI, USA

Abstract

Fitness landscapes map genotypes to their corresponding fitness under given environments and allow explaining and predicting evolutionary trajectories. Of particular interest is the landscape ruggedness or the unevenness of the landscape, because it impacts many aspects of evolution such as the likelihood that a population is trapped in a local fitness peak. Although the ruggedness has been inferred from a number of empirically mapped fitness landscapes, it is unclear to what extent this inference is affected by fitness estimation error, which is inevitable in the experimental determination of fitness landscapes. Here we address this question by simulating fitness landscapes under various theoretical models, with or without fitness estimation error. We find that all measures of landscape ruggedness are overestimated due to imprecise fitness quantification, but different measures are affected to different degrees. We devise a method to use replicate fitness measures to correct this bias and show that our method performs well under realistic conditions. We conclude that previously reported fitness landscape ruggedness is likely upward biased owing to the negligence of fitness estimation error and advise that future fitness landscape mapping should include at least three biological replicates in fitness estimation to permit an unbiased inference of the ruggedness.

Room 11-18

Genomics of cycads’ coralloid-root bacterial microbiome suggests adaptation from bacterial symbionts allowing holobiont to thrive in contrasting environments

Diego Garfias Gallegos, Angélica Cibrián Jaramillo, Francisco Barona Gómez
Cinvestav-Langebio, Irapuato, Guanajuato, Mexico

Abstract

The relationships between eukaryotes and prokaryotes have played a pivotal role in the evolution of macrobes. Cycads are an ancient lineage of gymnosperms distributed in a myriad of environments worldwide. These plants develop a special organ called coralloid-root where symbiotic bacteria thrive and which is considered a key trait for its survival. This bacterial microbiome can perform diverse processes for the plant by diverse metabolic functions encoded in their genomes as biosynthetic gene clusters (BGCs). BGC diversity is related to chemical-structural variations of their molecular products, resulting in different biological activities that can serve as adaptations for the holobiont.

We chose Dioon edule as a model to search for the relationship between the environmental differences where cycads thrive (soil type and pH), and the coralloid root microbiome’s genomic diversity (BGCs) across two contrasting environments. Using a culture-independent methodology, we found that Nostoc populations dominate over other bacterial lineages, contrary to previous observations. Differences in nitrogenase GCs and siderophore BGCs among the two environments suggest metabolic adaptation from the microbiome to the respective soil characteristics. Moreover, signatures of selection found in the dN/dS ratio point that differences in the BGCs integrity can be the result of positive selection acting on critical parts of the cluster. A terrestrial-symbiotic Nostoc lineage comparison suggests that these changes in dN/dS ratio are not balanced by genetic drift, expected in obligate symbionts. Together, this evidence confirms the main role that bacterial lineages have to help cycads in coping with the environment and holobioint adaptation to different niches.

Room 11-19

Investigating the emergence of protein sub-cellular localization by retracing the evolution of a *de novo* emerging ORF.

Carly J Houghton[1,2], Nelson Castilho Coelho[1], Nejla Ozbaki-Yagan[3], Allyson F O'Donnell[3], Anne-Ruxandra Carvunis[1]

[1]Department of Computational and Systems Biology, Pittsburgh Center for Evolutionary Biology and Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [2]Carnegie Mellon - University of Pittsburgh Joint PhD Program in Computational Biology, Pittsburgh, PA, USA. [3]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

A protein's sub-cellular localization is critical for defining its function. While the study of ancient proteins has revealed how sub-cellular localization changes during evolution, it does not provide a suitable paradigm for examining how this property emerges in the first place. Excitingly, novel protein-coding genes that arise *de novo* from previously non-genic sequences offer an unprecedented opportunity to study the emergence of localization. The protein encoded by *YBR196C-A*, an open-reading frame (ORF) in *S. cerevisiae* that is currently undergoing *de novo* gene birth, specifically localizes to the endoplasmic reticulum (ER) membrane. Did this novel protein acquire specific sequence signals that provide access to the ER? When and how did such signals arise during the process of *de novo* emergence? Are these signals recognized by the same central targeting pathways used by conserved ER proteins? To address these questions, we reconstructed *YBR196C-A*'s ancestors and determined their localization and the underlying cellular targeting and degradation pathways responsible using a combination of genetics and microscopy. Our results show that the ER localization potential was present in the incipient ORF and that *YBR196C-A* utilizes canonical ER targeting and degradation pathways. This study provides the first example of how a novel sequence exploits existing cellular infrastructure to gain access to a selective cellular compartment that offers a prospective niche on its journey to establishing a specific cellular function.

Room 11-20

Evolution in the brewery: Impact of structural variation on brewing yeast adaptation

Christopher R L Large[1,2], Noah Hanson[1], Andreas Tsouris[3], Omar Abou Saada[3], Jirasin Koonthongkaew[4], Yoichi Toyokawa[4], Tom Schmidlin[5], Daniela A Moreno-Habel[6], Hal McConnellogue[7], Richard Preiss[8], Hiroshi Takagi[4], Joseph Schacherer[3], Maitreya J Dunham[1]

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [2]Molecular and Cellular Biology Program, University of Washington, Seattle, Washington, USA. [3]Université de Strasbourg, CNRS, GMGM, UMR 7156, Strasbourg, France. [4]Division of Biological Science, Nara Institute of Science and Technology, Ikoma, Nara, Japan. [5]Postdoc Brewing Co., Redmond, Washington, USA. [6]Elysian Brewing Co., Seattle, Washington, USA. [7]Drake's Brewing Co., San Leandro, California, USA. [8]Escarpment Laboratories, Guelph, Ontario, Canada

Abstract

Ale brewing yeast are the result of admixture between a set of diverse strains of yeast, resulting in a heterozygous tetraploid that has since undergone numerous genomic rearrangements. As a result, comparisons between the genomes of modern related ale strains show both extensive aneuploidy and mitotic recombination that has resulted in a loss of intragenomic diversity. Despite these observations, it is currently unclear what impact these mutations have had on these yeasts' evolution and whether they have contributed to their domestication. Utilizing a common brewery practice known as 'repitching', in which yeasts are reused over multiple beer fermentations, we generated population time-courses from multiple breweries utilizing similar strains of yeast to investigate brewing yeast evolution in their modern, human-made environment. Applying whole-genome sequencing to the time-courses, we have found that the same aneuploidy and mitotic recombination events reproducibly rise to high frequency during adaptation to modern brewing conditions, indicating that domestication is an on-going process for ale brewing yeasts. Through genomic and phenotypic analysis of clones bearing a particular highly recurrent mitotic recombination event, we have determined that there is likely two loci that contribute to adaptive changes in cellular aggregation and growth in beer brewing conditions. Overall, we hope to garner a greater understanding of how the ale yeast's genome structure influences their adaptation to the brewery.

Room 12-01

Multiple simple genetic routes for the evolution of paralog specificity in a molecular complex

Carlos R Cortez-Romero[1], Arvind Pillai[2], Jixing Lyu[3], Arthur Laganowsky[3], Joseph W Thornton[1]
[1]University of Chicago, Chicago, IL, USA. [2]University of Washington, Seattle, WA, USA. [3]Texas A&M University, College Station, TX, USA

Abstract

Thousands of protein complexes are heteromers composed of paralogs. How do sister proteins after duplication evolve specific heteromeric interactions? Heteromeric interactions typically involve complementary interactions between many residues at the interface between proteins, suggesting that many mutations might be necessary to evolve a new specific interaction. Using ancestral protein reconstruction and biochemical experiments, we aimed to identify the genetic changes and biochemical mechanisms that caused the evolution of hemoglobin (Hb), a heterotetramer composed of two Hba and two Hbb subunits. We previously showed that Hb's ancient precursor was a homodimer at the time of the duplication that produced these paralogs; in the subsequent phylogenetic interval, a new interaction evolved to produce a tetramer, and the complex evolved its heterospecificity. Here we show that the evolution of heterotetramerization had a simple genetic basis. First, we found that a single historical substitution from this interval is sufficient to cause the evolution of the new tetrameric interaction with affinity comparable to modern Hb. Additionally, several nonhistorical mutations at the same site can also confer tetramerization. Second, we found that specificity for the heterotetramer was conferred strictly through substitutions at the ancestral dimer interface. Third, it takes only a few historical substitutions to confer heterospecificity at this interface, and substitutions that occurred in either paralog are sufficient for this effect. At both interfaces, residues that did not change during evolution contribute further to affinity or specificity. These results show that the elaborate interfaces that underlie molecular complexity can evolve in many simple ways.

Room 12-02

Reverse transcriptase-related genes and their possible role in the host cells

Irina A Yushenova, Fernando Rodriguez, Irina R Arkhipova
Marine Biological Laboratory, Woods Hole, MA, USA

Abstract

A distinct class of cellular reverse transcriptases (RTs), named reverse-transcriptase related (*rvt*) genes, is the only RT type that is found in both eukaryotes (fungi, plants, protists, certain invertebrates) and prokaryotes (selected bacteria). These RTs are immobilized in genomes as single-copy genes and are preserved by natural selection. Phylogenetic analysis of *rvt* sequences from all kingdoms of life indicates shared ancestry between prokaryotic and eukaryotic *rvts*. These characteristics of *rvt* genes may imply a biological function that is applicable to both prokaryotes and eukaryotes. RVT proteins from the filamentous gliding bacterium *Herpetosiphon aurantiacus* and the model ascomycete fungus *Neurospora crassa* display a peculiar property of initiating polymerization via protein priming, which is abolished by site-directed mutagenesis of the catalytic aspartate in the RT domain. The N-terminal coiled-coil domain allows RVT proteins to form multimers and could be a good candidate for interacting with metal ions. We find that the species with active *rvt* genes in their genomes (*H. aurantiacus*, *N. crassa,* and the bdelloid rotifer *Adineta vaga*) display signs of altered growth and behavior after treatment with varying concentrations of transition metal ions (nickel, cobalt, iron, etc.), and that expression of *rvt* genes in the above species can be strongly induced upon such treatments. When recombinant *H. aurantiacus* RVT is expressed in *E. coli*, it provides a growth advantage for bacteria in iron-rich environments. Participation of domesticated RTs in the response to environmental metals could potentially reveal an ancient function of these genes in the early evolution of life on earth.

Room 12-03

BALSAM: A Database for Exploring the Contrasting Evolutionary Paths of the Plant Genera *Hydrocera* and *Impatiens* (Family Balsaminaceae)

Alisha Harrison, Schuyler Humes, Dawniel Facque, Nicole Williams, Sudhindra R Gadagkar
Midwestern University, Glendale, AZ, USA

Abstract

Separated by 117 million years of evolution, the plant families Begoniaceae and Balsaminaceae each contain two contrasting genera: one expansively speciose and other monospecific. We focus on Balsaminaceae, which contains the genera *Impatiens* and *Hydrocera*. *Impatiens* is a large taxon, with hundreds of extremely diverse species spread across much of the old world and N. America. This genus has earned the epithet ‘notorious’ for the difficulty in morphological differentiation among its constituents. Nevertheless, not a year goes by without (typically several) new species of *Impatiens* being discovered. While there are larger plant genera (e.g., *Astragalus*), *Impatiens*, like *Begonia*, is unique in that its only sister genus, *Hydrocera*, consists of a single, phenotypically static species (*H. triflora*) that is also confined to a specific geographic region. These differences in evolutionary lability, phenotypic diversity, and geographic distribution between its two genera make Balsaminaceae an excellent model to study speciation. However, data necessary to undertake such studies lie scattered across books, book chapters, and thousands of papers spanning more than 250 years. Indeed, we are still relying on opinions and speculations on even the number of known species of *Impatiens*, currently a vague “more than 1000 species.” Therefore, as a first task, we began collecting the names and many relevant details of *H. triflora* and all known species of *Impatiens*. With a current count of at least 1100 known species, we present an easily searchable repository of this information that should aid further research into the evolutionary dynamics of this interesting family.

Room 12-04

Social evolution in the sea: the effect of eusociality on mitochondrial and nuclear genomes across snapping shrimps.

Solomon T. C. Chak
SUNY Old Westbury, Old Westbury, NY, USA

Abstract

Eusociality, often considered the pinnacle of animal social evolution, has convergently arisen at least 17 times in arthropods. Although a great deal of research has explored the genomic underpinning of eusociality, relative less has tested whether and how sociality may also affect genome evolution. Focusing on a group of snapping shrimps that exhibit multiple independent origins of eusociality, we explored the mitochondrial and nuclear genomes across species using ddRAD sequencing. We found that eusocial species have larger nuclear genomes with more transposable elements, as well as signals of relaxed purifying selection in the mitochondrial genes. Further, demographic inference showed that eusocial species had lower but more stable effective population sizes across 100,000 generations. Our results demonstrating that sociality can influence the evolution of the genome, likely through changes in demography related to patterns of reproductive skew.

Room 12-05

From Bats to Frogs: Uncovering an Antiviral Role for Receptor Transporter Proteins

Ian N Boys, John W Schoggins
UT Southwestern Medical Center, Dallas, TX, USA

Abstract

Viruses and their hosts are engaged in “genetic arms races” in which each side attempts to gain the advantage over evolutionary time. Results of these conflicts are wide-ranging: viruses diversify, hosts establish species-specific barriers to some viruses while remaining susceptible to others, and the lines for future genetic conflicts are drawn. In mammals, many antiviral effectors – proteins that directly inhibit viral infection – show species- or lineage-specific properties which are believed to be the result of past or ongoing conflicts. We identified Receptor Transporting Protein 4 (RTP4) from the bat *Pteropus alecto* as a potent inhibitor of flavivirus infection. Mechanistic studies determined that RTP4 is an RNA-binding protein that associates with flavivirus replication machinery, binds replicating viral RNA, and suppresses viral genome amplification. Phylogenomic analysis revealed that RTP4 has evolved under positive selection in several mammalian lineages, consistent with a model in which host-virus conflicts have shaped its evolution as a restriction factor. Indeed, we found that mammalian RTP4 orthologs exhibit striking patterns of antiviral specificity in cell culture. In follow-up work, we identified signatures of positive selection in several non-mammalian RTP homologs, indicative of a putative role in innate immunity. We screened a collection of vertebrate RTPs against a panel of viruses and identified antiviral RTPs in the African clawed frog, *Xenopus laevis*, that exhibit mosaic phenotypes which resemble those of mammalian RTP4 orthologs. Within the context of our findings with mammalian RTP4, these data suggest that Receptor Transporter Proteins are involved in host-virus genetic conflicts outside of Mammalia.

Room 12-06

Evidence of polygenic adaptation within Europeans and among continental populations

Minhui Chen, Charleston W.K. Chiang
University of Southern California, Los Angeles, CA, USA

Abstract

Human height had been one of the most well known examples of polygenic adaptation, but this conclusion was recently challenged as residual stratification from large-scale consortium studies was suggested to explain the previously noted difference in polygenic scores (PS) across European populations. It thus remains an open question whether a signature of polygenic adaptation at height-associated loci at all exists in any human population. We re-examined this question at two geographical scales: within Europe and across continents. We utilized height-associated loci ascertained from two biobank datasets: the UK Biobank and the Biobank Japan, to alleviate concerns of confounding caused by residual stratification in GWAS. Within Europe, we found that the Sardinians, one of the shortest European populations, remain significantly shorter than expected under neutrality (~0.22 standard deviation shorter than CEU by PS, p = 3.9e-4). We also found that height-associated SNPs showed a robust adaptive signals in UK population by tSDS (p = 9.1e-4), which was recently corroborated by large-scale family studies. Among continental populations (Africans, Europeans, and East Asians), we found that frequencies of height-associated SNPs are significantly more differentiated than non-associated SNPs (p = 0.0025). However, we found no significant difference among continental population based on PS, which we showed through simulation could lose power in detecting polygenic adaptation in presence of independent convergent selections. Taken together, by ascertaining height loci from Biobank datasets, ideally from an outgroup population, we further supported the evidence of polygenic adaptation at height-associated loci at least among some human populations.

Room 12-07

Evolution of chemoreceptors in social wasps

Andrew W Legan, Christopher M Jernigan, Sara E Miller, Matthieu F Fuchs, Michael J Sheehan
Cornell University, Ithaca, NY, USA

Abstract

Independent origins of sociality in bees and ants are associated with independent expansions of particular odorant receptor (OR) gene subfamilies. In ants, one clade within the OR gene family, the 9-exon subfamily, has dramatically expanded. These receptors detect cuticular hydrocarbons (CHCs), key social signaling molecules in insects. It is unclear to what extent 9-exon OR subfamily expansion is associated with the independent evolution of sociality across Hymenoptera, warranting studies of taxa with independently derived social behavior. Here we describe odorant receptor gene family evolution in the northern paper wasp, *Polistes fuscatus*, and compare it to four additional paper wasp species spanning ~40 million years of evolutionary divergence. We find 200 putatively functional OR genes in *P. fuscatus*, matching predictions from neuroanatomy, and more than half of these are in the 9-exon subfamily. Most OR gene expansions are tandemly arrayed at orthologous loci in *Polistes* genomes, and microsynteny analysis shows species-specific gain and loss of 9-exon ORs within tandem arrays. There is evidence of episodic positive diversifying selection shaping ORs in expanded subfamilies. Values of omega (d_N/d_S) are higher among 9-exon ORs compared to other OR subfamilies. Within the *Polistes* OR gene tree, branches in the 9-exon OR clade experience relaxed negative (relaxed purifying) selection relative to other branches in the tree. Patterns of OR evolution within *Polistes* are consistent with 9-exon OR function in CHC perception by combinatorial coding, with both natural selection and neutral drift contributing to interspecies differences in gene copy number and sequence.

Room 12-08

Transcription factors drive the relationship between gene age and tissue specificity in *Drosophila melanogaster*

Evan Witt, Nicolas Svetec, Sigi Benjamin, Li Zhao
Rockefeller University, New York, NY, USA

Abstract

The testis is a hotbed of evolutionary genetic novelty. Young genes are highly testis-biasedbiased, and the testis expresses most *de novo* genes, which are young genes born from ancestrally non-genic DNA. It was unknown, however, which cell types drive this testis-bias of young genes. Using single-cell RNA-seq, we found that *de novo* genes and duplicated genes both show highly dynamic expression patterns and can be found in various stages of spermatogenesis. Many *de novo* genes are enriched in meiotic cells, implying a functional role therein. We sought to characterize the genomic features that confer testis-biased expression on young genes. Using data from Flyatlas2, we noticed that while young genes are often testis-biased, old genes often show ovary-biased expression. Using a novel measure of a gene's tissue-specific upstream Transcription Factor (TF) expression, we found that gene expression is highly tied to upstream TF activity in ovary, but not in testis, suggesting that trans-regulation is highly important for ovary expression, but not testis expression. ATAC-seq of both tissues did not sufficiently explain age-related expression bias, suggesting an important but minor role in long-term gene expression evolution. Our results suggest that a relatively small set of TFs is sufficient to initiate expression of young genes in testis. Taken together, our work provides a deeper understanding of how the testis maintains its core reproductive function while being a hotbed of evolutionary innovation.

Room 12-09

Holosteans contextualize the role of the teleost genome duplication in promoting evolutionary novelty of the ray-finned fish innate immune system

alex Dornburg[1], Katerina Zapfe[1], Emma ferraro[2], Lindsay Roupe-Abrams[1], Dustin Wcisel[2], Andrew Thompson[3], Ingo Braasch[3], Tatsuya Ota[4], Jeffrey Yoder[2]
[1]UNC, Charlotte, NC, USA. [2]NC State University, Raleigh, NC, USA. [3]Michigan State University, E Lansig, MI, USA. [4]SOKENDAI, Tokyo, Japan

Abstract

Over 99% of ray-finned fishes (Actinopterygii) are teleosts, a clade that collectively comprises half of all living vertebrates that have successfully diversified across virtually all fresh and saltwater ecosystems. This diversity raises the question of how the immunogenetic diversity required to persist under heterogeneous pathogen pressures evolved. The teleost genome duplication (TGD) has been hypothesized as the evolutionary event that provided the genomic substrate for rapid genomic evolution and innovation. However, studies of teleost-specific immune receptors have been largely limited to comparisons either among teleosts or between teleosts and distantly related vertebrate clades such as tetrapods. Here we describe and characterize the receptor diversity of two clustered innate immune gene families in the teleost sister lineage: Holostei (bowfin and gar). Using genomic and transcriptomic data for spotted gar and bowfin, we provide the first detailed investigation of the phylogenetic history and conserved synteny of diverse immunoglobulin domain-containing proteins (DICPs) and novel immune-type receptors (NITRs). These data demonstrate an ancient linkage of DICPs to the major histocompatibility complex (MHC) and reveal an evolutionary origin of NITR variable-joining (VJ) exons that predates the TGD by at least 50 million years. Further characterizing the receptor diversity of Holostean DICPs and NITRs illuminates a sequence diversity that rivals the diversity of these innate immune receptor families in many teleosts. Taken together, our findings provide important historical context for the evolution of these gene families that challenge prevailing expectations concerning the consequences of the TGD during Actinopterygiian evolution.

Room 12-10

Coevolution is pervasive between unrelated glycosylation pathways and point to potential disease modifiers.

Holly Thorpe, Nathan L Clark, Clement Y Chow
University of Utah, Salt Lake City, Utah, USA

Abstract

Protein glycosylation is the most common post-translational modification, including N-linked glycosylation, O-linked glycosylation, and GPI anchor biosynthesis. These pathways are separate and have nearly no overlapping components. Congenital disorders of glycosylation (CDGs) are a group of rare disorders caused by defects in glycosylation. Patients present with seizures, hypotonia, and developmental delays; however, because of the ubiquity of protein glycosylation there can be symptoms associated with every organ system. Clinical variability between CDG patients, even those with the same mutation, is common, suggesting there are modifier genes affecting the phenotype. I am employing evolutionary approaches to identify modifier genes of CDGs. Evolutionary Rate Covariation (ERC) relies on the premise that proteins that interact physically or genetically or are functionally related coevolve at similar rates. ERC values are calculated using the correlation coefficient of evolutionary rates of gene pairs in a species tree. I pulled pairwise ERC values for 224 genes involved in protein glycosylation calculated from a species tree with evolutionary rates from vertebrates, worms, *Drosophila*, and yeast. As expected, we found enrichment for high ERC values within each glycosylation pathway. Surprisingly, we also identified a number of components of each pathway that showed high ERC with unrelated glycosylation pathways. For example, MAN2A2, a component of N-glycosylation localized to the Golgi, shows strong ERC with most components of the GPI anchor pathway in the ER. These types of coevolution signatures tell us that there are unappreciated connections between these unrelated pathways and have important implications for how they might be evolving.

Room 12-11

Widespread changes in gene expression accompany body size evolution in nematodes

Gavin C. Woodruff[1,2], Erik Johnson[1], John Willis[1], Patrick C. Phillips[1]
[1]University of Oregon, Eugene, OR, USA. [2]University of Oklahoma, Norman, OK, USA

Abstract

Body size is a fundamental trait that drives multiple evolutionary and ecological patterns. *Caenorhabditis inopinata* is a fig-associated nematode that is exceptionally large relative to other members of the genus, including *C. elegans*. We previously showed that *C. inopinata* is large primarily due to postembryonic cell size expansion that occurs during the larval-to-adult transition. Here, we describe gene expression patterns in *C. elegans* and *C. inopinata* throughout this developmental period to understand the transcriptional basis of body size change. We performed RNAseq in both species across the L3, L4, and adult stages. Most genes are differentially expressed across all developmental stages, consistent with *C. inopinata*'s divergent ecology and morphology. We also used a model comparison approach to identify orthologs with divergent dynamics across this developmental period between the two species. Among such genes were two transcription factors previously shown in *C. elegans* to be important for body size that are regulated by the TGF-β signaling pathway. Multiple hypodermal collagens were also observed to harbor divergent developmental dynamics across this period. *C. elegans*-specific ontology enrichment reveals such genes tend to be expressed in neurons and regulate behavior; they also include genes important for molting and body morphology. A comparison of such genes with previous *C. elegans* experiments reveals overlap with stress response, developmental timing, and small RNA/chromatin regulation. These results have identified candidate genes that will be further investigated to test their roles in cell size divergence and broaden our understanding of the genetic bases of body size evolution.

Room 12-12

Population genomics of variegated toad-headed lizard Phrynocephalus versicolor and its adaptation to the colorful sand of the Gobi Desert

Diana Aguilar Gómez
University of California, Berkeley, CA, USA

Abstract

The variegated toad-headed agama, Phrynocephalus versicolor lives in the arid landscape of the Chinese Gobi Desert. We analyzed populations from 3 different locations which vary in substrate color and altitude: Heishankou (HSK), Guazhou County (GZ), and Ejin Banner (EJN),. The substrate color is either light-yellow (GZ), yellow (EJN) or black (HSK); the corresponding lizard populations colors largely match their substrate in degree of melanism. We assembled the P. versicolor genome and sequenced over 90 individuals from the three different populations. Genetic divergence between populations corresponds to their geographic distribution. We infer the genetic relationship between these populations and use selection scans and differential expression to identify genes that appear to have been under selection during the differentiation of the populations. Slc2a11 and akap12 among other genes are highly differentiated and may be related to pigment adaptation to substrate color in P. versicolor.

Room 12-13

Molecular evolution of genes related to antioxidant response in cetaceans.

Giovanna Selleghin Veiga, Lucas Freitas, Érica Martinha Silva de Souza, Felipe André Silva, Letícia Magpali, Mariana Nery
State University of Campinas - UNICAMP, Campinas, São Paulo, Brazil

Abstract

Ischemia and reperfusion are a known pathological condition, mostly studied in human diseases, which leads to the production of ROS (reactive oxygen species) by many pathways, resulting in oxidative damage. Previous studies showed that cetaceans do not suffer from this issue, mainly due to their antioxidant enzymes. In this study, we aimed to investigate the molecular evolution of the antioxidant genes: CAT, GPX3, GSR, PRDX1, PRDX3, SOD1, and the production of ROS, by XDH. We used the rate of nonsynonymous (dN) to synonymous (dS) substitutions (ω = dN/dS) in the PAML and HyPhy packages to study adaptive molecular evolution on these genes in the cetacean group. We obtained at least 61 mammalian sequences, including 16 cetacean species. Our analysis showed that the genes GPX3, GSR, PRDX1, PRDX3 and SOD1 are positively selected in cetaceans, with GPX3 and PRDX1 showing significant differences of ω values between Mysticeti and Odontoceti. We identified 23 and 1 sites evolving under positive selection in PRDX1 and SOD1 respectively, with some sites at positions close to the active site of the protein, suggesting a possible change in functional activity. We also found episodic selection in the GSR gene in a specific clade of Odontoceti, which demands more investigation. Further, we will perform analyses of the physicochemical properties of amino acidic sites found to be evolving under positive selection and will employ the same methodological framework to study antioxidants in pinnipeds and other semi-aquatic mammalian species to search for a possible convergent evolution.

Room 12-14

Convergent pathways associated with adaptations to hypoxia in independent lineages invading high-altitude environments

Allie M Graham[1,2], Bogdan Kirilenko[3], Michael Hiller[4,5,6], Nathan L Clark[1,2]
[1]University of Utah, Salt Lake City, Utah, USA. [2]Eccles Institute for Human Genetics, Salt Lake City, Utah, USA. [3]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [4]LOEWE Centre for Translational Biodiversity Genomics, Frankfurt, Germany. [5]Senckenberg Society for Nature Research, Frankfurt, Germany. [6]Goethe University, Frankfurt, Germany

Abstract

A fundamental question in biological research is understanding the relationship between genotype and phenotype; computational approaches using patterns of convergent phenotypic evolution assist in providing a complementary approach to understanding predicting genotype-phenotype associations. Convergent evolution of adaptive traits in distantly related organisms inhabiting similar environments with the same selective pressures is common – with evidence of similarities in genomic variation, evolutionary-rate shifts, and pseudogenization across similar genes/pathways. Hypoxia is one selective pressure that stimulates a similar physiological response across metazoans allowing organisms to match O2 supply and demand. Thus, the ability to identify genes with such convergent rate signatures enable new insights into the molecular basis of these traits.

Using large-scale mammal genome alignments, we test the hypothesis that there will be similar genes and regulatory networks associated with life in oxygen limited environments across independent mammalian lineages. First, using individual genes via evolutionary-rate analysis, we uncover the convergent mechanisms associated with high-altitude lineages. Second, we use an expanded list mammal genomes to understand the role of how pseudogenization has played a role in response to hypoxia. Overall, we find evidence for convergence on pathways known to be involved in the response to hypoxia, nitric oxide, angiogenesis, mitochondrion, and inflammation. In addition, we found evidence of convergence on pseudogenized genes associated with the inflammatory pathway, respiratory electron transport, platelet homeostasis, and cardiac muscle contraction. Ultimately, this work shows significant evidence that independent high-altitude lineages are using shared adaptive pathways by virtue of the same pressure of oxygen limitation.

Room 12-15

Evaluation of methods for inference of ancestral recombination graphs

Débora Y. C. Brandt[1], Xinzhu Wei[2], Yun Deng[1], Andrew H. Vaughn[1], Rasmus Nielsen[1]
[1]UC Berkeley, Berkeley, CA, USA. [2]UCLA, Los Angeles, CA, USA

Abstract

The ancestral recombination graph (ARG) is a model that describes the genealogy of samples of DNA sequences, keeping full information of coalescence, mutation, and recombination events. Recent methods have made impressive progress towards scalably estimating whole-genome genealogies. In addition to inferring the ARG, some of these methods can also provide ARGs sampled from a defined posterior distribution. Obtaining good samples of ARGs is crucial for quantifying statistical uncertainty and for the downstream processing of ARGs to estimate parameters such as effective population size, mutation rate and age of variants. Here, we use simulations to benchmark three ARG inference programs: ARGweaver, Relate and tsdate. We use neutral coalescent simulations to 1) compare the true coalescence times to the inferred times at each locus; 2) compare the distribution of coalescence times to the expected exponential distribution across all loci ; 3) for ARGweaver and Relate, which both provide posterior samples, evaluate whether the sampled coalescence times have the properties expected of a valid posterior distribution. As all methods are Bayesian, they are not expected to yield unbiased estimates of coalescent times. Although biased, estimates from Relate and ARGweaver are more strongly correlated to the true values than those from tsdate. Finally, ARGweaver provides samples from a distribution that is closer to the correct posterior distribution than that produced by Relate, but this accuracy comes at a substantial computational cost. We conclude that the best choice of method will depend on the number of input sequences and on the specifics of the downstream analyses.

Room 12-16

Probing the dynamic properties of sequence space with an evolved GFP-like protein

Michael Shavlik, Michael Harms
University of Oregon, Eugene, OR, USA

Abstract

Understanding how proteins access new functions is critical in deciphering the evolutionary process and can inform applications such as protein engineering. A protein's ability to evolve depends on its local neighborhood of available states as it traverses sequence space. Key properties of neighborhoods such as robustness and evolvability directly shape a protein's evolutionary potential. Prior studies have assessed these neighborhood properties but have been limited in scope to an ancestral or extant genotype's neighborhood and only provide a static snapshot in sequence space. To date, little work has been done to show how these properties change as naturally evolving proteins acquire new functions. Characterizing changes in neighborhood properties across a realized evolutionary trajectory will reveal the connectivity and distribution of novel, neutral, and non-functional phenotypes, and the degree of contingency and random chance underpinning molecular evolution. To study how robustness and evolvability change across neighborhoods for evolving proteins, we used ancestral reconstruction to sample intermediate genotypes for a GFP-like protein that underwent a long-term, evolutionary color transition. We coupled random mutagenesis with flow cytometry to generate mutant protein libraries, characterize their fluorescent phenotypes, and then reconstruct local neighborhoods along the historical trajectory. These inferred neighborhoods showed stark qualitative differences in robustness and evolvability over evolutionary time. This suggests the distribution of function in sequence space is highly non-uniform, and that the nature of the local neighborhood around a genotype may depend on whether a sequence is squarely within its functional neutral network or in transition between two functional neutral networks.

Room 12-17

Large-scale phylodynamic analysis of human pathogenic RNA viruses

Pascal Mutz[1], Nash D Rochman[1], Yuri I Wolf[1], Guilhem Faure[2], Feng Zhang[2,3,4,5,6], Eugene V Koonin[1]
[1]National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, C, MA, USA. [4]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

RNA viruses possess a highly diverse range of “lifestyles” (host range, transmission mode, etc). We analyzed the phylodynamic properties of 20 human pathogenic monopartite RNA viruses, to gain insight into unifying trends and their implications for viral evolution.
Most of the phylogenetic trees based on complete viral coding sequences show “bushy” topologies with deep branches and parallel major lineages circulating globally for decades if not centuries, in contrast to Influenza A virus H3N2, which is known for a fast lineage turnover. The causes of the coexistence of parallel lineages likely include both spatial separation of lineages and periodic sylvatic transmissions from environmental reservoirs.
Single protein-coding genes from different groups of viruses typically have low dN/dS ratios, characteristic of strong purifying selection. The ratios of the number of non-synonymous to synonymous substitutions (nN/nS) from root to tip of viral evolutionary trees present a complex picture, suggesting temporal changes in the selection landscape.
Estimation of the effective population size (Ne) for most of the analyzed viruses yields 10- to 100-fold higher values than previously reported for Influenza A H3N2. The observed Ne trends could not be explained solely by viral transmission mode or spreading rate.
Altogether, our results show that “bushy” phylogenetic trees reflecting globally circulating parallel lineages and large Ne represent the prevailing evolutionary trend among human pathogenic RNA viruses. Whether novel emerging viruses, such as SARS-CoV-2, follow the same trend during their evolution subsequent to entering human populations remains to be monitored.

Room 12-18

Recurrent Duplication and Diversification of Acrosomal Proteins Mediating Fertilization in Abalone

Jolie A Carlisle[1], Megan Glenski[2], Willie J Swanson[1]
[1]University of Washington, Seattle, WA, USA. [2]Gonzaga University, Spokane, WA, USA

Abstract

Reproductive proteins mediating fertilization commonly exhibit rapid sequence diversification, often driven by positive selection. Gene duplication may also contribute to the diversification of fertilization proteins, however, the contribution of duplication events has not been well studied in most fertilization models. The marine mollusk abalone is a classic model for studying fertilization. Its two acrosomal proteins (lysin and sp18) are ancient gene duplicates (paralogs) with unique gamete recognition functions hypothesized to have arisen through subfunctionalization. Through detailed genomic and bioinformatic analyses we show how duplication events followed by sequence diversification have played an ongoing role in the evolution of abalone acrosomal proteins. The common ancestor of abalone is shown to have had four members of its acrosomal protein family in a tandem gene array that routinely experienced positive selection. We find that ancestral sp18 paralogs are both rapidly evolving, but with distinct regions undergoing positive selection consistent with a subfunctionalization model. We also describe more recent species-specific duplications of both lysin and sp18 in the European abalone H. tuberculata indicating that acrosomal protein duplications continue to play a role in the evolution of these rapidly evolving proteins. Despite having clade-specific acrosomal protein paralogs, H. tuberculata has no concomitant duplications of egg coat proteins, suggesting that duplications of egg proteins per se are not responsible for retention of duplicated acrosomal proteins. We hypothesize that, in a manner analogous to host/pathogen evolution, sperm proteins are selected for increased diversity through extensive sequence divergence and recurrent duplication driven by sexual conflict mechanisms.

Room 12-19

Disrupted satellite transcripts in the selfish *Segregation Distorter* system of *Drosophila melanogaster*

Xiaolu Wei[1], Vera Yu[2], Amanda Larracuente[2]
[1]University of Rochester Medical Center, Rochester, NY, USA. [2]University of Rochester, Rochester, NY, USA

Abstract

Meiotic drivers are selfish genetic elements that bias their transmission during gametogenesis, breaking Mendel's law of segregation. They drive rapid genome evolution and shape the evolution of genome structure. One of the well-known examples is *Segregation Distorter* (*SD*)--a drive system found in nearly all populations of *Drosophila melanogaster*. *SD* heterozygous male flies transmit the *SD* chromosome to nearly all of their progeny (>95%). The driver, *Segregation distorter* (*Sd*), encodes a truncated duplication of the gene *RanGAP*. The target, *Responder* (*Rsp*) corresponds to a block of tandem 120-bp satellite DNA (satDNA) repeats in the pericentric heterochromatin. *SD* targets *Rsp*-bearing sperm for destruction, however, its molecular mechanism is unknown. Here we found that not only *Rsp*, but also other satDNAs (*e.g. 359-bp*), show disrupted expression levels and localization in the presence of *SD*. In testes with *SD,* satDNA transcript signals are more aggregated, suggesting possible localization defects and/or phase separation of satellite-associated RNA-binding proteins. In addition to altered localization, we also found that satDNAs show reduced expression levels, and generate fewer piRNAs in the presence of *SD*. The expression level of other repeats like TEs and piRNA clusters remain unchanged, suggesting that the disrupted expression is exclusive to satDNA. A common theme in meiotic drive systems in Drosophila is that there is a link between meiotic drive, piRNAs, and heterochromatin. Our results not only provide insights into our understanding of the role of *Rsp* satDNA in *SD*, but also shed light on vulnerabilities of gametogenesis targeted by all meiotic drivers.

Room 12-20

Data-driven speciation tree prior for better species divergence times in calibration-poor molecular phylogenies

Qiqing Tao[1], Jose Barba-Montoya[1], Sudhir Kumar[1,2]
[1]Temple University, Philadelphia, PA, USA. [2]King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Precise time calibrations needed to estimate ages of species divergence are not always available due to fossil records' incompleteness. Consequently, clock calibrations available for Bayesian dating analyses can be few and diffused, i.e., phylogenies are calibration-poor, impeding reliable inference of the timetree of life. We examine the role of speciation birth-death tree prior on Bayesian node age estimates in calibration-poor phylogenies and test the usefulness of an informative, data-driven tree prior to enhancing the accuracy and precision of estimated times. We present a simple method to estimate parameters of the birth-death tree prior from the molecular phylogeny for use in Bayesian dating analyses. The use of a data-driven birth-death (ddBD) tree prior leads to improvement in Bayesian node age estimates for calibration-poor phylogenies. We show that the ddBD tree prior, along with only a few well-constrained calibrations, can produce excellent node ages and credibility intervals, whereas the use of an uninformative, uniform (flat) tree prior may require more calibrations. Relaxed clock dating with ddBD tree prior also produced better results than a flat tree prior when using diffused node calibrations. Our results have practical applications because the ddBD tree prior reduces the number of well-constrained calibrations necessary to obtain reliable node age estimates. This would help address key impediments in building the grand timetree of life, revealing the process of speciation, and elucidating the dynamics of biological diversification.

Room 13-01

Mutability of mononucleotide repeats explains the discrepancy between lab-accumulated mutations and the natural allele frequency spectrum of C. elegans

Moein Rajaei[1], Ayush Shekhar Saxena[1], Lindsay M. Johnson[1], Michael C. Snyder[1], Timothy A. Crombie[2], Robyn E. Tanny[2], Erik C. Andersen[2], Joanna Joyner-Matos[3], Charles F. Baer[1]
[1]University of Florida, Gainesville, FL, USA. [2]Northwestern University, Evanston, IL, USA. [3]Eastern Washington University, Cheney, WA, USA

Abstract

Mutation is the fuel of evolution, and is of fundamental importance in evolutionary biology. A usual and efficient way to estimate the properties of spontaneous mutation divorced from the influence of natural selection is by means of mutation accumulation (MA) experiments, in which the efficacy of selection is minimized. However, MA experiments potentially come with their own biases. Previous studies have consistently shown a discrepancy between the mutation spectrum observed in MA lines and the site frequency spectrum (SFS) of wild isolates. By applying MA experiments and whole genome sequencing to three strains of *C. elegans*, (N2, PB306, and an N2-derived strain carrying a defective allele at the *mev-1* gene), we investigated the property of spontaneous mutations in different part of the genome. We find that the mutational properties of mononucleotide repeats differ in both rate and spectrum from non-mononucleotide sequences, both for base-substitutions and insertion/deletion (indel) mutations. Comparison of the MA spectrum to that of segregating "private" alleles (which have presumably arisen recently as new mutations) reveals that the spectra at non-mononucleotides are similar between MA lines and wild isolates, whereas the mononucleotide spectra are very different, both for base-substitutions and indels. In addition, we compared the mutational properties of highly divergent regions of the *C. elegans* genome to those of weakly diverged regions. Our preliminary analysis suggests that the mutation rate is slightly higher in divergent regions, but the difference in mutation rate is not nearly large enough to explain the difference in nucleotide diversity.

Room 13-02

Leveraging Both Ancestral and Derived Information to Detect Local Introgression

Lesly Lopez-Fang[1], Diego Ortega-Del Vecchyo[2], Emily Jane McTavish[1], Emilia Huerta-Sánchez[3]
[1]University of California, Merced, Merced, CA, USA. [2]National Autonomous University Of Mexico, Querétaro City, Querétaro, Mexico. [3]Brown University, Providence, RI, USA

Abstract

Introgression is a common phenomenon that can reveal the shared evolutionary history between species across taxa. Modern-day non-African populations have approximately 1-2% of DNA inherited from the archaic humans Altai Neanderthals. Existing statistical methods such as Patterson's D statistic can detect introgression by measuring shared excess mutations between populations. The D statistic is effective genome wide but gives spurious inferences of introgression when applied to regions in the genome. In these regions D can also be undefined if there are no derived sites. We propose a statistical method D+, based on prior statistics, that can identify local introgression. Introgressed haplotypes are inherited in chunks that contain both ancestral and derived alleles. We increase the number of informative sites per region by incorporating both shared mutations and ancestral alleles to calculate D+. To test the robustness of D+ we ran coalescent simulations following the Out of Africa model with Neanderthal introgression into the non-African population. We calculated power of D and D+ using the simulated genetic variants and found that D+ has more precision than D in localized regions.

Room 13-03

Divergence time estimates for the hypoxia inducible factor-1 alpha (HIF1a) reveals an ancient origin of animals in low-oxygen environments

Flavia A. Belato[1], Beatriz Mello[2], Christopher J. Coates[3], Kenneth M. Halanych[4], Federico Brown[1], Juliana de M. Leme[1], Ricardo I. F. Trindade[1], Elisa M. Costa-Paiva[1]
[1]University of Sao Paulo, Sao Paulo, Brazil. [2]Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. [3]Swansea University, Swansea, United Kingdom. [4]Auburn University, Auburn, USA

Abstract

Defining in which geological period metazoans emerged is crucial to understand early animals evolution and physiology. Although the earliest accepted metazoan fossil record dates to the late Ediacaran period ~571 Ma, most molecular-clock estimates agree that the last common ancestor (LCA) of all extant animals emerged ~850 Ma, in the Tonian period, before the oldest evidence for widespread ocean oxygenation at ~635–560 Ma in the Ediacaran period. All metazoans are aerobic organisms, which means they are dependent on oxygen to survive. In low-oxygen conditions, most animals have an evolutionarily conserved pathway for maintaining oxygen homeostasis that triggers adaptive changes in gene expression via the hypoxia-inducible factor (HIFa). However, here we confirm the absence of a functional HIFa in sponges and ctenophores, corroborating that the metazoan LCA likely lacked the HIF pathway as well, and so could have aerobically metabolized under the very low-oxygen concentrations of their environments. Using Bayesian uncorrelated relaxed molecular-clock dating, we inferred that the lineage that originated HIFa arose deep earlier in the evolutionary history of animals, ~1,273 Ma in the Mesoproterozoic Era, consistent with the idea that an important fraction of the genetic toolkit required for animal development evolved deep earlier in their eukaryotic unicellular ancestors. Our data indicate at least two duplication events in the evolutionary history of HIFa, that generated three vertebrate paralogs, products of the two successive whole-genome duplications that occurred in the vertebrate LCA. Overall, our results support the hypothesis of a pre-Cryogenian emergence of metazoans under low-oxygen conditions.

Room 13-04

Estimating dispersal rates and locating genetic ancestors with genome-wide genealogies

Matthew Osmond[1], Graham Coop[2]
[1]University of Toronto, Toronto, Canada. [2]UC Davis, Davis, USA

Abstract

Spatial patterns in genetic variation reflect movements of the ancestors of the sample, suggesting we might infer past movements from contemporary genetic data. The recent ability to estimate genome-wide genealogies in recombining species opens up new possibilities on this front. In this joint work with Graham Coop, I explain how to use genome-wide genealogies and the locations of contemporary samples to infer dispersal rates and the locations of genetic ancestors. After validating our approach with simulations, I show how this method can be used to reconstruct range expansions in *Arabidopsis thaliana*.

Room 13-05

Impacts of Spatial Structure on the Evolution of Antibiotic Resistance

Mahfuza Akter[1], Susan Bailey[2]
[1]Clarkson University, Potsdam, NY, USA. [2]CLarkson University, Potsdam, NY, USA

Abstract

In natural environments, bacteria may evolve in different ways in response to antibiotics, depending on the other characteristics of their environment. Spatial structure in the environment is one characteristic that may drive a diverse range of evolutionary responses, depending on the impact of local neighborhood interactions and associated small-scale variations in the environment. Local interactions between bacterial cells can impact sensitivity to antibiotics and evolutionary dynamics in general. A single bacterial population may encounter a diverse range of spatial structure and heterogeneity depending on their habitats; and so even within a population, adaptive responses may vary widely. To test the impacts of spatial structure on the evolution of antibiotic resistance, I tracked the evolution of *Pseudomonas aeruginosa* exposed to a consistent sub-lethal concentration of the antibiotic ciprofloxacin across four different environments that differed only in their concentration of agar, thus capturing a range of viscosities that *P. aeruginosa* may encounter in its natural habitats. We found that degree of spatial structure and the antibiotic interacted to drive differences in evolution at both the phenotypic and genomic levels. Both antibiotics and spatial structure drove differences in genomic diversity, confirming the importance of environment in driving evolutionary dynamics in general. We also found that while spatial structure did not significantly affect evolved antibiotic resistance, it did impact other evolved traits, suggesting that spatial structure may have important implications for persistence of antibiotic resistance in bacterial populations and can play crucial role at molecular level for the evolution in general.

Room 13-06

Modeling Probabilities of Retention of Gene Copies after Consecutive Whole Genome Duplication Events

Amanda Erin Wilson, David A Liberles
Temple University, Philadelphia, PA, USA

Abstract

Gene duplication creates redundant gene copies which allows the genes to accumulate mutations under positive or neutral selection. Most genes have one of their copies pseudogenized through a process called nonfunctionalization. However, both copies could be retained in the genome if they take on new or more specialized functions (neofunctionalization, subfunctionalization). We often attribute the retention of a gene to the gene duplicability hypothesis, that some genes are inherently more likely to be retained. However, a recent study showed unexpected results in the Atlantic salmon genome that appeared to suggest that the probability of a gene being retained after one whole genome duplication event is independent of the probability that the gene is retained after a previous one. Here, we construct four mathematical models for different hypotheses that explain retention of gene copies resulting from consecutive gene duplication events. The four models include an independence hypothesis, a revised gene duplicability hypothesis, a novel mutational opportunity hypothesis, and a hybrid model of the latter two models. The novel mutational opportunity hypothesis explains that the probability of the retention of a gene after consecutive whole genome duplication events is affected by mechanisms that allowed for previous retention of these genes. For all the models, we incorporate time between duplication events, t_1, and time since the most recent duplication event, t_2. We applied statistical model testing with real world data sets to identify the best supported biological hypothesis. Our datasets include organisms that have had relatively recent consecutive gene duplication events.

Room 13-07

Examining the role of photoreceptor development genes in the evolution of transmuted reptile photoreceptors

Emily M Dong, Belinda SW Chang, Vincent Tropepe
University of Toronto, Toronto, Ontario, Canada

Abstract

Rods and cones are two distinct and highly specialized photoreceptor classes that allow vertebrates to see in an astonishingly wide range of visual environments. Among vertebrates, reptiles are unique in having repeatedly evolved secondary losses of either rods or cones. This loss has been hypothesized to have occurred not by the outright loss of one type, but by “transmutation” or the evolutionary transition between rods and cones, whereby one photoreceptor class evolves features of the other. Transmutation in reptile retinas and has been confirmed by the presence of molecular and morphological markers in adult eyes. However, it remains unresolved how these mature transmuted photoreceptors arise during development. Photoreceptor identity is tightly controlled by a network of genes largely conserved across vertebrates and we hypothesized that this network has undergone adaptive evolution in transmuted reptile photoreceptors. To address this, we use a comparative computational approach to examine evolutionary patterns of sequence constraint and selection on photoreceptor development genes combined with an examination of morphological and molecular markers in retinal tissues. Preliminary results reveal that reptile orthologs of many photoreceptor differentiation factors are present, and the evolutionary rates of these genes point to a divergence of selection pressures across a wide sampling of taxa.

Room 13-08

The effects of spatial structure on adaptation and repeatability in a microbial evolution experiment

Susan F Bailey, Katherine Tulowiecki, Morgan McGrath, Aria Belle, Herbert Fountain, Andrew Trudeau, Mahfuza Akter
Clarkson University, Potsdam, NY, USA

Abstract

Observations from both natural and experimental populations suggest that evolutionary dynamics and the repeatability of those evolutionary changes can vary tremendously across taxa and environment. While many microbial evolution experiments have been conducted under well-mixed homogeneous conditions, theoretical models suggest that local interactions and environmental heterogeneity can affect evolutionary dynamics in a number of important ways, including changes to rate of adaptation and increased probability of parallel evolution. Here we report results from an experiment comparing the evolutionary dynamics of replicate populations of *Pseudomonas fluorescens* in well-mixed environments versus spatially-structured ones, where structure is imposed through the addition of semi-solid agar. In contrast to the well-mixed environment, populations that evolved in the spatially-structured environment adapted more slowly, retained an ability to move more rapidly, and had a greater degree of parallel evolution across replicates. These results confirm model predictions that spatial structure can drive evolution to occur in a more repeatable and predictable way, and highlight the importance of considering spatial structure in evolution experiments.

Room 13-09

Worldwide Population Structure of Human Head Lice: Insights from Whole Genome SNPs

Niyomi House[1,2], Aida T Miró-Herrans[3], Mercedes Domingo[4], Natacha V. Wirdemark[5], Lajos Rozsa[6], Muhammad Ashfaq[7], Kosta Y. Mumcuoglu[8], Jan Stefka[9], Ariel Toloza[10], Heinz Mehlhorn[11], Alejandra Perotti[12], Henk Braig[13], Didier Raoult[14], Oleg Mediannikov[15], Julie Allen[16], Bret Boyd[17], David L Reed[2]
[1]Department of Biology, University of Florida, Gainesville, FL, USA. [2]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA. [3]George Smathers Libraries, University of Florida, Gainesville, FL, USA. [4]Pid Free Clinic, Milano, Italy. [5]LusFri Inspagat AB, Stockholm, Sweden. [6]MTA-ELTE-MTM Ecology Research Group, Budapest, Hungary. [7]Centre for Biodiversity Genomics, Biodiversity Institute of Ontario, University of Guelph, Ontario, Canada. [8]Department of Microbiology and Molecular Genetics, Hebrew University of Jerusalem, Jerusalem, Israel. [9]Laboratory of Molecular Ecology and Evolution, Institute of Parasitology, Univ. S. Bohemia, Bohemia, Czech Republic. [10]Heinrich-Heine-Universität, Düsseldorf, Gibraltar. [11]National Scientific and Technical Research Council, Buenos Aires, Argentina. [12]School of Biological Science, University of Reading, Reading, United Kingdom. [13]School of Natural Sciences, Bangor University, Bangor, Wales, United Kingdom. [14]Faculty of Medicine, Aix-Marseille University, Marseille, France. [15]Institute of Research for Development, Marseille, France. [16]University of Nevada, Reno, USA. [17]Virginia Commonwealth University, Richmond, USA

Abstract

Primate lice are obligate parasites that have lived among their hosts for millions of years. Particularly, human head lice have coevolved with their host, mirroring aspects of human evolution. Their reduced genome sizes and history of coevolution makes them an ideal candidate to better understand host-parasite co-evolution and co-demography. To date, most studies have used only a few mitochondrial and microsatellite markers with reduced geographic representation to assess louse diversity. These have found deeply divergent, geographically structured mitochondrial clades, as well as population structure based on microsatellite markers. To overcome the limitations of past studies, we sampled and sequenced 377 head lice from 43 countries to investigate their diversity and population structure using Single Nucleotide Polymorphisms (SNPs) from the whole nuclear genome. The resulting ~30K high quality SNPs showed continental structure as well as substantial geographic substructure within continents. Louse populations within continents showed high genetic differentiation, with F_{ST} values almost an order of magnitude larger than human populations within-continent F_{ST} values. Similar to human genetic clustering, lice show a distinction between African and non-African individuals and show additional structure within non-Africans by continent. Our results suggests that head louse populations have retained high levels of isolation in contrast to their human hosts. Although the louse populations as a whole seem to be evolving under mutation-drift equilibrium, we found some loci that are putatively under positive selection. These loci provide interesting examples of selection pressures imposed by the host environment possibly due to insecticides used to treat head louse infestations.

Room 13-10

Insight into pelagophytes: novel algal genomes and strain level genome variation in the harmful algal bloom causing species *Aureococcus anophagefferens*

Shannon J Sibbald[1,2], Maggie Lawton[1,2], Andrew J Roger[1,2], John M Archibald[1,2]
[1]Dalhousie University, Halifax, NS, Canada. [2]Centre for Comparative Genomics & Evolutionary Bioinformatics, Halifax, NS, Canada

Abstract

The Pelagophyceae are marine stramenopile algae that include *Aureoumbra lagunensis* and *Aureococcus anophagefferens*, two microbial species notorious for causing harmful algal blooms. Despite their ecological significance, relatively few genomic studies of pelagophytes have been carried out. To improve understanding of the biology and evolution of pelagophyte algae, we produced new high-quality reference genomes using Oxford Nanopore long-read sequencing technology for *A. lagunensis* (CCMP1510) (43 Mbp), *Pelagomonas calceolata* (CCMP1756) (34 Mbp), and re-sequenced *A. anophagefferens* (CCMP1984) (53 Mbp). This includes a fully resolved telomere-to-telomere genome assembly for *P. calceolata*, with 6 chromosomes ranging from 4 to 6 Mbp in size. Furthermore, to investigate intra-species variation we produced high-quality draft genomes for four additional *A. anophagefferens* strains (CCMP1707, CCMP1708, CCMP1850, and CCMP3368). The pan-genome refers to the sum of genes across all strains of a given species, only a subset of which reside in the genome of any given strain. The pan-genome concept readily applies to prokaryotes, where lateral gene transfer (LGT) can lead to enormous intra-species gene-content variability. However, the extent to which LGT-driven pan-genomes exist in eukaryotes is uncertain. Our comparative genomic investigation indicates strain level variation in gene content in *A. anophagefferens* (pan-genome), including genes predicted to be related to bloom conditions, providing insight into both bloom dynamics and how microbial eukaryotes adapt and diversify.

Room 13-11

Evolutionarily young microproteins drive rapid functional adaptation

Aaron C Wacholder[1,2], Anne-Ruxandra Carvunis[1,2]

[1]Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [2]Pittsburgh Center for Evolutionary Biology and Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Microproteins are a recently discovered yet highly prevalent class of small proteins (<100 amino acids) increasingly recognized as key players in cellular biology. However, the evolutionary forces acting on microproteins remain poorly understood. Here, we systematically analyzed the evolution of microproteins in *S. cerevisiae* by integrating an abundance of published genomes from fungal populations with translation and phenotype data. We find sharp differences in evolutionary dynamics between microproteins and larger proteins. While nearly all large proteins are well-conserved, a majority of microproteins appear evolutionarily young, having recently evolved *de novo* from noncoding sequence. We demonstrate that this result is not caused by detection bias, as we can trace back the evolutionary history of most identified microproteins to inferred homologous noncoding ancestral sequences. Furthermore, we find that the majority of microproteins lack signatures of codon-level selection, even when analyzed as a group to increase power. Importantly, microproteins with well-established phenotypes, high expression levels, and characterized roles in established biological pathways also are evolutionarily young and lack codon-level selective signatures. To explain these results, we propose a model in which a pool of evolutionarily transient microproteins are continually lost and replaced in response to rapidly changing selection pressures. This model predicts that hundreds of transient microproteins participate in rapid functional adaptation.

Room 13-12

Epistatic coefficients correlation between Hadamard-Walsh and “thermodynamic” representations

María Carolina Erazo[1], Laura Aviñó Esteban[2], Dmitry N Ivankov[1]

[1]Skolkovo Institute of Science and Technology, Moscow, Russian Federation. [2]Amalfi Analytics, Barcelona, Spain

Abstract

The classical definition of epistasis centers on the interaction between different genes; the current understanding of epistasis includes also interactions between positions in protein, RNA, or DNA sequences. The most straightforward approach to detect epistasis is to calculate epistatic coefficients from combinatorically complete datasets, which form hypercubes in sequence space. Most commonly, researchers apply Hadamard-Walsh transformation to get epistatic coefficients; however, the “thermodynamic” transformation could also be applied. Despite the formal link between the two approaches has been established, the empirical relationship has not yet been explored.

Here we study the correlation between epistatic coefficients obtained by the Hadamard-Walsh and “thermodynamic” approaches using as an example the experimental fitness landscape of the WW domain. Most values of Pearson correlation coefficients (PCC) between the set of epistatic coefficients were positive. However, negative values down to -30% also occurred. For randomized landscape, PCC ranged from -56% to 99%.

Further, we proved mathematically that the highest positive PCC between Hadamard-Walsh and “thermodynamic” epistatic coefficients is 100%, which occurs not only in the trivial case of absence of epistasis but also in a non-trivial case(s). The perfect anti-correlation is not possible, which means that the particular value of the most negative correlation seems to depend on the fitness values distribution.

Overall, we show that the epistatic coefficients calculated by Hadamard-Walsh and “thermodynamic” formalisms could differ dramatically, and both representations require attention.

Room 13-13

A computational screen for alternative genetic codes in over 250,000 genomes

Yekaterina Shulgina[1], Sean R. Eddy[1,2]
[1]Harvard University, Cambridge, MA, USA. [2]HHMI, Chevy Chase, MD, USA

Abstract

The genetic code was proposed to be a “frozen accident”, but the discovery of alternative genetic codes over the past four decades has shown that it can evolve to some degree. Since most examples were found anecdotally, it is difficult to draw general conclusions about the evolutionary trajectories of codon reassignment and why some codons are affected more frequently. To fill in the picture of genetic code diversity, we developed a computational method that can predict the amino acid meaning of all 64 codons from nucleotide sequence data. Our approach identifies stretches of sequence homologous to conserved protein domains and then analyzes patterns of codon occurrence at conserved amino acid positions to predict the amino acid meaning of all 64 codons. We surveyed the genetic code usage of over 250,000 bacterial and archaeal genomes on GenBank and discovered five new reassignments of arginine codons (AGG, CGA, and CGG), representing the first sense codon alterations in bacteria. The existence of tRNA genes consistent with the new translations further supports the computational predictions and provides insight into the mechanism of reassignment. In a clade of uncultivated Bacilli, AGG has been reassigned from arginine to become the dominant methionine codon, outnumbering AUG in conserved protein domains. Reassignments of the arginine codons CGA and/or CGG were found in four clades with low genomic GC content, an evolutionary force which may have facilitated codon reassignment by driving these codons to low frequency prior to reassignment by synonymous substitutions with more AT-rich arginine codons.

Room 13-14

A reproducible workflow for variant calling and comparative population genomics in nonmodel organisms

Sara JS Wuitchik[1,2], Allison J Shultz[3], Brian J Arnold[1], Cade Mirchandani[4], Russ Corbett-Detig[4], Timothy B Sackton[1]

[1]Harvard University, Cambridge, MA, USA. [2]Boston University, Boston, MA, USA. [3]Natural History Museum, Los Angeles, CA, USA. [4]University of California Santa Cruz, Santa Cruz, CA, USA

Abstract

Over the past decade, there has been a dramatic increase in the amount of publicly available sequencing data from a large variety of organisms. While large-scale data reuse in model systems is common, reuse and reanalysis of nonmodel sequencing data are hampered by inconsistencies in sequence archive information, differences in data quality, and computational batch effects. We have developed a reproducible workflow for variant calling that is optimized for nonmodel organisms and comparative population genomic analyses. We applied this workflow to over 70 non-mammalian vertebrate species, to produce a publicly available resource of consistently generated variant calls for comparative population genomics. We developed a robust, high throughput software package and Snakemake workflows for extracting synonymous and nonsynonymous mutations to perform McDonald-Kreitman tests and identify lineage-specific signatures of positive selection. We apply this workflow to high-quality public datasets in birds in order to test hypotheses about the role of convergent selection across lineages in shaping genome evolution. To aid in visualization, we created genome browser track files for each species, hosted on an assembly hub and viewable with the UCSC Genome Browser. By generating a non-mammalian vertebrate database, this workflow provides the foundation to expand to other taxa and comparative approaches. There are many questions we can ask using multispecies resequencing data and this managed pipeline will facilitate the reuse of public data in a consistent, reproducible workflow.

Room 13-15

Evolutionary history of domain expansion and diversification in a family of essential fertilization genes

Alberto M Rivera, Damien B Wilburn, Willie J Swanson
University of Washington, Seattle, WA, USA

Abstract

Many proteins contain functional domains, which can act as individual discrete units. Whole gene duplications and tandem duplications of individual domains, can create new genes of varying repeat structure, which can evolve novel functions. This pattern is observed in the proteins comprising the egg coat (known as a Zona Pellucida or ZP in mammals). Many of the egg coat components contain a ZP module, which consists of a paired ZP-N and ZP-C protein domain, but some genes have additional N-terminal ZP-N domains. ZP-N domains are important for mediating species-specific fertilization and they show strong structural conservation, despite substantial sequence divergence. While previous work built phylogenies from whole ZP modules, we extensively curated ZP-N domains from both within and outside the ZP module, to understand the duplication and diversification history of the individual domain. Our large-scale phylogenies reveal a robust divide between modular ZP-N domains and free N-terminal ZP-N domains, which fall outside of the module. This suggests a single origin of free N-terminal ZP-N domains, with some having additional lineage specific duplications. To understand the significance of this evolutionary divide, we used machine learning methods to uncover a conserved set of amino acid residues and structural features in the modular domains. Tests for positive selection further demonstrated the sequence conservation of modular ZP-N domains, while identifying two free N-terminal ZP-N domains with sites under positive selection. While modular ZP-N domains seem to have an important conserved structure, duplicated N-terminal domains have been able to diversify and evolve species-specific functions.

Room 13-16

Genomic landscape of introgression between blood flukes infecting livestock *(Schistosoma bovis)* and humans *(S. haematobium)*

Roy N Platt[1], Grace-Ann Arya[1], Elisha Enabulele[1], Bonnie L. Webster[2], Muriel Rabone[2], Fiona Allen[2], Adian Emery[2], David Rollinson[2], Timothy JC Anderson[1]

[1]Texas Biomedical Research Institute, San Antonio, TX, USA. [2]Natural History Museum, London, United Kingdom

Abstract

Hybridization between human and animal parasites may transfer novel pathogenic traits between species, increasing virulence, host range and negatively impacting human health. Knowing when and how often these events occurs is an essential step in achieving optimal health outcomes within a One Health frame work. The human parasitic blood fluke, *Schistosoma haematobium* infects millions of people across sub-Saharan Africa. *S. bovis*, a sympatric and closely related species, parasitizes livestock. Initial genetic analyses with rDNA and mtDNA markers showed discordance: this was interpreted to indicate that these species readily hybridize in the field. Laboratory crosses between these species also suggested few reproductive barriers. However, recent studies using larger numbers of genetic markers including microsatellites, whole-exome, and whole-genome single nucleotide variants have failed to identify recent or ongoing hybridization between *S. bovis* and *haematobium*. Instead, it appears that hybridization between these *S. bovis* and *haematobium* occurred in the relatively distant past with subsequent selection on introgressed alleles in *S. haematobium*. In particular, an invadolysin gene of *S. bovis* origin appears to have reached near fixation in west African populations of *S. haematobium*, but is absent in east Africa. Here, we scored 9.6 million genome-wide, single nucleotide variants in 161 *S. bovis* and *haematobium* samples collected from 18 countries across the African continent. Our goals were to (1) estimate the number of historical introgression events within a biogeographic context, (2) more precisely date the intogression event(s), and (3) identify introgressed regions that may be experiencing positive or negative selective pressures in different populations.

Room 13-17

Microbial eukaryote communities among differently-aged pitchers on Nepenthes tropical pitcher plants

Jailene C. Gonzalez[1], Laura A. Katz[1,2], Adri K. Grow[1]
[1]Smith College, Northampton, Massachusetts, USA. [2]University of Massachusetts Amherst, Amherst, Massachusetts, USA

Abstract

The majority of eukaryotes are single-celled microbes, referred to as protists, many of which remain understudied because they cannot currently be cultivated. Here, we survey protist diversity within phytotelmata (water cavities, including pitchers) of Nepenthes tropical pitcher plants. Our focus is on the diversity of the SAR (Stramenopiles, Alveolates, and Rhizaria) clade that includes photosynthetic lineages (e.g. diatoms, dinoflagellates), parasites (e.g. apicomplexans, oomycetes), heterotrophs (e.g. ciliates, most Cercozoa), as well as many other interesting and uncultivable lineages. We use SAR-specific primers designed to amplify a portion of the SSU-rRNA gene to characterize community diversity in pitchers sampled from the Nepenthes pitcher plants at the Smith College Lyman Plant House and Conservatory. Pitchers were sampled from different life stages to investigate whether unopened pitchers are microbially sterile or, instead, are seeded with a microbial community from their parent plant or environment. We aim to compare differences in SAR communities among closed juvenile, recently-opened adult, and open mature pitchers. Our preliminary results suggest that juvenile unopened pitchers harbor a less abundant and sometimes undetectable SAR community while open adult and mature pitchers harbor well-established and diverse SAR lineages.

Room 13-18

Epistasis creates invariant sites in molecular evolution

Ravi Patel[1,2], Sudhir Kumar[1,2]

[1]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA.

[2]Department of Biology, Temple University, Philadelphia, PA, USA

Abstract

Substitution rate variation is commonly observed in multispecies protein sequence alignments. This variability is frequently modeled by a gamma (Γ) distribution of site-wise rates, along with a class of invariant sites (I+). Through computer simulated sequence evolution including epistasis among residues, we show that epistasis provides a mechanistic explanation for the excess of invariant sites beyond those explained by a Γ-rate model for ten protein domain families of diverse sequence, biochemical, and functional properties. To that end, the creation of a class of invariant sites is a primary emergent property of epistasis. This is reasonable because the intensity of negative selection against residue changes is expected to be proportional to the degree of epistatic couplings. The most highly coupled positions are expected to evolve very slowly and rarely permit single-site substitutions, thus creating invariant sites. We show that epistasis predicts a I+Γ distribution of site-wise rates at high sequence divergences, while invariant plus single (I+S) rate model will provide an adequate fit for lower sequence divergences. At extremely high sequence divergences, a Γ model will best fit the site-wise molecular evolutionary rates.

Room 13-19

Investigating protist diversity in New England tide pools

Adri K Grow[1], Robin S Sleith[1], Rabindra Thakur[1,2], Jailene C Gonzalez[1], Michaela Labare[1], Laura A Katz[1,2]

[1]Smith College, Department of Biological Sciences, Northampton, MA, USA. [2]University of Massachusetts Amherst, Program in Organismic and Evolutionary Biology, Amherst, MA, USA

Abstract

Tide pools are extreme environments as they experience substantial changes in abiotic factors over short periods of time making them ideal locations to study the patterns and drivers of microbial diversity. Here we survey the diversity of microeukaryotes, referred to as protists, found within tide pools in New England, focusing on the diverse SAR (Stramenopiles, Alveolates, and Rhizaria) clade. Estimates are that SAR represents about half of eukaryotic diversity, and this clade includes many photosynthetic lineages (e.g. diatoms, dinoflagellates), parasites (e.g. apicomplexans, oomycetes, labyrinthulomycetes), heterotrophs (e.g. ciliates, most Cercozoa), among many other interesting and uncultivable lineages. In this study, we use SAR-specific primers designed to amplify a portion of the SSU-rRNA gene to characterize community diversity in tide pools sampled in Acadia National Park, Maine. Our goal is to investigate trends in abundance, richness, and overall SAR diversity in these dynamic pools. To further understand the diversity found within tide pools, we sampled the near-shore ocean for comparison to identify tide-pool-specific organisms. Our preliminary analyses indicate that: (1) Alveolates, specifically ciliates, dominate our samples, (2) there are many tide-pool-specific organisms that are able to persist throughout tidal cycles, and (3) organism size, sampling season, and DNA (present organisms) vs. RNA (active organisms) drive differences in our samples. Together, these data illuminate the biodiversity of protists in extreme environments.

Room 13-20

Allele frequency dynamics under sex-biased demography and sex-specific inheritance in a pedigreed bird population

Felix EG Beaudry, Rose Driscoll, Nancy Chen
University of Rochester, Rochester, NY, USA

Abstract

Sex-biased demography, including sex-biased survival or migration, can impact allele frequency changes across the genome. In particular, we expect different patterns of genetic variation on autosomes and sex chromosomes due to sex-specific differences in life histories, as well as differences in effective population size, transmission modes, and strength and mode of selection. Here, we demonstrate the role that sex differences in life history play in shaping short-term evolutionary dynamics across the genome. We use a 25-year pedigree and genomic dataset from a long-studied population of Florida Scrub-Jays (Aphelocoma coerulescens) to directly characterize the relative roles of sex-biased demography and inheritance in shaping genome-wide allele frequency trajectories. We use gene dropping simulations to estimate individual genetic contributions to future generations and to model drift and immigration on the known pedigree. We quantify differential expected genetic contributions of males and females over time, showing the impact of sex-biased dispersal in a monogamous system. Due to female-biased dispersal, more autosomal variation is introduced by female immigrants. However, due to male-biased transmission, more Z chromosome variation is introduced by male immigrants. Finally, we partition the proportion of variance in allele frequency change through time due to male and female contributions. We find most allele frequency change is due to variance in survival and births. Males and females have similar contributions to autosomal allele frequency change, but males have higher contributions to change on the Z. Our work shows the importance of understanding sex-specific demographic processes in accounting for genome-wide allele frequency change in wild populations.

Room 14-01

A phylogenetic approach to amplicon-based taxonomic assignments in South American Diptera

Aina Martinez Zurita[1], Jacob Tennessen[2,3], Meg Shieh[3,1], Angela Early[1,3], Daniel Neafsey[3,1]
[1]Broad Institute, Cambridge, Ma, USA. [2]Broad Institute, Cambridge, MA, USA. [3]Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

Taxonomic assignment via targeted amplicon sequencing has been used to address many biological questions, from the characterization of microbial communities to tracking animal biodiversity and new species discovery. Despite decades of work on DNA barcoding, there is no straightforward bioinformatic pipeline to rapidly survey biodiversity, even for well-studied taxa like Diptera. As a result, sequence-based taxonomic assignments often depend on ad hoc approaches such as matching the sample sequence to an existing database with BLAST. We present an alternate phylogenetically informed approach that combines an alignment match with an assignment derived from a phylogeny with curated reference sequences. We identify Shimodaira–Hasegawa supported clades that contain the sample sequence, providing an assignment as well as a measure of confidence. Furthermore, we integrated our taxonomic assignment into an easy to use computational pipeline that comprehensively analyzes amplicon data, from QC to amplicon processing with DADA2 and taxonomic id. We apply our method to a set of 1100 Diptera samples collected in Guyana for a vector surveillance study. We Illumina-sequenced the internal transcribed spacer 2 (ITS2) region and a portion of the COX1 mitochondrial gene in all samples, then proceeded with our taxonomic assignment pipeline. Our approach readily assigns specimens to species or to narrow clades within Anopheles, Culex, Aedes with well-defined boundaries of taxonomic certainty in each case. In addition to validating our approach, these results characterize vector species abundance in a less well-studied world region. The method is designed to be target agnostic and could be extended to other sequences or species

Room 14-02

mCHH islands: a non-conserved but ubiquitous feature of plant genomes

Galen Martin[1], Brandon Gaut[1], Danelle Seymour[2]
[1]UC Irvine, Irvine, CA, USA. [2]UC Riverside, Riverside, CA, USA

Abstract

mCHH islands are peaks of CHH methylation that occur primarily upstream to genes. These regions are actively targeted by the methylation machinery, occur at chromatin boundaries, and tend to be near expressed genes. Here we took an evolutionary perspective by studying upstream mCHH islands across a sample of eight grass species. Using a statistical approach to define mCHH islands as regions that differ from genome-wide background CHH methylation levels, we demonstrated that mCHH islands are common and associate with 39% of genes, on average. We hypothesized that islands should be more frequent in larger genomes, because they have more heterochromatin. We found, however, that smaller genomes tended to have a higher proportion of genes associated with 5’ mCHH islands. Consistent with previous work suggesting that islands reflect the silencing of the edge of transposable elements (TEs), genes with nearby TEs were more likely to have mCHH islands. However, the presence of mCHH islands was not a function solely of TEs, both because the underlying sequences of islands were often not homologous to TEs and because other genic properties strongly predicted the presence of 5’ mCHH islands. These genic properties include length and gene-body methylation. In contrast, gene expression was a weak predictor of the presence of an island. Finally, we assessed whether mCHH islands were evolutionarily conserved by focusing on a set of 2,720 orthologs across the eight species. They were generally not conserved. Overall, our data suggest that mCHH islands are not just a consequence of near-gene TE silencing.

Room 14-03

Introgression biases incomplete lineage sorting at linked loci

Miriam Miyagi[1], Andrew J Blumberg[2], Nick Patterson[1], John Wakeley[1]
[1]Harvard University, Cambridge, MA, USA. [2]Columbia University, New York City, NY, USA

Abstract

We extend the Markov model of Slatkin and Pollack to model the joint distribution of gene trees at two linked loci sampled from three species with a pulse of introgression. Using this, we show how various summaries, such as concordance between the two trees, are affected by the introgression event and the interlocus recombination rate. We find that the probability two gene trees are concordant with each other is higher for introgressed loci than for those that follow the species tree, but the total concordance probability may be non-monotonic or decreasing with the introgression fraction. In addition, we describe how even if incomplete lineage sorting occurs, information about the species network topology is preserved in the joint distribution of gene tree frequencies, unlike at a single locus.

Room 14-04

Structural variation in *Drosophila santomea*

Brandon Turner[1], Theresa Miorin[2], Rebekah Rogers[1]
[1]University of North Carolina - Charlotte, Charlotte, NC, USA. [2]University of Georgia, Georgia, GA, USA

Abstract

The appearance of new genes through the reorganization of existing genetic material is a unique source of genetic variation within populations. Chromosomal rearrangements act as a source of this genetic variation and divergence by shuffling DNA throughout the genome. We used a population genomic approach to explore how these mutations reshape genetic diversity during local adaptation in *D. yakuba* and *D. santomea* on the island of Sao Tome. These closely related Drosophila species provide an excellent way to examine genome structure changes and their contribution to differentiation between island and mainland. We find 5,993 mutations with strong differentiation between populations and an increase in allele frequencies on the island. 3,337 mutations are associated with new TE insertions. Rearrangements on the X chromosome are common, and more likely to show differentiation between *D. santomea* and *D. yakuba*. We observe stronger differentiation from mainland populations in *D. santomea* than in island *D. yakuba*, consistent with the shorter timeframe since island invasion. These results suggest that genome structure changes, including those associated with TE movement, can reshape genetic diversity during local adaptation as organisms experience environmental shifts in nature.

Room 14-05

The evolution of the centromere-associated retrotransposon *G2/Jockey-3* in *Drosophila melanogaster* populations and *Drosophila simulans* clade species

Lucas Hemmer[1], Xuewen Geng[1], Sheif Negm[1], Eddyson Altidor[1], Jim Chaffer[1], Iain Speece[1], Cécile Courret[1], Chingo-Ho Chang[1,2], Amanda Larracuente[1]

[1]University of Rochester, Rochester, NY, USA. [2]Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Centromeres are chromosomal regions necessary for kinetochore attachment and cell division in eukaryotic organisms. Little is known about centromere organization because they reside in rapidly evolving, repeat- rich regions of the genome. The centromere islands of *D. melanogaster*, while unique to each chromosome, share one particular non-LTR retrotransposon named *G2/Jockey-3*. Studies in other organisms including plants, mammals, and fungi have detected centromere- associated retroelements suggesting a role in centromere function, maintenance, or establishment. However, *G2/Jockey-3* is not exclusive to centromeres in *D. melanogaster* and our work indicates natural selection does not act to conserve centromeric insertions. This suggests that *G2/Jockey-3* is acting like a selfish genetic element rather than an important component of centromeres. Centromeres evolve rapidly but it is unclear if centromere-associated transposable elements evolve in a similar manner or if they are conserved between species. We find *G2/Jockey-3* in all three species within the *simulans* clade, a sister group to *D. melanogaster*, and are unable to detect full-length elements in other species suggesting *G2/Jockey-3* is a relatively new element. Conservation within the sequence is limited to the open-reading frame while the 5’-end is highly divergent. We also discovered a diversification of *G2/Jockey-3* that is unique to the *simulans* clade where many of the copies are much younger than those in *D. melanogaster*. *G2/Jockey-3* is an active and dynamic element within the *simulans* clade in comparison to *D. melanogaster* and could have implications for centromere evolution in *Drosophila* and the role of retrotransposons in centromeres.

Room 14-06

Extreme Codon Usage Bias in Tubothalamid Foraminifera

Hannah B. Rappaport[1], Mattia Greco[1], Auden Cote-L'Heureux[1], Laura A. Katz[1,2]
[1]Smith College, Northampton, MA, USA. [2]University of Massachusetts Amherst, Amherst, MA, USA

Abstract

Compositional bias varies tremendously in eukaryotic genomes, and drives biases in codon usage that are most evident in the G+C content at third position of four-fold degenerate sites (i.e. GC3s). Here, we investigate the relationship between compositional bias and codon usage in Foraminifera, an ancient clade of amoebae with tests (i.e. shells). We assess codon usage in transcriptome data of ~30 species of Foraminifera. While species within "Monothalamea" (i.e. the non-monophyletic lineages with single-chambered tests) and the clade Globothalamea (i.e. multi-chambered 'globular' tests) have ~40-50% GC3-content, Tubothalamid lineages (i.e. Foraminifera with tubular tests) show extreme AT bias (0-10% at GC3s). In fact, codons ending in G or C are virtually absent from the 170 genes investigated thus far. We are working to refine these data in order to uncover the mechanisms (i.e. mutation bias vs. selection) that drive extreme codon usage bias in the Tubothalamea. Our results are indicative of potential differences in translation machinery among major clades of Foraminifera, and expand our knowledge of genome organization across the eukaryotic tree of life.

Room 14-07

A New Set of Old Genes: A Taxon-Rich Approach to the Identification and Analysis of Inter-Domain Non-Vertical Gene Transfer Events

Auden E Cote-L'Heureux[1], Xyrus Maurer-Alcala[2], Laura A Katz[1]
[1]Smith College, Northampton, MA, USA. [2]Institute of Cell Biology, University of Bern, Bern, Switzerland

Abstract

Vertical inheritance of genetic material is foundational to Darwinian evolution, but it fails to neatly explain important trends observed in evolution. For example, the important role of lateral gene transfer (LGT) in the evolution of prokaryotes has been widely accepted as a driving force behind adaptive change. Similarly, the transfer of organellar genes to the nuclear genome of eukaryotes, or endosymbiotic gene transfer (EGT), is well documented in the literature. Gene transfer into specialized eukaryotes (e.g. anaerobic parasites) is also relatively well-studied, but the lack of large-scale analyses of inter-domain LGT hinders further recognition. A notable critique of past studies of inter-domain LGT has been their reliance on deviation from the expected eukaryotic monophyly in gene trees. Here, we present an alternative approach to illuminating inter-domain LGT on a broad scale. We start with over 13,600 genes that are relatively conserved across the tree of life and apply a lineage-rich pipeline to identify over 350 putative inter-domain EGTs and LGTs. Taking a conservative approach and intensely curating our data to mitigate the effects of contamination and poorly-resolved single gene trees, we rely heavily on the exclusive presence of specific eukaryotic lineages in gene trees with diverse prokaryotes. We also evaluate the compositional bias and codon usage of LGTs, map them onto chromosomes for taxa with whole genome data, and characterize these genes by function across recipient categories. Our analyses expand the taxonomic scope of previously identified inter-domain LGTs and provide rigorous methodologies and criteria for future studies.

Room 14-08

Detecting species boundaries within uncultivable ciliate species by comparative analyses of single cell WGAs and WTAs

Ragib Ahsan[1,2], Wumei Blanche[2], Laura A. Katz[2,1]

[1]University of Massachusetts Amherst, Amherst, Massachusetts, USA. [2]Smith College, Northampton, Massachusetts, USA

Abstract

Ciliates, an ancient (>1 billion years old) clade of eukaryotes, are usual because they contain two distinct nuclei, a somatic macronucleus and germline micronucleus, within each individual. Because of this, the few cultivable ciliates have been studied as models for germline/soma differentiation. Recent advancements in single-cell 'omics, namely whole genome and whole transcriptome amplifications (WGAs and WTAs), all allow us to study the many uncultivable species. Here, we use a Python-based pipeline to compare WTAs and WGAs within species, with the aim of identifying germline-specific sequences that delineate species. Specifically, we focus on multiple individuals within the genera Chilodonella (Phyllopharyngea), Halteria (Spirotrichea) and Loxodes (Karyorelictea). These data will further our understanding of cryptic ciliates species samples from diverse freshwater and marine habitats.

Room 14-09

Characterization of alternative splicing in the ciliate genus *Spirostomum* (Heterotrichea) based on phylogenomic analyses of single-cell transcriptomic data

Shahed Uddin Ahmed Shazib[1,2], Mann Kyoon Shin[2], Laura A. Katz[3,4]

[1]Smith College, Northampton, Massachusetts, USA. [2]University of Ulsan, Ulsan, Korea, Republic of. [3]Smith College, Northampton, USA. [4]University of Massachusetts Amherst, Amherst, Massachusetts, Korea, Republic of

Abstract

Alternative splicing is common among eukaryotes and provides an additional mechanism for regulating gene expression and generating protein diversity. Intron retention is an alternative splicing mode that is well documented in animals, plants and fungi, but relatively little is known about this process in microbial eukaryotes. Ciliates are one of the most diverse clades of unicellular eukaryotes as they are important component of aquatic food webs; ciliates are also characterized by nuclear dimorphism where a transcriptionally active somatic macronucleus, and the germline micronucleus both develop from a zygotic nucleus. Here we focus on the genus *Spirostomum* (Cl: Heterotrichea), a model organism for applied ecology (i.e. assessment of water quality) and symbiosis research. In this study, we performed single-cell transcriptome amplification of multiple individuals from seven morphospecies of *Spirostomum*, including a new species. Using a phylogenomic pipeline, we characterize patterns of intron retention at species level. Our preliminary results indicate that *Spirostomum* species have tiny introns (15 bp) and that these introns are alternatively processed to generate proteins differing by a mere five amino acids. These data expand our understanding of intron retention across the eukaryotic tree of life.

Room 14-10

Odorant binding proteins in the *Drosophila* post-mating response

Nora C Brown[1], Benjamin J Gordon[1], Geoffrey D Findlay[1,2], Caitlin E McDonough-Goldstein[3], Andrew G Clark[1], Mariana F Wolfner[1]

[1]Cornell University, Ithaca, NY, USA. [2]College of the Holy Cross, Worcester, MA, USA. [3]Syracuse University, Syracuse, NY, USA

Abstract

During mating in many species, males transfer sperm in the ejaculate to females, as well as a suite of non-sperm components that includes seminal fluid proteins (SFPs). In *Drosophila melanogaster* and other insects, SFPs are essential for fertility and induce profound effects on female physiology and behavior post-mating. Genes encoding SFPs include some of the fastest evolving in the genome, likely the consequence of sperm competition and/or sexual conflict. The SFP suite in *D. melanogaster* includes several members of large gene families, including the odorant binding protein (Obp) family. Previous work in *Drosophila* has shown that some Obps are highly expressed in the antennae and mediate behavioral responses to odorants, potentially by carrying them to odorant receptors. However, the function of the seven seminal Obps remains uncharacterized. Using RNAi and CRISPR/Cas9 generated mutants, we examined the reproductive functions of these genes. Strikingly, we find males lacking Obp56g fail to induce the post-mating response in their mates. We further find that Obp56g is expressed in the male's ejaculatory bulb and is an important component of the mating plug in the female reproductive tract, where it promotes proper sperm storage, fertility, and the post-mating response. Comparative analysis of RNAseq data from multiple *Drosophila* species suggests that male reproductive tract expression of Obp56g is derived in a subset of species, indicating "co-option" of this protein for reproductive function over evolutionary time. Together, this work uncovers a novel role for Obps in reproduction and enhances our understanding of seminal fluid evolution.

Room 14-11

Reverse plasticity underlies rapid evolution by clonal selection within populations of fibroblasts propagated on a novel soft substrate

Purboja Purkayastha[1], Kavya Pendyala[1], Ayush S Saxena[2], Hesamedin Hakimjavadi[2], Srikar Chamala[2], Purushottam Dixit[2], Charles F Baer[2], Tanmay P Lele[1]
[1]Texas A&M University, College Station, Texas, USA. [2]University of Florida, Gainesville, Florida, USA

Abstract

Mechanical properties such as substrate stiffness are a ubiquitous feature of a cell's environment. Many types of animal cells exhibit canonical phenotypic plasticity when grown on substrates of differing stiffness, in vitro and in vivo. Whether such plasticity is a multivariate optimum due to hundreds of millions of years of animal evolution, or instead is a compromise between conflicting selective demands, is unknown. We addressed these questions by means of experimental evolution of populations of mouse fibroblasts propagated for ~90 cell generations on soft or stiff substrates. The ancestral cells grow twice as fast on stiff substrate as on soft substrate and exhibit the canonical phenotypic plasticity. Soft-selected lines derived from a genetically diverse ancestral population increased growth rate on soft substrate to the ancestral level on stiff substrate and evolved the same multivariate phenotype. The pattern of plasticity in the soft-selected lines was opposite of the ancestral pattern, suggesting that reverse plasticity underlies the observed rapid evolution. Conversely, growth rate and phenotypes did not change in selected lines derived from clonal cells. Overall, our results suggest that the changes were the result of genetic evolution and not phenotypic plasticity per se. Whole-transcriptome analysis revealed consistent differentiation between ancestral and soft-selected populations, and that both emergent phenotypes and gene expression tended to revert in the soft-selected lines. However, the selected populations appear to have achieved the same phenotypic outcome by means of at least two distinct transcriptional architectures related to mechano-transduction and proliferation.

Room 14-12

Genome-wide characterization of genetic variation for pollen expression

Meng Yuan, Stephen I Wright, John R Stinchcombe
University of Toronto, Toronto, Ontario, Canada

Abstract

The studies of plant evolutionary biology have been mostly focused on the diploid sporophytic phase of the lifecycle, while the haploid gametophytic phase remains largely neglected. With the common genome shared between gametophytes and sporophytes and potential scope for gametophytic selection, a central question arises as how does gametophytic selection affect the sporophyte? Theoretical work has suggested that antagonistic pleiotropy between gametophyte and sporophyte can generate balancing selection and maintain genetic variation. Empirical tests of the genomic prevalence of this antagonistic pleiotropy between life stages is lacking. Here, we examined whether genes with overlapping (more likely to be subject to pleiotropy) or limited expression in different life stages show different signatures of selection, to estimate the genomic prevalence of genes under balancing selection due to antagonistic pleiotropy between gametophyte and sporophyte. We used gene expression data in pollen, pollen tube and leaf, combined with neutral diversity statistics in *Rumex hastatulus* and found different patterns of diversity for pollen-biased, leaf-biased genes and genes with relative equal expression between leaf and pollen. This research furthers our understanding of the genome-wide characterization of plant gametophytic selection.

Room 14-13

Comprehensively mapping the landscape of 3D structural constraint in the human proteome

Bian Li[1], John A. Capra[2]
[1]Vanderbilt University, Nashville, TN, USA. [2]UCSF, San Francisco, CA, USA

Abstract

Quantification of patterns of protein-coding genetic variation within and between species is a cornerstone of evolutionary and functional analyses. However, current approaches for quantifying constraint on proteins either focus on individual sites or the whole protein, without accounting for the functional context of the sequence: 3D structure. Recent growth in databases of genetic variation and protein 3D structure enable the synthesis of protein spatial context into the estimation of site-level constraint.

Here, we describe a new framework, called COSMIS, for quantification of the constraint on genetic variation in 3D neighborhoods of each protein site based on a mutation-spectrum-aware statistical model of the expected number of variants. We define a comprehensive map of protein spatial constraint by applying COSMIS to the 3D distribution of >1.88 million human missense variants from gnomAD, covering 47% of all canonical human transcripts.

We demonstrate that the COSMIS score is accurate in predicting gene essentiality and variant pathogenicity. We further show that COSMIS performs significantly better than a range of 1D sequence-based metrics, such as the MTR score, while also providing biophysical insight into the potential functional roles of constrained sites. We then demonstrate the utility of the COSMIS framework by applying it to detect constrained sites in ion channels and predict the pathogenicity of recently characterized variants using custom-built homology models.

Looking forward, we anticipate that the structural landscape of constrained sites identified by COSMIS will facilitate interpretation of patterns of protein-coding constraint in human evolution and prioritization of sites for mechanistic or functional investigation.

Room 14-14

An ontogenetic qPCR analysis of non-visual opsin expression in *Anoplopoma fimbria*, a deep-sea fish with 36 opsin genes.

Hayley V Barnes, Niloufar Mokariasl, John S Taylor
University of Victoria, Victoria, British Columbia, Canada

Abstract

Light regulates many biological processes through light-sensitive proteins called opsins. Opsins have been extensively studied according to their roles in the visual system, however, the 'visual opsins' only account for a small portion of all opsins. Opsins with extraretinal roles are referred to as non-visual opsins. Many organisms, fish in particular, have large repertoires of these opsin genes expressed across many tissues and, as it stands, there is little explanation as to what their functions may be. We introduce sablefish (*Anoplopoma fimbria*) as a model for opsin research. Our survey of a draft genome, multi-tissue transcriptome, and ESTs showed that they have a large opsin repertoire (36 genes) despite the fact that they spend the vast majority of their lives in the aphotic zone. This observation suggests that a large opsin repertoire might be essential for sablefish development. To test the hypothesis that opsin expression is higher during the comparatively brief period of time when this species is exposed to light (ages 40 days to 2 years), we quantified expression of five paralogs from a well-studied non-visual opsin family, OPN4 (melanopsin), in a diversity of tissues across five developmental stages. These observations provide insight into the roles of development and light exposure on non-visual opsin expression in a deep-sea fish. Preliminary data show high levels of variation among OPN4 paralogs at all life stages, and low levels of variation for each opsin across approximately one year of development.

Room 14-15

Intercalary heterochromatin and other genome features drive pairing loss in interspecies Drosophila hybrids

James G Baldwin-Brown, Nitin Phadnis
University of Utah, Salt Lake City, UT, USA

Abstract

Homologous chromosome pairing is essential in eukaryotes, but difficult to understand mechanistically. Pairing occurs primarily during meiosis in most eukaryotes, but dipterans such as *Drosophila* also pair their somatic chromosomes. Reproducible patterns of lost somatic pairing in interspecific *Drosophila* hybrids are an opportunity to understand the mechanism of pairing. Understanding what makes some regions pair while others do not is an avenue toward understanding the molecular components that drive pairing generally. Hybrids of more diverged *Drosophila* species show a greater degree of pairing loss, making a causal relationship between somatic pairing and reproductive incompatibility possible.

We crossed *Drosophila melanogaster* and *Drosophila simulans*, then measured the genome-wide rate of pairing at high resolution using Hi-C. We found dramatic regions of high and low pairing. This reduction in pairing is unique to polytene tissues in hybrid individuals, and there is no signal of reduced pairing around inversion breakpoints in a hybrid cross, in opposition to existing hypotheses.

We mapped a complex landscape of high- and low- pairing regions of varying sizes and intensities. The narrow peaks are consistent with the "button" model of pairing, rather than the "zipper" model. We identified local pairing maxima and minima and correlated them with genome features such as intercalary heterochromatin. Future work will contrast hybridization-driven pairing changes with knockdowns of the pairing-influencing gene complex Condensin II, as well as knockdowns of known *Drosophila speciation* genes such as *Lhr*. We will quantify transvection (trans-control of genes by cis-enhancers

during pairing) genome-wide using allele-specific expression assays.

Room 14-16

Galactose-evolved *Saccharomyces cerevisiae* populations across 4,000 generations

Artemiza A Martinez, Andrew Conboy, Gregory I Lang
Lehigh University, Bethlehem, PA, USA

Abstract

All cells respond and adjust their growth and behavior depending on the specific nutrients in the environment. In yeast, as in most organisms, glucose is the preferred carbon source. When glucose is absent, and galactose is present, yeast turns on a set of GAL genes to catabolize this galactose. We asked how cells respond to long-term propagation in galactose. Experimental evolution has proven to be a valuable strategy for understanding the different mechanisms and pathways important for adaptation under specific conditions. Here, we perform a 4,000-generation evolution experiment using isogenic *Saccharomyces cerevisiae* populations to study adaptation in galactose media. Whole-genome sequencing of 48 diploid evolved populations revealed a complex interplay of de novo single nucleotide mutations, copy number variation, and ploidy changes that led to an increase in fitness in galactose as a sole carbon source. We find three candidate drivers of galactose adaptation: SNT1, SEC23, and PEP5. We reconstructed some of these alleles as haploids, heterozygous diploids, and homozygous diploids and performed competitive fitness assays in galactose and glucose sources.

Room 14-17

Estimating historic human generation times with the mutation spectrum

Richard J Wang[1], Jeffrey Rogers[2], Matthew Hahn[1]
[1]Indiana University, Bloomington, IN, USA. [2]Baylor College of Medicine, Houston, TX, USA

Abstract

Generation time is a key parameter linking population genetics, demographic history, and phylogenetics. Various attempts to estimate the interval between generations have been made from genetic data, through models that include some combination of molecular and recombination clocks. Human generation time is a topic of considerable anthropological interest and is speculated to have varied through developments in the course of human history, including for example, the rise of agriculture and civilization.

Here, we apply a novel technique to estimate the human generation time up to 200,000 years into the past using mutation spectrum data derived from the 1000 Genomes Project. Our model for the mutation spectrum relies on the parental age effects observed in human mutation studies. We find a recent increase in the human generation time and estimate a male generation time that is consistently higher than the female generation time.

Room 14-18

A comparative transcriptomic approach to identify novel and known genes involved in gonad development in sea lamprey

Tamanna Yasmin[1], Sara V Good[1,2], Margaret F Docker[1]
[1]University of Manitoba, Winnipeg, Manitoba, Canada. [2]University of Winnipeg, Winnipeg, Manitoba, Canada

Abstract

Lampreys are important model species for understanding early vertebrate evolution, yet the genetic basis of sex determination and the genes involved in gonadal development in lampreys are still poorly understood. RNA-Seq analysis was performed on sea lamprey gonads from 10 definitive females, 11 definitive males, two undifferentiated larvae, and five prospective males (i.e., after completion of ovarian differentiation but prior to testicular differentiation). A genome-guided *de novo* assembly pipeline was employed to create a comprehensive superTranscriptome for differential expression analysis and Orthofinder pipeline was implemented to identify putative orthogroups of sea lamprey genes in 10 other chordate species. Of the 19,489 transcripts retained for analysis, 28 male-specific, and 1,644 male-biased transcripts (~25% novel) were identified, while only 25 female-biased transcripts were observed, with only 28% of the sex-biased genes being assigned to orthogroups. Some female-biased genes were associated with ovarian expression, while others had testis-biased expression in later-evolving vertebrates. Female-biased genes included paralogues of *zona pellucida* (*zp*2, *zp3, zp4*), which were both highly expressed and found in nine different lamprey linkage groups, compared to four in higher vertebrates. Male-biased genes were enriched for sex determination and differentiation as well as testicular development pathways. Also, the transcripts exhibiting male-biased expression in spermatogenesis showed overlap with sex-biased genes expressed in prospective males. Overall, we found a large number of known and novel gene candidates for sea lamprey gonad development with modest orthology relationship with later vertebrates. This information will help us understand the process and genes involved in gonad differentiation and development.

Room 14-19

Molecular evolution of dolphin hearing genes: different pathways for echolocation in marine and freshwater environments?

Letícia Magpali, Mariana Freitas Nery, Lucas Freitas, Érica Martinha Silva de Souza, Elisa Karen da Silva Ramos
Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

Abstract

Echolocation, a high-frequency biosonar, is employed by toothed whales for navigation in various aquatic ecosystems with contrasting features. High frequency hearing is mediated by cochlear hair cells, which express several genes known to be evolving under positive selection in echolocating mammals. Increasing evidence suggests that multiple environmental constraints shaped echolocation variability among odontocetes, regarding sound emission and reception. Therefore, the echolocation systems of riverine, coastal, and marine odontocetes might have undergone different evolutionary pathways, resulting in distinct selective pressures and divergent/convergent evolution within hearing genes. We investigated the coding sequences of five hearing genes for a range (7 - 22) of cetaceans, including two Sotalia dolphins from newly sequenced whole genomes. Selection analyses on HyPhy and codeML recovered evidence of differential/positive selection among all toothed whales compared to other mammals, for genes CDH23 and SLC26A5. Branch-wide positive selection and selection intensification were restricted to marine lineages, with no evidence among river dolphins. Lineages under selection include the coastal/estuarine guiana dolphin, recently diverged from its freshwater sister tucuxi, the sperm whale, which produces a specialized sonar with hypothesized adaptations to deep diving, and the vaquita, the most endangered cetacean species. Sites under positive selection among toothed whales were identified in four genes, showing no substantial differences between marine and riverine branches. However, riverine, coastal and marine dolphins showed distinct substitution rates (dN/dS) in individual sites of TMC1 (2), SLC26A5 (5) and CDH23 (7), suggesting that divergent selective pressures acted among those groups in the evolution of hearing and echolocation.

Room 14-20

Identifiability of Species Network Topologies from Genomic Sequences

Hector Banos[1], John A. Rhodes[2], Elizabeth S. Allman[2]
[1]Dalhousie University, Halifax, NS, Canada. [2]University of Alaska, Fairbanks, AK, USA

Abstract

It is known that hybridization plays an important role during the evolutionary process of some species. Therefore phylogenetic trees are sometimes insufficient to describe species-level relationships. We show that most topological features of a level-1 species network are identifiable under the network multi-species coalescent model (NMSC) using the log-det distance between aligned DNA sequences of concatenated genes.

Room 15-01

Forecasting Viral Life History Traits through Endemicity or Vaccination

Nash D Rochman, Yuri I Wolf, Eugene V Koonin
NIH, Bethesda, MD, USA

Abstract

Unprecedented sequencing efforts have enabled the observation of the molecular evolution of SARS-CoV-2 in real time; however, the prediction of emergent variant characteristics remains extremely challenging. When faced with a new pathogen, it is highly desirable to be able to forecast the case fatality rate (CFR) into the future. Using a compartment model approach, we reveal general kinetic constraints among human pathogenic respiratory viruses where the variation of multiple parameters in concert leads to decreased virulence and increased pathogen fitness, whereas independent variation of these parameters decreases pathogen fitness. We show that highly virulent viruses, such as Smallpox, are likely often constrained by host behavior, whereas moderately virulent viruses, such as SARS-CoV-2, appear to be typically constrained by the relationship between the duration of immunity and CFR. When the immune population is rapidly expanded through vaccination, more specific, short-term predictions can be made in the face of dramatically altered selective pressures. Optimal vaccine distribution is constrained by the potential emergence of vaccine resistance. Analogous to chronic low-dose antibiotic exposure, recently vaccinated, partially immunized individuals play an outsized role in the emergence of resistance. We demonstrate when an escape variant is only modestly less infectious than the originating strain, there exists an optimal rate of vaccine distribution. Exceeding this rate increases the cumulative number of infections due to vaccine escape. Critically, modulating the rate of host-host contact for the recently vaccinated population by less than an order of magnitude can alter the cumulative number of infections by more than 20%.

Room 15-02

Genome variation through Mexican indigenous populations

Judith Ballesteros[1,2,3], Israel Aguilar[3,4], Fernando Perez[5,3], Humberto García[3], Francisco Barajas[3], Ram Gonzalez[2], Cristobal Fresno[3], Alejandro Garcíarrubio[5], Juan Carlos Fernández[3], Hugo Tovar[3], Enrique Hernández[3], Lorena Orozco[3], Xavier Soberón[3,5], Enrique Morett[5]
[1]CINVESTAV - LANGEBIO, Irapuato, Guanajuato, Mexico. [2]BUAP, Puebla, Puebla, Mexico. [3]INMEGEN, CDMX, CDMX, Mexico. [4]UNAM, Cuernavaca, Morelos, Mexico. [5]IBT UNAM, Cuernavaca, Morelos, Mexico

Abstract

In the present work, a bioinformatic tool was developed to allow the study of genomic variation in population-based projects, validating it with the complete genome data reported in the 100G-MX project, which has aimed to characterize native Mexican genomes to explore the component native by whole genome sequencing and seek to infer the demographic history of native Mexican ancestors, as well as biomedically relevant variants. A bibliographic review of these variants cataloged as biomedically relevant was made, the impact of the gene to which it and a summary of the associated phenotype with these variants. Interestingly, variants related to phenotypes were found within the Mexican population, such as increased body mass index, differences in cholesterol levels. Unfortunately, the majority of these studies were from European individuals and only one included individuals with American ancestry, which highlights the importance and need to have more projects for sequencing the genomes of native and mestizo populations that contribute to the development of Genomic Medicine in Mexico, for which it is expected that the development of this tool will facilitate the identification of variants of interest biomedical in future population projects.

Room 15-03

Gene family amplification facilitates adaptation in freshwater Unionid bivalve Megalonaias nervosa

Rebekah L Rogers[1], Stephanie L Grizzard[2], Katherine Bockrath[3], Sagar Patel[1], John P Wares[4], Jeffrey T Garner[5], Cathy C Moore[1]

[1]UNC Charlotte, Charlotte, NC, USA. [2]Old Dominion Univiersity, Norfolk, VA, USA. [3]US Fish and Wildlife Services, Onalaska, WI, USA. [4]University of Georgia, Athens, GA, USA. [5]Alabama Department of Conservation and Natural Resources, Florence, AL, USA

Abstract

As organisms are faced with intense rapidly changing selective pressures, new genetic material is required to facilitate adaptation. Among sources of genetic novelty, gene duplications and transposable elements (TEs) offer new genes or new regulatory patterns that can facilitate evolutionary change. With advances in genome sequencing it is possible to gain a broader view of how gene family proliferation and TE content evolve when populations become threatened. Freshwater bivalves (Unionidae) currently face severe anthropogenic challenges. Over 70% of species in the United States are threatened, endangered or extinct human encroachment. We have created a reference genome for M. nervosa to determine how genome content has evolved in the face of these widespread environmental challenges. We observe a burst of recent transposable element proliferation causing a 382 Mb expansion in genome content. Gene family expansion is common, with a duplication rate of 1.16 x 10^-8 per gene per generation. Cytochrome P450, ABC transporters, Hsp70 genes, von Willebrand proteins, chitin metabolism genes, mitochondria eating proteins, and opsin gene families have experienced significantly greater amplification and show signatures of selection. We identify signatures of strong, recent selective sweeps reshaping genetic diversity for large swaths up to 1Mb in this species, including genes related to detox, shell formation, and parasitic larvae. Hence, we suggest that gene family evolution is a source of "hopeful monsters" within the genome that facilitate adaptation under ecological upheaval. These indicator species can serve as a case study to frame expectations for other freshwater organisms experiencing anthropogenic challenges.

Room 15-04

Reconstructing the adaptive history of a functional variant from a spatially distributed wild *Arabidopsis* population

Ahmed F Elfarargi[1], Andrea Fulgione[1], Célia Neto[1], Nina Küppers[1], Elodie Gilbault[2], Andreas P.M. Weber[3], Olivier Loudet[2], Angela Hancock[1]

[1]Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. [2]Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France. [3]Institute of Plant Biochemistry & CEPLAS Plant Metabolism and Metabolomics Laboratory, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

Abstract

Learning how wild plants adapted to their local environments can enable us to forecast the genomic modifications necessary to tailor crop plants to specific locales. Plant stomata govern CO_2 uptake for photosynthesis and transpiration, determining plant water use efficiency (WUE) and hence plant productivity. Plant stomatal adjustments are crucial to balance the need for photosynthesis with desiccation avoidance, especially in water-limited environments. *Arabidopsis* from the Cape Verde Islands face prolonged dry periods with limited and highly variable rainfall, where the main source of precipitation comes from humid trade-winds. To learn how plants adapt to this unique precipitation regime, we examined the genetic architecture and evolutionary history of variation in stomatal conductance (SC) and WUE. We conducted phenomics for SC and WUE under highly controlled mild drought and well-watered conditions in a panel of Cape Verdean and Moroccan outgroup lines. Genome-wide association mapping revealed a nonsynonymous SNP that severely reduces the function of MAP KINASE 12 (*MPK12* G53R) and explains 57% and 66% of the variation in SC and WUE, respectively. We reconstructed the spatially explicit evolutionary history of *MPK12* G53R and inferred that this allele increased in frequency in the island due to positive selection as *Arabidopsis* expanded into the drier regions of the island. These findings advance our understanding of the evolutionary dynamics impacting a functional variant across space and time in its natural environment. They further suggest that a simple genetic change (loss of *MPK12* function) could enhance productivity in crop species in similar local environments.

Room 15-05

Convergent molecular evolution in visual systems of deepwater lake fishes

Alexander Van Nynatten[1], Nathan R Lovejoy[1], Nicholas E Mandrak[2], Belinda SW Chang[3]
[1]University of Toronto Scarborough, Toronto, ON, Canada. [2]University of Toronto Scarborough, Toronto, Toronto, Canada. [3]University of Toronto, Toronto, ON, Canada

Abstract

The evolution of fish vision is known to be strongly influenced by the depth-dependent reduction in the of the visible spectrum of light. As a result, spectral tuning substitutions in rhodopsin, the dim-light sensitive visual pigment, are common among fishes inhabiting deepwater environments. Here we investigate the phylogeography of spectral tuning substitutions in the rhodopsins of deepwater fishes inhabiting lakes in previously glaciated North America. We find substitutions at spectral tuning sites in populations inhabiting deep and clear lakes that blue-shift rhodopsin spectral sensitivity, matching the underwater light environment. We also conducted more targeted sampling of deepwater fishes from Lake Huron which has recently undergone a dramatic shift in water clarity as a result of invasive zebra mussels. We found spectral tuning substitutions only in individuals collected in deepwater trawls and at very low frequency, consistent with the depth-dependent attenuation of light. We further developed and evaluated a metabarcoding approach for more efficient monitoring of rhodopsin variants in Lake Huron. Our approach was effective in returning accurate estimations of the relative frequencies of different rhodopsin sequence variants from bulk DNA and tissue samples comprised of up to 32 individuals. In investigating the phylogeographic distribution of adaptive variation in rhodopsin we have identified important evolutionary differences in populations relevant to the conservation of biodiversity in Canadian fishes. Lastly, we demonstrate that metabarcoding can be a powerful tool for identifying functional adaptive variation in ecologically relevant traits and can be used to inform conservation and restorations efforts.

Room 15-06

The transmissibility of major SARS-CoV-2 clades through time

Damien Richard[1,2], Liam P Shaw[3], Rob Lanfear[4], Mislav Acman[1], Christopher J Owen[1], Cedric CS Tan[1], François Balloux[1], Lucy van Dorp[1]

[1]UCL Genetics Institute, University College London, London, United Kingdom. [2]Division of Infection and Immunity, University College London, London, United Kingdom. [3]Department of Zoology, University of Oxford, Oxford, United Kingdom. [4]Department of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, Australia

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in late 2019 and spread globally to cause the COVID-19 pandemic. Despite the constant accumulation of genetic variation in the SARS-CoV-2 population, there was little evidence for the emergence of significantly more transmissible lineages in the first half of 2020. Around November 2020, several more contagious and possibly more virulent ‘Variants of Concern’ (VoCs) were detected near-simultaneously in various regions of the world. These VoCs share some mutations and deletions that haven arisen recurrently in distinct genetic backgrounds. Here, we build on our previous work modelling the association of mutations to SARS-CoV-2 transmissibility and characterise the contribution of individual recurrent mutations and deletions to estimated viral transmissibility. We estimate enhanced transmissibility associated to mutations characteristic of VoCs and identify a tendency for C to T substitutions to lead to a reduction in estimated transmissibility. We then assess how patterns of transmissibility in all SARS-CoV-2 clades have varied over the course of the pandemic by combining our estimates in a simple multiplicative multilocus model based on the mutations carried by any sequenced genome. Such an approach recovers 501Y.v1 (B.1.1.7) as the most transmissible clade currently in circulation. By assessing transmissibility over time, we observe a tendency for estimated transmissibility to decrease for most clades analysed, consistent with a decay in fitness in mainly non-recombining lineages caused by the accumulation of weakly deleterious mutations. This pattern may, at least in part, explain the turnover of global viral lineages observed over the pandemic so far.

Room 15-07

Introgression and adaptation to deserts in the African golden wolf, a resilient top predator

Carlos Sarabia[1], Shyam Gopalakrishnan[2], Bridgett vonHoldt[3,4], Jennifer Leonard[1]
[1]EBD-CSIC, Sevilla, Spain. [2]GLOBE Institue, University of Copenhagen, Copenhagen, Denmark. [3]Faculty of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, USA. [4]Lewis-Sigler Institute, Princeton, New Jersey, USA

Abstract

Animals living in hot and arid environments receive strong abiotic selective pressures, which are expected to increase in the next decades with climate change. In subtropical deserts high rates of transpiration, limited access to food and water and high UV radiation are important drivers of adaptation in mammals. This is especially important in the case of top predators, whose relatively smaller effective population sizes make selection less effecient. Most studies of desert-adapted animals using genomics have been focused on camelids, wild ungulates and domestic goats. Here we used a set of African golden wolf (*Canis lupaster*) genomes and detected a large number of candidate genes under selection. Most significant GO categories and pathways were related to homeostasis, urine concentration, fat and sugar storage and metabolism, heat shock proteins and UV damage repair, some of which are possibly introgressed from gray wolves (*Canis lupus*) from the Middle East. In particular, some key genes under selection could have accumulated nonsynonymous mutations that dramatically alter protein function. Our results indicate that African golden wolves show signatures of adaptation to hot and arid environments and present key features that ensure their future survival atop trophic chains across extensive landscapes in north Africa.

Room 15-08

Selection against codons prone to nonsense mutations in prokaryotic genes

Dorota Mackiewicz[1], Kuba Nowak[2], Julia Muniak[1], Monika Grzybowska[1], Paweł Mackiewicz[1], Paweł Błażej[1]

[1]Department of Genomics, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland. [2]Faculty of Mathematics and Computer Science, University of Wrocław, Wrocław, Poland

Abstract

It is well known that the non-random synonymous codon usage in genes is very common phenomenon in many genomes. The main reason of this codon bias is the selection for the effectiveness of translation. The codon usage is also under the influence of mutational pressure associated with GC/AT content or selection for amino acid composition of coded proteins. It has been also suggested that the usage of preferred codons can minimize the rate of missense or nonsense errors and, in consequence, protect against a premature termination of synthetized proteins. The understanding of this phenomena seems particularly important because nonsense mutations are the molecular basis of many genetic diseases. It was found that there is a selection in eukaryotic genomes for avoidance of dangerous codons that can be changed into a stop codon by single substitution. However, analyses for prokaryotic genes are ambiguous. Here, we studied the distribution of dangerous codon usage in three regions of genes: 5’ region, middle and 3’ region, from 1207 bacterial and archaeal genomes. We noticed that the dangerous codons are significantly less preferred in the middle of the genes than in their beginning and their end in more than 76% of the genomes. The results did not depend on the length, function of gene and CG content in genes. The avoidance of the dangerous codons in the middle part of genes and their more common occurrence in other regions can be associated with a trade-off between translational costs and the length of synthetized proteins.

Room 15-09

Adaptation under the sea: the genomics of ecological adaptation in the Northern krill (*Meganyctiphanes norvegica*)

Per Unneberg[1], Chiara Papetti[2], Katharina Michael[3], Bettina Meyer[4], Andreas Wallberg[1]
[1]Uppsala University, Uppsala, Sweden. [2]University of Padova, Padova, Italy. [3]ICBM, University of Oldenburg, Oldenburg, Germany. [4]Alfred Wegener Institute, Bremerhaven, Germany

Abstract

Ocean warming has strongly impacted plankton communities, putting important ecosystems at risk. It is still unclear if genetic adaptation in zooplankton can keep pace with continued change and what genetic mechanisms promote survival in warmer oceans. This limits our ability to model the future distributions of zooplankton.

Krill are crustacean keystone macrozooplankton with huge populations and notoriously large and repetitive genomes. We assembled the 19.1 Gbp genome of the Northern krill, a widespread species native to both warm and cold environments across North Atlantic and Mediterranean waters. We used Nanopore long-reads, linked-read and RNA-seq for assembly, polishing, scaffolding and annotation to produce a draft genome with scaffold N50=220kbp, recover 20,000 complete genes (BUSCO: 80% complete genes, with little duplication) and find that ~70% of the genome is repetitive.

We re-sequenced 70 specimens from different latitudes and thermal regimes. We detected 280 M SNPs across all contigs with genes and estimate nucleotide diversity to 1.2%, intermediate among arthropods despite being superabundant. Population differentiation is low (FST~0.03), suggesting extensive gene flow. Comparing variation between lower and higher latitudes, we detect outlier SNPs across ~200 genes. Top selective candidates include genes that govern circadian rhythm and heat-shock responses, implying functions involved in photo-periodic expression and thermal tolerance to be targets for local adaptation and possibly important under climate change. We find no fixed alleles, suggesting selection on widespread standing variation.

We hope these findings will help develop genetic tools for monitoring and forecasting the status and distributions krill stocks in the future.

Room 15-10

Quantifying protein diversity in quickly evolving genomes

Erik Lavington, Siobain Duffy
Rutgers University, New Brunswick, NJ, USA

Abstract

While RNA and ssDNA virus population are often described as ‘diverse,’ there has been limited quantification of this diversity to compare groups of viruses over all but short evolutionary timescales (i.e., before measures of nucleotide diversity become saturated). Given the high nucleotide substitution rates of many viruses, logically we should focus on the level of amino acids. Indeed, several measures of protein diversity have been developed and implemented. However, these methods are used largely to measure diversity along the primary protein structure to predict attributes of the secondary, tertiary, and quaternary structure. We would like to make quantitative comparisons of protein diversity between different proteins within a taxonomic group and between different taxa. Here, we explore the behavior of Shannon Entropy (SE), Von Neumann Entropy (VNE), and DIVAA measures, as well as average pairwise differences (APD) using amino acid identity and amino acid similarity utilizing values from an amino acid substitution matrix, on the biallelic-locus level and at the full-protein level with data from HIV-1. Based on theoretical maxima of these measures over possible allele counts, we show that DIVAA and VNE measures are driven mostly by allele counts at any frequencies. Using empirical data, we show that SE and APD (by identity and by similarity) are more similar to each other ($\rho=0.918,0.925,0.999$, $\rho<0.883$ otherwise) and DIVAA and VNE are more similar to each other ($\rho=0.978$, $\rho<0.883$ otherwise). We suggest that an extension of average pairwise identity by incorporating biologically realistic amino acid substitution matrices could be useful in quantitative comparisons between proteins within and between taxa and that this measure may be helpful in better understanding evolvability in quickly evolving genomes.

Room 15-11

How Does Regulatory Evolution Differ Between Related Groups? Diverse, Repeatable, and Undiscovered Routes to Restore Motility Through Mutations in Gene Regulatory Networks in Two Immotile Bacterial Strains

Louise M Flanagan, James Horton, Tiffany B Taylor
University of Bath, Bath, United Kingdom

Abstract

Gene regulatory networks are essential to organism survival as they allow rapid adaptation to changing environments. But how do these networks evolve? And are there predictable mutational biases in their evolution? Removing a key component of a regulatory network and testing how an organism adapts can help us answer this. Deletion of the master regulator (FleQ) for flagellar synthesis genes renders *Pseudomonas fluorescens* strains (Pf0-1 and SBW25) immotile, but after being put under strong selection to swim, they reliably regain motility after a few days, through indirect co-option of a distantly related master regulator, NtrC, which controls the nitrogen regulatory pathway. However, the mutations that restore motility differ markedly between strains and across nutrient environments.

In Pf0-1, a huge variety of mutations were observed across a range of loci. In a complex nutrient environment (LB) 55% of restorative mutations occur in *glnA*, a component of the nitrogen regulatory pathway. In a minimal nutrient environment (M9) 54% of these mutations occur in another nitrogen regulatory gene, *ntrB*. These findings contrast strongly with SBW25, where the same motility-restoring SNP (A289C *ntrB*) is repeatedly found across all environments. Interestingly, this SNP has never been seen in Pf0-1. When it is engineered into Pf0-1 however, swimming is restored.

These results highlight how mutational biases can vary across environments, that not all viable routes may be uncovered, and that strain-to-strain differences can have a major impact on types of mutations that are uncovered and selected for by evolution.

Room 15-12

Role of ancient duplicates in the metabolic switching in *Saccharomyces cerevisiae*

Beatriz Sabater-Munoz[1], Florian Mattenberger[1,2], Mario A. Fares[3,4], Christina TOFT[5]

[1]Integrative and Systems Biology lab, Department of Plant stress, Institute of Plant Molecular and Cellular Biology (IBMCP CSIC-UPV), Valencia, Spain. [2]Viral Biology, Program for Systems Biology of Molecular Interactions and Regulation, Institute for Integrative Systems Biology (I2SysBio CSIC-UV), Paterna, Spain. [3]Posthumous author, Dublin, Ireland. [4]Smurfit institute of genetics, Trinity College Dublin, Dublin, Ireland. [5]ePanomics and Evolutionary Systems Microbiology lab, Program for Systems Biology of Molecular Interactions and Regulation, Institute for Integrative Systems Biology (I2SysBio CSIC-UV), Paterna, Spain

Abstract

Gene duplication events have been associated with increasing biological complexity throughout the tree of life, but also with illnesses, such as cancer. Early evolutionary theories indicated that duplicated genes could explore alternative functions due to the relaxation of selective constraints in one of the copies, as the other remains an ancestral-function backup. In unicellular eukaryotes like yeasts, it has been demonstrated that the fate and persistence of both duplicated copies in the genome depend on the duplication mechanism (whole-genome or small-scale events). Although it has been shown that small-scale duplicates tend to innovate and whole-genome duplicates specialize in ancestral functions, the implication of ancient duplicates' transcriptional plasticity and transcriptional divergence on environmental and metabolic responses remains largely obscure. Here we subject *Saccharomyces cerevisiae* to a metabolic switch by enforcing acute and chronic growth on a non-fermentative carbon source (ethanol) unrevealing the central role, the ancient duplicates have in metabolic shifts. In particular, the duplicates respond by transcriptional rewiring, depending on their transcriptional background. Our results shed light on the mechanisms that determine the role of duplicates, and on their continued evolvability.

Room 15-13

Population structure and selection among combatants of a Bronze Age battlefield

Vivian Link[1,2], Jens Blöcher[3], Joachim Burger[3], Daniel Wegmann[1,2], Zoé Pochon[1,2]
[1]University of Fribourg, Fribourg, Switzerland. [2]Swiss Institute of Bioinformatics, Fribourg, Switzerland.
[3]Johannes Gutenberg University Mainz, Mainz, Germany

Abstract

Numerous weapons, human and horse remains have been discovered along the Tollense River in northeastern Germany. Dating from about the 13th century BC, these findings are interpreted as the result of a single event, probably a battle as indicated by the large number of visible wounds on the bones. To determine whether the combatants formed genetically distinct groups, we analysed a sample of 19 individuals. Our results suggest that they came from the same population clustering around central and northern Europe. Since having a sufficiently large sample of one population at a given time point in history is a rare opportunity, we decided to use these data to measure the coefficient of selection of c. 450 loci of interest, using Central Europeans (CEU) individuals from the 1000 Genomes project as a second time point (1000 Genomes Project Consortium, 2010). We found strong and significant selection at two loci associated with lactase persistence. This shows that natural selection had a strong effect on these variants long after the Neolithic and the advent of agriculture.

Room 15-14

Tracing the evolution of human gene regulation and its association with shifts in environment

Laura L Colbran[1,2], Iain Mathieson[2], John A Capra[1,3]
[1]Vanderbilt University, Nashville, TN, USA. [2]University of Pennsylvania, Philadelphia, PA, USA.
[3]University of California, San Francisco, San Francisco, CA, USA

Abstract

As humans populated the world, they experienced shifts in many environmental factors, including climate, diet, and lifestyle. Traits relevant to these environmental shifts, like metabolism and skin pigmentation, have experienced recent selection. However the precise mechanisms underlying their evolution remain poorly understood, especially traits affected by changing gene regulation. We adapted PrediXcan, which imputes gene regulation from genotype data, to apply to low-coverage ancient human DNA. First, we performed a transcriptome-wide scan to identify regulatory differences between ancient populations with hunter-gatherer, pastoralist, or agricultural lifestyles. Predicting gene regulation in 49 tissues in 490 ancient Eurasians, we identified dozens of genes with significant regulatory differences among the three groups, which were enriched for immune and metabolic genes. Our results suggest explanations for previously observed signals of selection, e.g., in *FADS1*, *GPX1*, and *LEPR*. We next quantified gene regulatory changes associated with known changes in skin pigmentation over time. We trained PrediXcan models in melanocytes and applied them to a time series of 2999 ancient Europeans spanning ~38,000 years. Most pigmentation genes show no regulatory change over time, suggesting that adaptation mainly involved large-effect coding variants. The exception to this is *TYR*, where gene regulatory changes may buffer the functional effect of a coding variant. Overall, this work demonstrates a novel methodology for studying the interaction of gene regulation, ancient population structure, and environmental change. Our results identify systems influenced by gene regulatory changes in recent evolutionary history and highlight the potential impact of these differences on present-day complex traits.

Room 15-15

Reconstruction the evolutionary history of the tundra vole (Alexandromys oeconomus) based on mitochondrial DNA and radiocarbon dating in the context of climate changes in the late Pleistocene and Holocene.

Aleksandra Żeromska[1], Mateusz Baca[2], Danijela Popović[2], Anna Lemanik[3], Tatyana Fadeeva[4], Alexander Agadzhanya[5], Ivan Horáček[6], Sara Rhodes[7], Nicolas Conard[7], Magdalena Krajcarz[8], Emmanuel Desclaux[9], Aurélien Royer[10], Leonid Rekovets[11], Natalia Serdyuk[5], Adam Nadachowski[3], Paweł Mackiewicz[1]

[1]Faculty of Biotechnology, University of Wrocław, Fryderyka Joliot-Curie 14a, Wrocław, Poland. [2]Centre of New Technologies, University of Warsaw, Banacha 2c, Warszawa, Poland. [3]Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Sławkowska 17, Kraków, Poland. [4]Mining Institute, Ural Branch, Russian Academy of Science, Sibirskaya, 78a, Perm, Russian Federation. [5]Borissyak Paleontological Institute, Mammalogy Laboratory, Russian Academy of Sciences, Profsoyuznaya ul. 123, Moscow, Russian Federation. [6]Department of Zoology, Charles University, Viničná 7, Prague, Czech Republic. [7]Institut für Naturwissenschaftliche Archäologie, Universität Tübingen, Tübingen, Germany. [8]Institute of Archaeology, Faculty of History, Nicolaus Copernicus University, Szosa Bydgoska 44/48, Toruń, Poland. [9]Laboratoire départemental de Préhistoire du Lazaret, CEPAM, UMR 7264 CNRS, Nice, France. [10]Biogéosciences, UMR 6282, CNRS, Université Bourgogne Franche-Comté, 6 Boulevard Gabriel,, Dijon, France. [11]University of Environmental and Life Sciences, C. K. Norwida 25, Wrocław, Poland

Abstract

Ancient DNA (aDNA) isolated from remains of organisms living thousands of years ago is an invaluable source of information about both extant and extinct species. The studies of aDNA enable us to place genetic results in chronological context and compare them with climatic changes during the Pleistocene and the Holocene. We used this approach to reconstruct the evolutionary history of the European populations of tundra vole (*Alexandromys oeconomus*). This species had a wide geographic distribution in the Late Pleistocene, nowadays it occurs in wet grasslands from Northern Europe eastwards to Northern and Central Asia, as well as Alaska and Canada. We examined the 142 complete mitochondrial genomes obtained from paleontological specimens from various periods and sites. Our research shows that samples from Central Asia and the Ural Mountains are representative of the earliest diverged lineages, which corroborates the hypothesis that Asia was the ancestral region of this species evolution. The Asiatic populations spread to the east giving rise to the European lineages. One of them was distributed in Central Europe but, before the Last Glacial Maximum, migrated eastwards and replaced the lineage that inhabited the area of the Ural Mountains. In turn, another European lineage was initially present mainly in Western Europe but next expanded to Central Europe partially replacing the native populations. Our results indicate intensive movements and replacements of tundra vole populations across the North and East European Plains. These movements were most likely driven by the climatic and environmental changes during the Late Pleistocene.

Room 15-16

A flexible snakemake pipeline to map ancient DNA data

Samuel Neuenschwander[1,2], Diana Ivette Cruz Dávalos[1,2], Lucas Anchieri[1,2], Anna Sapfo Malaspinas[1,2]
[1]University of Lausanne, Lausanne, Switzerland. [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

Abstract

Ancient DNA is degraded and contaminated. Most standard bioinformatics tools to align sequenced reads have been designed for modern data and cannot be used “out of the box” to accommodate the features typical of ancient DNA. In this work, we propose a robust pipeline to align ancient DNA data and to have a first rough idea of the authenticity of the data. The implemented modules consist of the steps needed to go from a simple ‘fastq' file to a final ‘bam' file. The steps include quality control, mapping, filtering, duplicate removal, damage pattern inference, realignment, statistics to assess authenticity and inference of the sex of the organism. A final graphical report summarizes the statistics of the different modules allowing a quick overview of the data. The pipeline is implemented in the workflow manager snakemake providing the flexibility to run the pipeline on a workstation, a high memory server or a cluster. The installation of the pipeline and the underlying programs is made easy using conda. The pipeline may be used out of the box, or may be adapted with little knowledge of python and snakemake.

Room 15-17

Benchmarking methods using time-series data to infer selection

Lucas Anchieri, C. Eduardo Guerra Amorim, Anna-Sapfo Malaspinas
University of Lausanne, Lausanne, Switzerland

Abstract

With the increasing availability of ancient DNA (aDNA) data, it has become possible to reconstruct sampled allele trajectories in human populations over time (time-series data). These allele trajectories can be used to estimate selection on specific loci within a population. In order to do that, likelihood and Bayesian methods have been developed. Although these methods have been validated at the time of publication, few have been compared for realistic ancient DNA-like scenarios. Here, we assess the performance of those methods through simulations matching the characteristics of ancient DNA studies, in particular the reduced sample size. In this preliminary study, we simulate allele trajectories under parameters realistic for human populations, and test the methods using sampled time-series data. We consider positive and balancing selection and vary the age of the alleles, initial allele frequency, selection coefficient, and dominance coefficient to determine under which conditions specific methods perform better.

Room 15-18

Imputation of ancient genomes

Bárbara Sousa da Mota[1,2], Martin Sikora[3], Simone Rubinacci[1,2], Eske Willerslev[3], Anna-Sapfo Malaspinas[1,2], Olivier Delaneau[1,2]

[1]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [2]Swiss Institute of Bioinformatics, University of Lausanne, Lausanne, Switzerland. [3]Centre for GeoGenetics, Natural History Museum of Denmark and Department of Biology, University of Copenhagen, Copenhagen, Denmark

Abstract

After death and without repair mechanisms, DNA undergoes extensive damage, including fragmentation and C-to-T substitutions. Therefore, ancient genomes often have low sequencing depth, which complicates their analysis. Imputation has been proposed as a solution to this problem. However, it is unclear whether low-coverage ancient genomes can be accurately imputed and how imperfect imputation affects downstream analyses. To address these two questions, we downsampled 43 high-coverage ancient human genomes (>10x) to low-coverage and subsequently imputed them with GLIMPSE, using 1000 Genomes as a reference panel. For most individuals, we recovered original high-coverage genotypes with low error rates (<5% for genotypes with at least one copy of the minor allele). Error rates were higher for African genomes, likely due to underrepresentation in the reference panel. GLIMPSE outperformed the standard imputation tool, Beagle4.1, across multiple coverages (0.1x to 4x). Transversion sites were more accurately imputed than transitions at rare variants, but no significant difference was observed for common variants. This is a remarkable result, as distinguishing true mutations from damage-inflicted C-to-T substitutions is particularly challenging. Finally, we observed no major biases for imputed genomes in the first dimensions in MDS analysis. Runs of homozygosity (ROH; an inbreeding measure requiring diploid data) were highly consistent between imputed and high-coverage data. Again, the ROH overlap was smaller for the African genomes, reflecting the previously seen higher error rates. Altogether, these results suggest that, depending on the studied population, imputation is a reliable method with potential to expand the scope of ancient DNA studies.

Room 15-19

A framework for simulating spatio-temporal population genomic data on a real geographic landscape

Martin Petr[1], Benjamin C. Haller[2], Peter L. Ralph[3], Fernando Racimo[1]
[1]University of Copenhagen, Copenhagen, Denmark. [2]Cornell University, Ithaca, New York, USA.
[3]University of Oregon, Eugene, Oregon, USA

Abstract

One of the goals of population genetics is to understand how different evolutionary forces shape the patterns of genetic variation over time. Because species evolve across both time and space, most evolutionary processes in nature have an implicit spatial dimension. Despite its key role in population genetics, however, the spatial dimension of evolution is often completely neglected, and the tools for building and evaluating complex spatio-temporal population genetic models and inference methods in a reproducible way are lacking. To address these issues, we have developed a new programming framework for simulating spatially-explicit genomic data. Its core component is an R package, spannr (github.com/bodkan/spannr), which leverages real cartographic data for Earth to programmatically and visually encode spatial population boundaries and temporal dynamics (i.e., population displacements and expansions) across a real geographic landscape, and specify population divergences and gene flow events based on an arbitrary admixture graph. Defined using straightforward R code, a spatial demographic model is then compiled into a custom-built SLiM script which can be extended by the user to perform arbitrary simulations, including simulations of natural selection. To aid model design, the R package includes functionality to inspect the encoded spatio-temporal dynamics of populations using interactive graphics, and extract and visualize the admixture graph implied by the specified demographic model. We demonstrate the utility of this software by using it to simulate a complex demographic model of spatial population dynamics of Eurasians over the last 10,000 years, including the Neolithic expansion and the Bronze Age steppe migration.

Room 15-20

Evolutionary history and population genetics of *Mycobacterium leprae* in the Pacific Islands

Adele E Crane[1,2], Kelly E Blevins[3,4], Christopher Lum[5], Kanako Furuta[6], Keolu Fox[7], Anne C Stone[3,4,2]
[1]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [2]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. [3]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA. [4]Center for Bioarchaeological Research, Arizona State University, Tempe, AZ, USA. [5]Department of Pathology, John A Burns School of Medicine, University of Hawaii, Honolulu, HI, USA. [6]Hawaii Pathologists Laboratory, Honolulu, HI, USA. [7]Departments of Anthropology and Global Health, University of California, San Diego, CA, USA

Abstract

Leprosy, or Hansen's disease, is one of the oldest known human diseases and remains a public health issue today, with over 200,000 new cases reported yearly. The disease is caused by the intracellular, obligate pathogen *Mycobacterium leprae* and the more recently discovered *M. lepromatosis*, which is primarily found in Mexico and the Caribbean. The global patterning of genomic variation in *M. leprae* is not well defined, particularly in the Pacific Islands where the origins of the pathogen are disputed. To investigate this, we extracted DNA from 39 formalin-fixed paraffin-embedded biopsy blocks collected between 1992 to 2016. We used whole-genome enrichment and next-generation sequencing to generate 9 *M. leprae* genomes ranging from 4-63x depth of coverage. Phylogenetic analyses place these strains in branches 0 and 5, the basal lineages of the *M. leprae* phylogeny. The phylogeographical patterning and evolutionary dating analysis of these strains support a pre-modern introduction of *M. leprae* into the Pacific Islands. We have expanded this work by including time-series samples from patients during treatment and will use both empirical data and modelling to identify ongoing selection.

Room 16-01

Simple Tools for Rapid Measurements of Evolving Fitness Distributions

Abe Pressman[1], Drew Tack[2], David Ross[1]
[1]National Institute of Standards and Technology, Gaithersburg, MD, USA. [2]AbSci, Vancouver, WA, USA

Abstract

Optimizing genetic engineering requires detailed knowledge of how fitness is distributed among populations of organisms or gene variants. But measuring even simple distributions of fitness requires characterizing the activity of thousands of individual variants, typically only possible with genes that produce a visible signal. Through the introduction of fluorescent genes, we show how cytometry can be used to measure the distribution of fitness for a wide range of genes. Additional analysis highlights other tools for approximating fitness distributions, and shows how these measurements can directly improve protocols for directed evolution.

Room 16-02

Modelling the spatiotemporal spread of beneficial alleles using low-coverage ancient genomes

Rasa A. Muktupavela[1], Martin Petr[1], Laure Segurel[2], Thorfinn Korneliussen[1], John Novembre[3], Fernando Racimo[1]

[1]University of Copenhagen, Copenhagen, Denmark. [2]Universit´e de Paris, Paris, France. [3]University of Chicago, Chicago, USA

Abstract

The advancement of ancient genome sequencing technologies provides the opportunity to study natural selection with unprecedented detail. Rather than investigating indirect patterns left by selection on present-day genomes, we can directly observe whether a given allele was present or absent in a particular region of the world at almost any period of human history within the last 10,000 years. Methods for studying selection, however, often rely on grouping individuals, time stretches or sections of a map into discrete units. These approaches often fail to account for the fact that selection is a spatiotemporal process occurring on a landscape, which limits our ability to perform inference about parameters of interest. Here, we extend a previously developed framework for inferring the spread of beneficial alleles on present-day data using two-dimensional partial differential equations. Under this framework, we can now handle time-stamped ancient samples, as well as genotype likelihoods and pseudo-haploid sequences from low-coverage genomes. We apply the new method to a panel of published ancient West Eurasian genomes and provide dynamic maps showcasing the spread of candidate beneficial alleles over time and space. We also provide estimates for the geographic origin of the mutation, strength of selection and diffusion rate for each of these alleles.

Room 16-03

Characterizing Signatures of Archaic Introgression in Pre-Contact Indigenous Americans

David Peede[1], Viridiana Villa Islas[2], Elizabeth Mejia Perez[2], Fernando Villanea[3], María Ávila-Arcos[2], Emilia Huerta-Sánchez[4,1,5]

[1]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, USA. [2]International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México, Querétaro, Mexico. [3]Department of Anthropology, University of Colorado Boulder, Boulder, USA. [4]Department of Ecology, Evolution, and Organismal Biology, Providence, USA. [5]Center for Computational Molecular Biology, Brown University, Providence, USA

Abstract

Over the past decade, there has been a recent proliferation of genomic studies aiming to characterize the genetic legacy of Neanderthals and Denisovans, in both contemporary and extant populations. Such studies have shown that some introgressed segments have been lost due to natural selection, however, there are multiple examples of archaic variants segregating at high frequencies in both European and Asian populations. Recently, there has been an emphasis on identifying archaically introgressed regions in populations from the Americas, but as a result of European colonization in the recent past, present-day populations in the Americas are considered to be highly admixed, derived from both African and European ancestries. This genomic mosaicism of ancestries observed in present-day Indigenous Americans makes it uniquely difficult to confidently differentiate between an introgressed segment originating from the European lineage compared to the American lineage. This ambiguity inherently hinders studying the evolutionary history of introgressed segments private to the American lineage. To combat this we utilized ancient genomes from individuals that predate European colonization to interrogate the introgression landscape in Pre-Contact Indigenous Americans. We recover candidate genes that exhibit signatures consistent with introgression, as well as identify regions of the genome that exhibit signatures consistent with positive selection. Our study highlights the importance of studying Pre-Contact Indigenous Americans to better understand the evolutionary history of the Americas.

Room 16-04

Signatures of selection on individual loci and pathways spanning 10,000 years in central Europe

Garrett Hellenthal[1], Jens Blöcher[2], Louise Ormond[1], Joachim Burger[2]

[1]University College London, London, United Kingdom. [2]Johannes Gutenberg University Mainz, Mainz, Germany

Abstract

The increasing propagation of ancient DNA samples from multiple time periods offers a powerful means to infer selection by directly modeling observed allele frequency changes over time. However, reliable inference requires overcoming several challenges, such as variable data quality and latent demography. Addressing some of these issues, we describe an efficient new statistical approach to perform genome-wide selection scans using Single-Nucleotide-Polymorphisms (SNPs), which (i) accounts for varying admixture proportions across individuals, (ii) copes with highly variable coverage information across individuals and SNPs, and (iii) infers selection coefficients.

We apply our model to genome-wide SNP data from over 800 ancient individuals, including data from 450 remains found in Europe dated to 800-4800 years ago. We use these data to infer adaptation patterns in central Europe over a time period of more than 10,000 years ago into the Middle Ages. In addition to showing strong evidence of selection at individual SNPs, we leverage results from genome-wide association studies (GWAS) and other analyses to demonstrate how different traits may have been subjected to selection at different times. In particular we assess the evidence for selection in sets of SNPs reported in recent studies involving pharmacogenomics, cancer, immunity, metabolism, aging, and many other health factors. We demonstrate the utility of our new approach, and provide new insights into the selective forces shaping genetics and phenotypes in central Europe.

Room 16-05

Environmental genomics of Late Pleistocene black bears and giant short-faced bears

Bianca De Sanctis[1], Mikkel W Pedersen[2], Nedda Saremi[3], Martin Sikora[2], Emily E Puckett[4], Zhenquan Gu[5], Katherine L Moon[3], Joshua D Kapp[3], Lasse Vinner[2], Zaruhi Vardanyan[2], Ciprian Ardelaan[6], Joaquin Arroyo-Cabrales[7], James A Cahill[8], Peter D Heintzman[9], Grant Zazula[10], Ross MacPhee[11], Beth Shapiro[3], Richard Durbin[1], Eske Willerslev[2]

[1]University of Cambridge, Cambridge, United Kingdom. [2]University of Copenhagen, Copenhagen, Denmark. [3]University of California, Santa Cruz, Santa Cruz, USA. [4]University of Memphis, Memphis, USA. [5]State Key Laboratory of Tibetan Plateau Earth System Science (LATPES), Beijing, China. [6]University of Toronto, Toronto, Canada. [7]nstituto Nacional de Antropologia e Historia, Mexico City, Mexico. [8]University of Rockefeller, New York, USA. [9]The Arctic University of Norway, Lagnes, Norway. [10]Yukon Palaeontology Program, Department of Tourism & Culture, Whitehorse, Canada. [11], American Museum of Natural History, New York, USA

Abstract

Analysis of ancient environmental DNA (eDNA) has revolutionized our ability to describe biological communities in space and time by allowing for parallel sequencing of DNA from all trophic levels. However, because environmental samples contain sparse and fragmented data from multiple individuals, and often contain closely related species, the field of ancient eDNA has so far been limited to organellar genomes in its contribution to population and phylogenetic studies. This is in contrast to data from fossils, where full-genome studies are routine, despite these being rare and their destruction for sequencing undesirable. Here, we report the retrieval of three low-coverage (0.03×) environmental genomes from American black bear (Ursus americanus) and a 0.04× environmental genome of the extinct giant short-faced bear (Arctodus simus) from cave sediment samples from northern Mexico dated to 16–14 thousand calibrated years before present (cal kyr BP), which we contextualize with a new high-coverage (26×) and two lower-coverage giant short-faced bear genomes obtained from fossils recovered from Yukon Territory, Canada, which date to ~22–50 cal kyr BP. We show that the Late Pleistocene black bear population in Mexico is ancestrally related to the present-day Eastern American black bear population, and that the extinct giant short-faced bears present in Mexico were deeply divergent from the earlier Beringian population. Our findings demonstrate the ability to separately analyze genomic-scale DNA sequences of closely related species co-preserved in environmental samples, which brings the use of ancient eDNA into the era of population genomics and phylogenetics.

Room 16-06

Network Analysis of Complex Trait Evolution

Elli Cryan, Daniel Kliebenstein, Jeffrey Ross-Ibarra
UC Davis, Davis, CA, USA

Abstract

A complex trait can be represented by a network of interactions between metabolites and proteins underpinned by a network of genes that encode the proteins. Studying these genes in a network context allows complex trait evolution to be represented by both changes in network architecture and gene sequences across evolutionary time. Network architecture changes include the rewiring of links between nodes as genes are exapted to new functions and the creation of new nodes in cases of gene duplications. Additionally, network architecture properties such as centrality can give insight into the ‘choice’ of which gene in a family will be exapted. Established comparative measures of gene sequence evolution like Dn/Ds can be mapped onto the network to provide a more holistic view of complex trait evolution. An example of an important complex trait in the green tree of life is C4 photosynthesis, which is an adaptation that reduces the incidence of photorespiration by partitioning gas exchange and carbon capture to different cells. C4 photosynthesis is a complex trait that has evolved independently from a C3 photosynthesis ancestral state in around 70 documented events. By comparing the gene networks of pairs of species on either side of independent C4 evolution events, we hope to better understand the process of complex trait evolution.

Room 16-07

A high-density lineage tree reveals dynamics of expression differences accumulation in nondifferentiating clonal expansion

Jian-Rong Yang, Feng Chen, Zizhang Li, Xiaoyu Zhang, Peng Wu, Wenjing Yang, Xiaoshu Chen
Sun Yat-Sen University, Guangzhou, China

Abstract

Differences in gene expression levels are initiated by the division of a single cell and accumulate during further divisions, resulting in either expression noise or differentiation. Nevertheless, how transcriptome-wide noise accumulation is constrained to maintain homeostasis during such continuous cell divisions has remained largely unresolved. Here, we developed a novel technique referred to as the “single-cell transcriptome and dense tree” (STADT) method that simultaneously determines the transcriptomes and lineage tree of >50% single cells in a colony seeded from a single ancestral HEK-293 cell. The comparison of cell pairs with different divergence times revealed gradual accumulation of transcriptome differences that became saturated upon four cell divisions, suggesting the existence of a biological constraint on expression noise with a role in homeostatic maintenance. Further analyses based on a geometric model of gene expression found reduced expression noise in sub-trees/sub-colonies closer to a biological expression boundary, which is a pattern compatible with constrained expression noise due to a limited range of expression levels. Moreover, correlated expression noise (i.e., co-fluctuation) between genes, constituting another type of noise constraint, was mechanistically resolved by our data as either caused by transcriptional control or asymmetric cell division via phylogenetically independent contrast analysis. Collectively, our work not only introduced a novel effort for the reconstruction of lineage trees towards high coverage of cells but also revealed the constrained accumulation of expression noise during the continuous division of nondifferentiating cells for the first time, therefore providing a novel mechanistic understanding of homeostatic maintenance in proliferating cellular colonies.

Room 16-08

Evolution of fungal lifestyles and associated effector proteins in the genus *Fusarium* (Ascomycota)

Rowena Hill[1,2], Richard JA Buggs[1,2], Ester Gaya[1]

[1]Royal Botanic Gardens, Kew, Richmond, United Kingdom. [2]Queen Mary University of London, London, United Kingdom

Abstract

The fungal genus *Fusarium* (Ascomycota) is well known for its role worldwide in plant and animal diseases but many of its species are also found ubiquitously as asymptomatic-plant inhabitants (endophytes), insect-mutualists, and decomposers (saprotrophs). This diversity of lifestyles makes the genus an attractive case-study for fungal lifestyle evolution. To complement existing genomic resources for pathogenic *Fusarium* species, we performed whole genome sequencing, assembly, and structural annotation of five endophytic *Fusarium* strains belonging to the *Fusarium fujikuroi* and *Fusarium incarnatum-equiseti* species complexes. Using predicted protein sets from these and other publicly available *Fusarium* strains, orthologous proteins were inferred for phylogenomic reconstruction of the genus and closely related taxa. Different lifestyles were found scattered in the phylogeny, suggesting frequent evolutionary switches among lifestyles. We annotated putative effector repertoires via computational prediction of candidate secreted effector proteins (CSEPs) – small, secreted proteins known to be involved in the host-fungal interaction. We found that, compared to lifestyle, phylogenetic relatedness better described orthogroup and CSEP variance across taxa, with no significant difference in orthogroup or CSEP content between pathogenic, endophytic, and saprotrophic strains. We did identify, however, specific CSEPs estimated to be under positive selection in endophytic versus pathogenic lineages using an adaptive branch-site model. We found large numbers of accessory CSEPs (i.e., present in more than one taxon but not all) and a comparatively low number of species-specific CSEPs, suggesting that there is limited host specialisation/coevolution among plant-associated fusaria, which may help to explain the apparent broad host range of many *Fusarium* taxa.

Room 16-09

Modular Evolution of the Drosophila Metabolome

Benjamin R Harrison[1], Jessica M Hoffman[2], Ariana Samuelson[1], Daniel Raftery[1], Daniel E L Promislow[1]
[1]University of Washington, Seattle, WA, USA. [2]University of Alabama, Birmingham, AL, USA

Abstract

Comparative phylogenetic studies offer a powerful approach to study the evolution of complex traits. While much effort has been devoted to the evolution of the genome, relatively little work has been done on the evolution of the metabolome, despite the fact that it describes the basic structural and functional building blocks of all organisms. Here we explore the coevolution of the metabolome across 50 million years of evolution in the genus *Drosophila*. We employ a common garden design to measure variation in the metabolome within and among 11 species of *Drosophila*. We find that both sex and age have dramatic and evolutionarily conserved effects on the metabolome. We also find substantial evidence that metabolome evolution is modular. Phylogenetically independent contrast reveal modules made up of groups of covarying metabolites. These modules show different evolutionary trajectories, with some showing signs of stabilizing selection, and others showing more neutral or complex patterns of divergence. These modular coevolutionary patterns also differ between sexes and are affected by age. We explore the relevance of modular evolution to fitness by associating modules with lifespan variation measured in the same common garden. We find several modules associated with lifespan, particularly in the metabolome of older flies. A canonical longevity regulating pathway is enriched in the lifespan-associated module of the female metabolome along with several other pathways that point toward metabolic associations that may underlie 50 million years of lifespan evolution.

Room 16-10

Transient wing polymorphisms in aphid males

Omid Saleh Ziabari[1], Binshuang Li[1], Nate B Hardy[2], Jennifer A Brisson[1]
[1]University of Rochester, Rochester, NY, USA. [2]Auburn University, Auburn, AL, USA

Abstract

The occurrence of polymorphic phenotypes within species has long been used to study the causes and evolutionary consequences of genetic variation. The process of speciation tends to deplete genetic variation, so when these polymorphisms persist between species we can directly study how evolutionary forces maintain genetic variation through species boundaries. Here we compiled monograph descriptions of wing states in male aphids together in an evolutionary framework. Our macroevolutionary analyses revealed two primary results: first, that the polymorphism is evolutionarily transient and second, that polymorphic species have an elevated speciation rate. In most trans-species polymorphisms studied, it's challenging to distinguish between factors that promote persistence, such as shared ancestry or introgression. In aphids, however, the evolutionary distance between polymorphic species, and its evolutionary transience, makes it very likely that polymorphisms evolved independently through de novo mutations. We also found an association between polymorphic species and elevated speciation rates, which in conjunction with the phylogenetic dispersion of polymorphisms, supports a model where polymorphism promotes speciation. We use life history and ecological associations with wing state to propose a model of how wing polymorphisms can, but not always will, act as a precursor to speciation. While we show how the male polymorphism has evolved repeatedly, reversible, and independently; the conditions which lead polymorphisms to fall under balancing selection and subsequent species divergence requires more species-level analyses.

Room 16-11

An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related *Aspergillus* fungi

Matthew E Mead[1], Jacob L Steenwyk[1], Lilian P Silva[2], Patrícia A de Castro[2], Nauman Saeed[3], Falk Hillmann[3], Gustavo H Goldman[2], Antonis Rokas[1]

[1]Vanderbilt University, Nashville, USA. [2]Universidade de São Paulo, Ribeirão Preto, Brazil. [3]Hans Knöll Institute (HKI), Jena, Germany

Abstract

Aspergillosis is an important opportunistic human disease caused by filamentous fungi in the genus *Aspergillus*. Roughly 70% of infections are caused by *Aspergillus fumigatus*, with the rest stemming from approximately a dozen other *Aspergillus* species. Several of these pathogens are closely related to *A. fumigatus* and belong in the same taxonomic section, section *Fumigati*. Pathogenic species are frequently most closely related to non-pathogenic ones, suggesting *Aspergillus* pathogenicity evolved multiple times independently. To understand the repeated evolution of *Aspergillus* pathogenicity, we performed comparative genomic analyses on 18 strains from 13 species, including 8 species in section *Fumigati*, which aimed to identify genes, both ones previously connected to virulence as well as ones never before implicated, whose evolution differs between pathogens and non-pathogens. We found that most genes were present in all species, including approximately half of those previously connected to virulence, but a few genes were section- or species-specific. Evolutionary rate analyses identified over 1,700 genes whose evolutionary rate differed between pathogens and non-pathogens and dozens of genes whose rates differed between specific pathogens and the rest of the taxa. Functional testing of deletion mutants of 17 transcription factor-encoding genes whose evolution differed between pathogens and non-pathogens identified eight genes that affect either fungal survival in a model of phagocytic killing, host survival in an animal model of fungal disease, or both. These results suggest that the evolution of pathogenicity in *Aspergillus* involved both conserved and species-specific genetic elements, illustrating how an evolutionary genomic approach informs the study of fungal disease.

Room 16-12

Evidence for natural selection on low CR1 expression in malaria-endemic regions

Paolo Lorenzini[1], Elena Gusareva[1], Amit Gourav Ghosh[1], Hie Lim Kim[1,2]
[1]Asian School of the Environment, Nanyang Technological University, Singapore, Singapore. [2]Singapore Centre for Environmental Life Sciences Engineering, Singapore, Singapore

Abstract

Malaria has coevolved with humans as its host and proved to be a strong driving force of natural selection on the human genome. To identify genes under natural selection, we analyzed whole-genome sequence datasets from 907 healthy and ethnically diverse individuals to detect positive selection in populations living in malaria-endemic regions but not in non-endemic regions. We found strong positive selection on the Complement Receptor Type 1 (*CR1*) gene, which is well known to be associated with malaria infection. The gene encodes a single chain transmembrane glycoprotein that plays a crucial role in malaria parasite invasion into red blood cells. Previous studies report contradicting results whether a low expression of CR1 is protective against malaria or not. Our analysis identified an adaptive haplotype associated with low expression and a slower rate of erythrocyte sedimentation. This haplotype is likely advantageous against severe malaria by reducing the blockage of small blood vessels. Interestingly, this new haplotype is found in high frequency particularly in the indigenous populations living in Eastern India and Papua New Guinea, where malaria endemicity is high. We also detected signals of positive selection in different CR1 gene variants in other populations, suggesting population-specific manners of positive selection depending on their malaria environments. Remarkably, this study proves that we can identify variants responsible for malaria susceptibility without patient samples or medical records using the population genetics approach.

Room 16-13

Selection inference based on genealogy branch length distributions

Benjamin Wölfl
University of Vienna, Vienna, Austria. Vienna Graduate School of Population Genetics, Vienna, Austria

Abstract

Efficient backward-time and forward-time simulation of simple to complex evolutionary scenarios are combined in order to describe the *branch length distributions* of branches with i underlying leaves in the extant sample in the correlated coalescent trees across linked loci in the genetic basis of a single independent trait. This takes into account the effect of genetic linkage on the decay of coalescent tree correlation across neighboring loci. Specifically, also the distributional shape under *polygenic adaptation* is investigated. Generally, there is no analytical expression for these branch length distributions which raises the importance of a computational insight. Ultimately, this distribution is used in order to construct a statistical test of selection versus no selection which does not only make use of singletons as in the singleton density score (SDS), but generally the density of i-tons. In this way, attention is put on the characteristics of this distribution under different evolutionary scenarios, in particular when we are not only facing a simple recent hard selective sweep. Among other organisms, this method may then for instance be applied to human genetic data sets in order to infer targets of selection.

Room 16-14

Direct estimate of the distribution of fitness effects (DFE) of spontaneous mutations in *C. elegans*

Timothy A Crombie[1], Moein Rajaei[2], Ayush S Saxena[2], Lindsay M Johnson[2], Roby E Tanny[1], Erik C Andersen[1], Charles F Baer[2], Jose Miguel Ponciano[2]
[1]Northwestern University, Evanston, IL, USA. [2]University of Florida, Gainesville, FL, USA

Abstract

The distribution of fitness effects (DFE) of new mutations is a fundamental parameter in population genetics, and has practical application in the context of modeling the genetic basis of complex heritable disease in humans. However, the DFE is very difficult to estimate empirically. At present, nearly all estimates of the DFE rely on either statistical inference from the standing site-frequency spectrum or from laboratory estimates of fitness in mutation accumulation (MA) lines, with no reference to the underlying specific mutations. We report results from a new method to estimate the DFE from competitive fitness data from a set of recombinant inbred lines (RILs) derived from a cross between two *C. elegans* MA lines, combined with whole-genome sequence data from the set of >500 RILs and the parental lines, as well as from ~40 fully-sequenced MA lines. Of five distributions fit to the data, all produce similar results, but the best fit is to a normal distribution nearly centered at 0, with almost no weight beyond +/- 2%. Averaged over two fitness assays, nearly ten years apart, the average fitness effect of a new mutation is about -0.2%. These results are in strong contrast with estimates of the average mutational effect inferred from MA line fitness data not informed by sequence data, which are much larger. We suspect that a few mutations of large effect may have been lost during the inbreeding phase of the construction of the RILs.

Room 16-15

The durability of immunity against reinfection by SARS-CoV-2

Jeffrey P. Townsend[1], Hayley B. Hassler[1], Zheng Wang[1], Sayaka Miura[2], Jaiveer Singh[1], Sudhir Kumar[2], Nancy H. Ruddle[1], Alison P. Galvani[1], Alex Dornburg[3]

[1]Yale University, New Haven, Connecticut, USA. [2]Temple University, Philadelphia, Pennsylvania, USA. [3]University of North Carolina, Charlotte, North Carolina, USA

Abstract

Among the most consequential unknowns of the devastating COVID-19 pandemic are the durability of immunity and time to likely reinfection. There is limited direct data on SARS-CoV-2 long-term immune responses and reinfection. However, the durability of immunity among evolutionarily close coronavirus relatives of SARS-CoV-2 has been assayed, making it possible to estimate its duration of immunity by a comparative evolutionary analysis of related viruses SARS-CoV-1, MERS, HCoV-229E, HCoV-OC43, and HCoV-NL63. We integrated comparative phylogenetic approaches with analysis of Spike protein immunoglobulin G antibody levels and reinfection data on multiple coronaviruses. We estimated the expected decline in antibody levels over time, the probability of reinfection based on antibody level, and the anticipated time to reinfection. Reinfection by SARS-CoV-2 under endemic conditions would likely occur between 3 and 65 months, with a median of 17 months. This protection is of less than half the duration revealed for the seasonal coronaviruses circulating among humans. The time frame for reinfection is fundamental to myriad aspects of public health decision-making. As the pandemic continues, reinfection is likely to become increasingly common. Maintaining public health measures that curb transmission—including among individuals who were previously infected by SARS-CoV-2—coupled with persistent efforts to accelerate vaccination are critical to prevention of COVID-19 morbidity and mortality.

Room 16-16

Resolving Metazoan Opsin Evolution Reveals the Artefactual Nature of Novel Photopigment Subfamilies

Benjamin C Schlossman[1], Davide Pisani[1], Roberto Feuda[2]
[1]University of Bristol, Bristol, Bristol, United Kingdom. [2]University of Leicester, Leicester, Leicestershire, United Kingdom

Abstract

Aim

Opsins are light sensitive photopigment proteins responsible for phototransduction, the conversion of light to an electrical potential. Opsins are present in every animal except sea sponges and are crucial in the role of animal vision. Early opsin evolution is critical to our understanding of how nervous systems evolved in the early animals, although no agreement has been reached on the relationships between opsin subfamilies.

Methods

We constructed a dataset from several previous studies that proposed differing hypotheses of opsin evolution, with particular focus on early marine Phyla: Cnidaria and Ctenophora. We applied an innovative, permutative approach to sequence alignment and implemented a novel protein mixture model in a Maximum Likelihood framework.

Results

We show that the use of destabilising outgroups and inadequate models of substitution play a role in the erroneous proposal of additional opsin paralogs in the Bilateria.

Main conclusion

The new dataset improves the taxonomic sampling of non-bilaterian, marine animals, providing evidence for the diversification of all opsin subfamilies in the eumetazoan and a single origin for the evolution of nervous systems in Metazoa.

Room 16-17

METAGENOMICS SHEDS LIGHT ON THE EVOLUTION OF ‘SUNSCREEN’ PIGMENT METABOLISM IN LICHEN-FORMING ASCOMYCETE FUNGI

Theo Llewellyn[1,2], Reuben W Nowell[3], Andre Aptroot[4], Marina Temina[5], Thomas K Prescott[1], Timothy G Barraclough[2,3], Ester Gaya[1]

[1]Royal Botanic Gardens, Kew, London, United Kingdom. [2]Imperial College London, London, United Kingdom. [3]University of Oxford, Oxford, United Kingdom. [4]Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil. [5]University of Haifa, Haifa, Israel

Abstract

Despite representing almost one-third of known Ascomycete fungi, lichens are under-represented in fungal molecular evolution studies. Lichens are obligate mutualisms between a fungus, one or more photosynthetic algal/cyanobacterial species, and a series of additional microbes. They have successfully colonised all continents and are well-known for their ability to survive inhospitable or extreme conditions. One group, the Teloschistales (an order of >1000 species), has been particularly successful in these ecosystems and is a key element of desert and high-altitude ecosystem biodiversity. Part of their success comes from producing UV-protective secondary metabolites called anthraquinones. Recent work has shown that the evolution of anthraquinones, in conjunction with ecological factors, may have allowed these lichens to colonize unexploited habitats worldwide, facilitating an adaptive radiation. Although anthraquinone chemistry is reasonably well understood, we know very little about the genetic basis of anthraquinone biosynthesis. Here we implement a metagenomics approach to assemble and analyse 24 new high-quality Teloschistales genomes. Using comparative genomics, we investigate the diversity and evolution of secondary metabolite gene clusters and putatively identify those involved in anthraquinone biosynthesis. We find that although secondary metabolite gene clusters in lichens are largely species-specific, the core biosynthetic genes are shared across lichens. Our findings suggest that secondary metabolism diversification in lichens can occur through neofunctionalization of pre-existing biosynthetic genes combined with changes to surrounding accessory genes. This implies that the same ancestral genes have evolved to produce very different compounds in different clades of the main lichen class Lecanoromycetes.

Room 17-01

Resolving the early divergence pattern of teleost fish using genome-scale data

Naoko Takezaki
Kagawa University, Kitagun, Kagawa, Japan

Abstract

Regarding the phylogenetic relationship of the three primary groups of teleost fishes, Osteoglossomorpha (bonytongues and others), Elopomorpha (eels and relatives), Clupeocephala (the remaining teleost fish), early morphological studies hypothesized the first divergence of Osteoglossomorpha, whereas the recent prevailing view is the first divergence of Elopomorpha. Molecular studies supported all the possible relationships of the three primary groups. This study analyzed genome-scale data from four previous studies: (1) 412 genes from 12 species, (2) 772 genes from 15 species, (3) 1,062 genes from 30 species, and (4) 491 UCE loci from 27 species. The effects of the species, loci, and models used on the constructed tree topologies were investigated. In the analyses of the datasets (1) - (3), although the first divergence of Clupeocephala that left the other two groups in a sister relationship was supported by concatenated sequences and gene trees of all the species and genes, the first divergence of Elopomorpha among the three groups was supported using species and/or genes with low divergence of sequence and amino-acid frequencies. This result corresponded to that of the UCE dataset (4), whose sequence divergence was low, which supported the first divergence of Elopomorpha with high statistical significance. The increase in accuracy of the phylogenetic construction by using species and genes with low sequence divergence was predicted by a phylogenetic informativeness approach and confirmed by computer simulation. These results supported that Elopomorpha was the first basal group of teleost fish to have diverged, consistent with the prevailing view of recent morphological studies.

Room 17-02

Ongoing Global and Regional Adaptive Evolution of SARS-CoV-2

Nash D Rochman[1], Yuri I Wolf[1], Guilhem Faure[2], Pascal Mutz[1], Feng Zhang[2,3,4,5,6], Eugene V Koonin[1]
[1]NIH, Bethesda, MD, USA. [2]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, MIT, Cambridge, MA, USA. [4]McGovern Institute for Brain Research, MIT, Cambridge, MA, USA. [5]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA. [6]Department of Biological Engineering, MIT, Cambridge, MA, USA

Abstract

We analyzed more than 300,000 genomes of SARS-CoV-2 variants available as of January 2021 to construct a global topology and complete ancestral reconstruction leveraging a divide-and-conquer approach. We demonstrate the ongoing evolution of SARS-CoV-2 during the pandemic is characterized primarily by purifying selection, with a set of sites evolving under positive selection. The receptor-binding domain of the spike protein and the nuclear localization signal (NLS) associated region of the nucleocapsid protein are enriched with positively selected mutations. These replacements form a strongly connected network of apparent epistatic interactions and are signatures of major partitions in the SARS-CoV-2 phylogeny. Analysis of the phylogenetic distances between pairs of regions reveals four distinct periods of the pandemic linked to the emergence of key mutations. First, rapid diversification into region-specific phylogenies ending February 2020. A major extinction event and global homogenization concomitant with the spread of D614G in the spike protein followed, ending March 2020. NLS associated variants across multiple partitions rose to global prominence March-July, during a period of stasis in terms of inter-regional diversity. Finally, beginning July 2020, multiple mutations, some of which enable antibody evasion, began to emerge associated with ongoing regional diversification. Understanding these trends, which might be indicative of speciation, are paramount to both ongoing and future public health responses.

Room 17-03

An Evaluation of Phylogenetic Workflows in Viral Molecular Epidemiology

Colin Young, Sarah Meng, Niema Moshiri
UC San Diego, San Diego, CA, USA

Abstract

The use of computational techniques to analyze viral sequence data has become increasingly common in epidemiology. These methods typically attempt to make epidemiological inferences based on multiple sequence alignments and phylogenies estimated from the raw sequence data. Like all estimation techniques, multiple sequence alignment and phylogenetic inference tools are error-prone, and the impacts of such imperfections on downstream epidemiological inferences are poorly understood. To address this, we executed multiple commonly-used workflows for conducting viral phylogenetic analyses on simulated viral sequence data modeling HIV, HCV, and Ebola, and we computed multiple methods of accuracy motivated by transmission clustering techniques. For multiple sequence alignment, MAFFT consistently outperformed MUSCLE and Clustal Omega in both accuracy and runtime. For phylogenetic inference, FastTree 2, IQ-TREE, RAxML-NG, and PhyML had similar topological accuracies, but branch lengths and pairwise distances were consistently most accurate in phylogenies inferred by RAxML-NG. However, FastTree 2 was orders of magnitude faster than the other tools, and when the other tools were used to optimize branch lengths along a fixed topology provided by FastTree 2 (i.e., no tree search), the resulting phylogenies had accuracies that were indistinguishable from their original counterparts, but with a fraction of the runtime. Our results indicate that an ideal workflow for viral phylogenetic inference is to (1) use MAFFT to perform MSA, (2) use FastTree 2 under the GTR model with discrete gamma-distributed site-rate heterogeneity to quickly obtain a reasonable tree topology, and (3) use RAxML-NG to optimize branch lengths along the fixed FastTree 2 topology.

Room 17-04

A GPU-Powered Phylogenetic Analysis for Large-scale Genomic Sequences

Jaeyoung Kang[1], Colin Young[1], Justin Morris[1], Ameen Akel[2], Sean Eilert[2], Justin Eno[2], Ken Curewitz[2], Niema Moshiri[1], Tajana Rosing[1]

[1]UC San Diego, La Jolla, California, USA. [2]Micron Inc., Boise, Idaho, USA

Abstract

Phylogenetic inference is a standard procedure in many viral molecular epidemiological workflows. Currently, the state-of-the-art phylogenetic inference methods highly rely on CPU-based tools, and while these tools are able to run in reasonable amounts of time on smaller viral datasets of the past, the massive global sequencing efforts of the SARS-CoV-2 pandemic have yielded ultra-large datasets that yield real-time phylogenetic inference infeasible using existing CPU-based tools. In this work, we propose a maximum likelihood (ML)-based phylogenetics analysis acceleration strategy using graphics processing units (GPU). Based on our analysis of IQ-TREE 2, we offload and parallelize the ML scoring function to the GPU, which is shown to be the main bottleneck of the analysis. The proposed tool converts and handles the tree topology and sequences in a GPU-friendly manner, which maximizes memory coalescing. Through the parallelization of the bottom-up tree reconstruction and scoring per site, the proposed solution leads to significant speedup of the scoring function. Our evaluation results show that the tree scoring function can be run 32x faster on the GPU, as compared to the CPU-based implementation. Further, on a benchmarking experiment measuring overall phylogenetic inference runtime on simulated data modeling SARS-CoV-2 whole genome sequences, we show over an order of magnitude of speedup in our GPU-scoring IQ-TREE 2 implementation compared to the official CPU-based implementation, and the speedup widens as the number of sequences grows.

Room 17-05

Exaptation of transposable elements into the proteome of Drosophila melanogaster

Changcheng Wu[1], Shiqi Luo[2], Hong Zhang[1], Jian Lu[1]

[1]State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Beijing, China. [2]China Agricultural University, Beijing, China

Abstract

Transposable elements (TEs) can be exapted into the proteome of host organisms. Despite our expanding understanding of this field, the prevalence and underlying evolutionary consequences of TE exaptation remain to be further examined. Here we identified 133 exapted TE fragments that are translated in the protein-coding sequences of *D. melanogaster,* as evidenced with ribosome profiling. Furthermore, 64 of these fragments were supported by mass spectrometry data. Our results reveal TEs contribute to the proteomic diversity of *D. melanogaster* through diverse mechanisms. We found 41 TE specific domains (TESDs) that are enriched in DNA binding functions and putatively expand the gene regulatory networks. Most exapted TE fragments are ancient and highly conserved. Nevertheless, some recently exapted TE fragments are lowly expressed and likely targeted by small RNAs. The *dN/dS* analysis suggests most exapted TE fragments are under functional constraints, whereas the McDonald-Kreitman test suggests some of them are driven by positive selection. We also found the exapted TE fragments optimized their codon usage to get adapted to the translational systems of *D. melanogaster* during long-term evolution. Our study systematically reveals that exaptation of TE plays important roles in expanding the proteome of *Drosophila*.

Room 17-06

The phylogenetic information content of multiple sequence alignments

Cassius Manuel[1], Arndt von Haeseler[1,2]
[1]CIBIV, University of Vienna & Medical University of Vienna, Vienna, Vienna, Austria. [2]Faculty of Computer Science, University of Vienna, Vienna, Vienna, Austria

Abstract

A little discussed topic in molecular systematics is the phylogenetic information content of multiple sequence alignments. Possible saturation effects tend to be superficially studied and poorly supported by statistical methods. Using the asymptotics of the likelihood function, we have developed a framework where we compute, for every clade in a phylogenetic tree, its phylogenetic information content. Because every branch of a tree splits the alignment into two clades, the joined phylogenetic information content can be used to estimate branch lengths.

This quantification of information content is obtained in several steps: First we describe the evolutionary process between two sequences. The spectral decomposition of the GTR model, well known in mathematics, allows us to characterize the growth of the likelihood function when the expected number of substitutions (evolutionary time) approaches infinity. From this we develop a statistical test to reject saturation. This test can be applied to two sequences, but also to single edges in a phylogenetic tree.

In parallel, our tools induce efficient estimates of branch lengths in a phylogenetic tree. This method, essentially a first moment estimate that depends on the dominant exponential decay of the substitution model, is called Dominant Exponential Decay Estimate (DEDE). In this context, we show that the DEDE measures the information content of a clade.

The saturation test and the DEDE of branch length are implemented in IQ-TREE. We will also discuss the robustness and efficiency of the DEDE using simulations and biological data.

Room 17-07

TopStrains: A Robust Approach for Building the Phylogeny of Dominant Strains of SARS-CoV-2

Marcos A. Caraballo-Ortiz, Sayaka Miura, Sudhir Kumar
Temple University, Philadelphia, PA, USA

Abstract

Building reliable phylogenies from very large collections of sequences with low variation and significant sequencing error have been challenging. The sequencing error interferes with phylogenetic signal, making robust phylogenetic inference difficult well-documented for SARS-CoV-2. Massive global sequencing of SARS-CoV-2 genomes has produced very tall sequence alignments in which the number of phylogenetically informative positions are orders of magnitude smaller than the number of sequences, and the amount is sequencing error is significant. For such datasets, we developed a simple approach, named *TopStrains*, to build a phylogenetic tree of strains consisting of common variants. This approach dramatically improved the signal-to-noise ratio and produced a robust phylogeny when applied to a multiple sequence alignment of >300,000 SARS-CoV-2 genomes. The bootstrap test of the *TopStrains* phylogeny, in which we resampled genomes instead of positions, produced a well-supported SARS-CoV-2 tree. The root of the SARS-CoV-2 phylogeny, the sequence of the most recent common ancestor, and the orientation of mutational changes in the *TopStrains* phylogeny were the same as those produced by an independent mutation ordering analysis. *TopStrains* offers additional benefits: it is computationally efficient, reveals recurrent mutations, and enables quick placement of newly emerging variants in the existing phylogeny.

Room 17-08

The Evolutionary Origins of Halophilic Archaea

Yutian Feng[1], Uri Neri[2], Artemis Louyakis[1], Sean Gosselin[1], Thane Papke[1], Uri Gophna[2], Johann P Gogarten[1]
[1]UConn, Storrs, USA. [2]TAU, Tel Aviv, Israel

Abstract

The phylogeny of Archaeal halophiles has recently been thrown into controversy, with the characterization of the Nanohaloarchaea (NanoH) and Methanonatronarchaeaia. The NanoH are often considered part of the DPANN superphylum, while the Mtna have been placed at the base of the Methanotecta. However, there is also evidence that both groups share a common ancestor with the classical Haloarchaea. These opposing hypotheses have far-reaching implications on convergent evolution and genome reduction. Many recent analyses have considered large groups of Bacteria and Archaea composed exclusively of MAGs and SAGs as deep branching groups in their respective domains (DPANN is the Archaeal group). These groups display characteristics distinct from other members of their domain, which can attract unrelated lineages into those groups. The groups themselves may reflect artifacts of phylogenetic reconstruction. In this study, we create phylogenies and evaluate the suitability of the placement of the Nanohaloarcaea into DPANN. We also cluster gene families in the NanoH to uncover evolutionary trajectories that exist in the core genome.

We reconstructed phylogenies from 3 marker sets (ATPases, a concatenated set of 44 ribosomal proteins, and a concatenated set of 282 core genes) to test where the NanoH place. In addition, we assembled 8 new genomes for the study. We find the concatenated datasets recover the NanoH in the DPANN, while the ATP synthase dataset recovers the NanoH-Haloarchaea sister group. Through clustering of the gene families in the NanoH, we provide evidence that the ATPase genes have not been transferred.

Room 17-09

Finding a place for tardigrades: assessment of their phylogenetic signal at the site level

Mattia Giacomelli[1], Jesus Lozano-Fernandez[2], Philip C.J. Donoghue[1], Davide Pisani[1]
[1]University of Bristol, Bristol, United Kingdom. [2]Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain

Abstract

A constrained tree analysis is a complementary approach to the classic phylogenomic inference methods to resolve contentious relationships in the tree of life. It allows to understand how the phylogenetic signal is distributed across genes or sites in a phylogenomic alignment, constraining the tree search to two or more competing topologies and measuring the difference in likelihood (per gene or per site) among the set of alternative hypotheses.

Here we use this procedure to investigate the signal at the site-level for two different evolutionary histories of tardigrade evolution, which were recovered under a Bayesian concatenation analysis (under the CAT-GTR model in Phylobayes) with and without recoding our dataset: one where tardigrades are sister to nematodes and the other where they are sister to velvet worms and arthropods.

We show that the inconsistency of the phylogenetic position of this group within ecdysozoans is mainly driven by a small minority of influential sites (~8% of the total). These sites, which have a significant better likelihood for one or the other topology, are characterized by a significantly higher rate in comparison to the others in the alignment. Moreover, if these positions are removed and a new tree search is performed, the signal for both the topologies is lost, and a new unresolved topology is found.

In conclusion, our study shows that the investigation of a phylogeny at the site level reveals a remarkable amount of useful information to understand sources of inconsistencies, especially in hard to resolve phylogenetic problems.

Room 17-10

SLiM-Tree: Simulating Molecular Evolution Along Phylogenies with Pure Population Genetics Models and Hyper-Realistic Fitness Functions

Erin Brintnell, Roman Frolov, Dave W Anderson, A.P. Jason de Koning
University of Calgary, Calgary, Alberta, Canada

Abstract

Current tools for simulating sequence evolution along phylogenetic trees typically take a phenomenological approach and assume continuous-time Markovian models of sequence evolution where events are substitutions of one state for another (fixations). These models ignore the effect of population size and polymorphism, have no site-heterogeneity in sequence-fitness relationships, usually have identical stationary distributions at every position, and have no explicit population genetic basis. It has become appreciated, however, that modelling the population genetics of molecular evolution can be fruitful. For example, site-heterogeneous mutation-selection models overcome many but not all of these limitations of classical models.

Here, we present SLiM-Tree, a software package that uses SLiM (Haller and Messer) to automate forward evolution simulations along phylogenies under general models of population genetics, site-heterogeneous models of sequence-fitness relationships, and non-equilibrium demographies. Fitness profiles can be specified for each genomic position (as in mutation-selection models), or whole molecule fitness can be evaluated with respect to a protein structure.

SLiM-Tree simulates evolution for multiple populations by traversing a Newick formatted phylogenetic tree with branch-lengths defined in terms of generations, and runs simulations along parallel branches on different CPUs when possible. SLiM-Tree allows users to specify a mutation rate, recombination rate, population size and genome size, either globally for an entire phylogenetic tree or locally for each branch of a tree. Here, we use the generality of the underlying models to highlight how classical phylogenetic models of sequence evolution can misrepresent important patterns of molecular evolution and lead to bias in a variety of ways.

Room 17-11

Relative model selection can be sensitive to multiple sequence alignment uncertainty

Stephanie J Spielman[1], Molly Miraglia[2]
[1]Rowan University, Glassboro, NJ, USA. [2]Fox Chase Cancer Center, Philadelphia, PA, USA

Abstract

Multiple sequence alignments (MSAs) represent the fundamental unit of data for phylogenetic analyses. Errors in MSA reconstruction have the potential to induce further errors in downstream analyses such as phylogenetic reconstruction, ancestral state reconstruction, and divergence estimation. However, the MSA is not the only input for these phylogenetic analyses; researchers must also identify a suitable evolutionary model for their analyses. Most commonly, researchers apply relative model selection to select a model from candidate set, and then provide both the MSA and the selected model as input to subsequent analyses. While the influence of MSA errors has been explored for most stages of phylogenetics pipelines, the potential effects on the relative model selection procedure itself have not been explored. In this study, we assessed the consistency of relative model selection when presented with multiple perturbed versions of a given MSA. We found that while relative model selection is mostly robust to MSA uncertainty, in a substantial proportion of circumstances, relative model selection identifies distinct optimal models from different MSAs created from the same set of sequences. While this issue was more pervasive for nucleotide data compared to amino-acid data, we also found that it is extremely challenging to predict whether relative model selection will be robust or sensitive to uncertainty in a given MSA of either data type. We conclude that MSA uncertainty can affect virtually all steps of phylogenetic analysis pipelines to a greater extent than has previously been recognized, including relative model selection.

Room 17-12

Phylogenetic analysis of Dicer's helicase domain

Adedeji M Aderounmu[1], Bryan Kolaczkowski[2], Brenda Bass[1]
[1]University of Utah, Salt Lake City, Utah, USA. [2]University of Florida, Gainesville, Florida, USA

Abstract

Dicer is a ribonuclease that produces microRNAs (miRNAs) and small interfering RNAs (siRNAs) from double-stranded RNA (dsRNA) precursors. While Dicer is conserved across eukaryotes, the enzyme has different functions in different organisms. For example, *Drosophila melanogaster* has two Dicer enzymes; Dicer-1 is dedicated to miRNA processing, while Dicer-2 cleaves viral replication intermediates into siRNAs that silence viral transcripts. Viral defense is mediated by processive cleavage of viral dsRNAs coupled to ATP hydrolysis in Dicer-2's helicase domain. In contrast, humans have a single Dicer but human Dicer's helicase does not hydrolyze ATP *in vitro* and is not thought to be essential for antiviral defense *in vivo*.

Using phylogenetic analysis of animal Dicers, we have begun to elucidate the evolutionary history of Dicer's helicase to understand how amino acid substitutions have contributed to gain/loss of helicase function. We used the maximum likelihood method to construct a phylogenetic tree that describes the evolutionary relationship between Dicer helicase domains across the animal kingdom. We performed ancestral protein reconstruction at select ancestral nodes to assess divergence of helicase function across different organisms. Ongoing biochemical assays using purified recombinant ancestral proteins reveal robust ATP hydrolysis in the most recent common ancestor of arthropod Dicer-2 and Human Dicer, which further diverged to either acquire dsRNA-dependence in invertebrates or lose ATP hydrolysis capability in vertebrates.

Our work will not only provide a coherent understanding of the divergence in Dicer's function across different organisms, but will also allow mechanistic studies of Dicer's helicase domain.

Room 17-13

Transcription start sites are mutation cold spots in humans

David Castellano, Miguel Rodriguez-Galindo, Donate Weghorn
Center for Genomic Regulation, Barcelona, Spain

Abstract

David Castellano, Miguel Rodriguez-Galindo, Donate Weghorn

Although most of the human genome is transcribed (~80%) very few sites show signals of selective constraint (~5%), suggesting that most of the human transcriptome is molecular noise. This, however, does not imply that transcription is irrelevant to evolution. There is a strong negative correlation between transcription and somatic mutation rate in humans. Recent reports suggest that germline mutation rate is also affected by transcription. In this work, I revisit the relative impact of transcription in human mutation patterns. I use millions of somatic mutations from whole cancer genomes and thousands of germline mutations from trio data. I build a statistical model with well-known mutation rate covariates such as mutation sequence context, origin of replication timing, DNase I hypersensitivity sites, transcription level, and others. I find that mutation rate is substantially depressed around transcription start sites (TSS) even for nonCpG>TpG mutations after accounting for the rest of mutation rate covariates. I show that this depression is deeper around the TSS of highly expressed genes but transcription itself is very likely not generating the observed pattern. Finally, I discuss potential mutational processes and DNA repair pathways behind this enigmatic lack of mutations at TSSs and the evolutionary consequences of these mutation cold spots.

Room 17-14

Attribution of Cancer Origins to Endogenous, Exogenous, and Actionable Mutational Processes

Vincent L Cannataro[1], Jeffrey D Mandell[2], Jeffrey P Townsend[2]
[1]Emmanuel College, Boston, MA, USA. [2]Yale University, New Haven, CT, USA

Abstract

Mutational processes in tumors leave tell-tale genomic signatures composed of “passenger” mutations and mutations that have quantifiable effects on the proliferation and survival of cancer cell lineages. We identify the contributions of mutational processes to each oncogenic variant, quantifying responsibility for origination of changes at oncogenic variant sites contributing to tumorigenesis in 23 cancer types. We demonstrate that the variants driving melanomas and lung cancers are predominantly attributable to the actionable, preventable, exogenous mutational processes of ultraviolet light and tobacco exposure, whereas gliomas and prostate adenocarcinomas are largely attributable to endogenous processes associated with aging. Preventable mutations associated with pathogen exposure and APOBEC activity account for a large proportion of the cancer effect within head and neck, bladder, cervical, and breast cancers. These attributions complement epidemiological approaches—revealing the burden of cancer driven by single-nucleotide variants caused by either endogenous or exogenous, non-actionable or actionable processes, and crucially inform cancer prevention.

Room 17-15

Molecular dating of the blood pigment hemocyanin provides new insight on the origin of animals

Elisa M Costa-Paiva[1,2], Beatriz Mello[3], Bruno S Bezerra[1], Christopher J Coates[4], Kenneth M Halanych[5], Federico Brown[1], Juliana M Leme[6], Ricardo I. F. Trindade[2]

[1]Institute of Biosciences, Zoology Department, University of Sao Paulo, Sao Paulo, Brazil. [2]Institute of Astronomy, Geophysics and Atmospheric Sciences, University of Sao Paulo, Sao Paulo, Brazil. [3]Biology Institute, Genetics Department, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. [4]Department of Biosciences, College of Science, Swansea University, Swansea, United Kingdom. [5]Department of Biological Sciences, Molette Biology Laboratory for Environmental and Climate Change Studies, Auburn University, Auburn, AL, USA. [6]Geoscience Institute, University of Sao Paulo, Sao Paulo, Brazil

Abstract

The Neoproterozoic included changes in oceanic redox conditions, the configuration of continents and climate, extreme ice ages (Sturtian and Marinoan), and the rise of complex life forms. A much-debated topic in geobiology is related to the influence of atmospheric oxygenation on Earth and the origin and diversification of animal lineages, with the most widely popularized hypotheses relying on causal links between oxygen levels and the rise of animals. The vast majority of extant animals use aerobic metabolism for homeostasis, hence the binding and transportation of oxygen represents a vital physiological task. Considering the blood pigment hemocyanin (Hc) is present in sponges and ctenophores, and likely to be present in the common ancestor of animals, we investigated the evolution of Hc emergence using bioinformatics approaches on both transcriptomic and genomic data. Bayesian molecular dating suggested that the ancestral Hc gene arose approximately 871 Ma during the Tonian Period (1000-720 Ma), prior to extreme glaciation events of the Cryogenian Period (720-635 Ma). This result is corroborated by the appearance of the earliest animals into the pre-Sturtian Neoproterozoic as attested by purported animal biomarkers (<800-740Ma) and by modern molecular dating for the origin of metazoans of about 1,000-650 Ma (but does contradict previous inferences regarding the origin of Hc ~700-600 Ma). Our data reveal that crown-group animals already possessed hemocyanin-like blood pigments, which may have enhanced the oxygen-carrying capacity of these animals in hypoxic environments at that time or acted in the transport of hormones, detoxification of heavy metals, and immunity pathways.

Room 17-16

A Phylogenetic pipeline for ancient proteomes.

Ioannis Patramanis, Jazmín Ramos-Madrigal, Enrico Cappellini, Fernando Racimo
GLOBE Institute, Copenhagen, Denmark

Abstract

Ancient proteins preserve longer than DNA in fossil remains. Thus, they can conserve phylogenetic information for up to millions of years, enabling a deeper look into the evolutionary past of extinct species. Recent publications have shown that the enamel proteome can be used to reconstruct the phylogenetic history of extinct taxa, supplementing fossil morphology as an additional source of information. However, the limited availability of ancient and present-day protein datasets heavily restricts the types of analyses that can be performed with paleo-proteomes. Here, we present a pipeline under development that aims to simplify and streamline these analyses, providing the user with a readily deployable toolbox for the evolutionary analysis of ancient proteomes. This includes the in-silico translation of protein sequences from ancient and present-day genomes, thus expanding the available sequences that can be compared to ancient proteomes. Our pipeline allows the user to easily align and prepare ancient proteomic datasets, while considering characteristic damage patterns in ancient proteins, and to construct phylogenetic trees. The pipeline can work with individual proteins as well as with concatenated datasets that include several ancient proteins. As proof of principle, we deploy this pipeline in the reconstruction of ancient hominid history using the publicly available proteomes of Homo antecessor and Gigantopithecus blacki, in combination with translated genomes from hundreds of present-day and ancient hominid samples.

Room 17-17

A vicariant origin of the crown Testudinata during the Pangaea breakup

Alexandre P Selvatti, Filipe R R Moreira, Claudia A M Russo, Ana C M Junqueira
Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Abstract

All turtles derive from a wide-ribbed reptile that lived in the Triassic. Later, the ancestor split into two monophyletic groups: the hidden-necked Cryptodira and the side-necked Pleurodira. The time of their split, however, is yet to be settled. Many authors have tackled the problem using time-trees, but the error of those estimates is far too high. As continents faced major reshaping in the Mesozoic, the inconsistency between a Jurassic or a Triassic split prevents a detailed scenario for the evolution of turtles. Recently, a more accurate time dating method has become available, particularly well-suited for rich fossil record groups, such as turtles. Hence, we used the new Fossilized Birth-Death (FBD) using complete mitochondrial genomes (147 taxa) and a set of nuclear orthologs (10 million bp for 25 taxa) to build the time-tree. Our results set the main turtle split at the Early Jurassic (197–185 Ma), suggesting that was caused by the breakup of the supercontinent Pangaea. This agrees with geographical distribution data for turtles (Pleurodira in Gondwana related continents and early fossil Cryptodira confined to Asia). Paleoecological data indicates that the early lineages of crown testudines were not tolerant to salt water. Hence, the opening of the Atlantic Ocean and the formation of the Turgai Strait in Eurasia would well account for the Cryptodira and Pleurodira split, confining the early Cryptodira stocks in eastern Laurasia. With a split in the Jurassic rather than in the Triassic, our results support the vicariant driven evolution proposed by paleontological studies.

Room 18-01

Perception, selection, and breeds: the genetic basis of canine behavioral variation

Kathleen Morrill[1,2], Diane Genereux[2], Jessica Hekman[2], Kathryn Lord[1], Xue Li[2,1], Brittney Kenney[1], Elinor Karlsson[2,1]

[1]University of Massachusetts Medical School, Worcester, MA, USA. [2]Broad Institute, Cambridge, MA, USA

Abstract

Modern dog breeds exhibit conformed physical characteristics and are purported to express distinct and heritable behaviors. The public tends to link breed with behavior, and even scientific studies apply inter-breed comparisons to interrogate the basis of canine behavioral variation. Our surveys of thousands of pet dogs, both pure- and mixed-breed, upend these assumptions. Even dogs of the same breed vary widely in their behavior. Only a handful of breeds differed dramatically in behavioral scores from dogs overall, with border collies ($p=0.00001$) as more biddable than all other dogs and breeds, and labrador and golden retrievers, plus American pit bull terriers, as the most people-friendly breeds ($p=0.00003,0.0017,0.026$). We next assessed whether heritable factors, rather than socio-environmental effects, drive breed differences by modeling the relationship between behavior scores, age, and ancestry inferred from low-coverage sequencing in admixed dogs. Ancestry from herding-type breeds like border collies and Australian cattle dogs contributed to more biddable scores in mutts ($p=0.00018,0.01$), whereas ancestry from beagles and chihuahuas had the opposite effect ($p=0.04,0.04$). By contrast, reports for aggressive behavior in purebreds showed weak concordance with breed stereotypes, but ancestry from purportedly aggressive breeds had no effect on behavior. We suppose that population differentiation in breeds might capture genetic loci involved in behavior; however, selection signals failed to explain association signals from our cohort ($N=2,155$). Rather, selection appeared to highlight morphological genes, suggesting that modern breeds are strongly selected for aesthetics and that inter-breed comparisons is an ineffective strategy for discerning the genetics of canine behavior.

Room 18-02

Within-gene epistatic selection in genetically diverse populations

Anastasia V Stolyarova[1], Tatiana V Neretina[2,3], Elena A Zvyagina[4], Alexey S Kondrashov[5,3], Georgii A Bazykin[1,6]

[1]Skolkovo Institute of Science and Technology, Moscow, Russian Federation. [2]White Sea Biological Station, Biological Faculty, Lomonosov Moscow State University, Moscow, Russian Federation. [3]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russian Federation. [4]Surgut State University, Surgut, Russian Federation. [5]Department of Ecology and Evolutionary Biology, University of Michigan, Michigan, USA. [6]Institute for Information Transmission Problems of the Russian Academy of Science, Moscow, Russian Federation

Abstract

As proposed by Fisher and Dobzhansky, within-population variation can be shaped by coadaptation between chromosomal regions. This may lead to increased correlations between alleles at polymorphic interacting positions, or linkage disequilibrium (LD).

The extent of this increase is expected to depend on the amount of variability present in the population. The basidiomycete fungus *Schizophyllum commune* possesses the highest genetic diversity among the studied eukaryotic species, with ~20% of neutral sites differing between any two individuals. Here, we study the LD patterns in 55 complete genomes of *S. commune* from North America and Europe. In both populations, the LD between nonsynonymous SNPs is higher than that between synonymous SNPs; furthermore, the LD between nonsynonymous SNPs within a gene is higher than between those in different genes at the same nucleotide distance.

Simulations show that the elevated nonsynonymous LD cannot result from differences in negative selection, selective sweeps, or background selection, but implies abundant epistasis between nonsynonymous sites. LD is increased between sites for which interactions are more likely a priori, e.g. those encoding amino acids adjacent in the protein structure. Furthermore, LD between pairs of shared nonsynonymous polymorphisms is correlated in the two populations of *S. commune*, but only if they are located within the same gene.

As expected, the elevated intragenic LD between nonsynonymous mutations is undetectable in humans and barely detectable in fruit flies. Our results suggest that the presence of coadapted allele combinations in many well-studied populations can be limited by low polymorphism, and such coadaptations may emerge whenever sufficient variability is present.

Room 18-03

Northern hemisphere fungal specimens unravel dynamic and polymorphic nature of mating loci

David Peris[1,2], Dabao Sun Lu[1], Vilde Bruhn Kinneberg[1], Ine-Susanne Methlie[1], Malin Stapnes Dahl[1], Håvard Kauserud[1], Inger Skrede[1]

[1]Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, Oslo, Oslo, Norway. [2]Department of Health, Valencian International University (VIU), Valencia, Valencia, Spain

Abstract

Balancing selection is an important evolutionary force that retains genetic diversity. When balancing selection persists for long periods, it can predate speciation events, generating trans-species polymorphisms. Balancing selection has been detected in multiple genes and organisms, such as the sexual mating loci in fungi. In tetrapolar basidiomycete fungi, sexual type is determined by two unlinked loci, *MATA* and *MATB,* which usually are diverse but with conserved domains important to complete the sexual life cycle. Classical studies have determined that the white-rot *Trichaptum* species*,* Hymenochaetales, are tetrapolar fungi, with multiple alleles for *MATA* and *MATB*. Here, we hypothesize that the retention of multiple alleles are due to the action of balancing selection in both mating loci, necessary to promote outcrossings. To support our hypotheses, we sequence the complete genome of a hundred and eighty specimens of three *Trichaptum* species, collected from the northern hemisphere. Using the power of comparative genomics and phylogenetics, we quantify the number of alleles, their divergence, dynamic of those mating regions, and how protein diversity affects successful mating events. We conclude, the genetic diversity of *Trichaptum* mating loci is due to balancing selection, with limited recombination and duplication activity. Balancing selection was persistent before the diversification of the included *Trichaptum* species and in some cases beyond Hymenochaetales, leaving signatures of trans-species polymorphisms. Exploring a large number of specimens, we demonstrate that despite the huge diversity in mating genes, conserved domains and motifs within these genes are important in *Trichaptum*.

Room 18-04

Heterogeneity in viral infections increases the rate of deleterious mutation accumulation

Brent E Allman, Katia Koelle, Daniel B Weissman
Emory University, Atlanta, GA, USA

Abstract

RNA viruses have high mutation rates, with the majority of mutations being deleterious. Using agent-based model simulations, we examine patterns of deleterious mutation accumulation over multiple rounds of viral replication, with a focus on how cellular coinfection and heterogeneity in viral output affect these patterns. When viruses co-infect cells, the protein products produced from their genomes are treated as public goods when producing progeny virus. As such, group-level phenotypes can emerge when more than one genotype infects a host cell. We find, in agreement with previous studies, that coinfection of cells by viruses relaxes the strength of purifying selection, and thereby increases the rate of deleterious mutation accumulation. We further find that cellular heterogeneity in viral output exacerbates the rate of deleterious mutation accumulation, regardless of whether this heterogeneity in viral output is stochastic or is due to variation in cellular multiplicity of infection. These results highlight the need to consider the unique life histories of viruses and their population structure to better understand observed patterns of viral evolution.

Room 18-05

Modeling, simulating and infering Transmission of Reproductive Success: Are whole-genome tree reconstuction tools reliable ?

Jeremy Guez[1,2,3], Ferdinand Petit[1,3,2], Romain Laurent[1,2,3], Bruno Toupance[1,2,3], Evelyne Heyer[1,2,3], Frederic Austerlitz[1,2,3], Flora Jay[4,2,5,6]
[1]MNHN, Paris, France. [2]CNRS, Paris, France. [3]University of Paris, Paris, France. [4]LISN, Orsay, France. [5]Paris Saclay University, Orsay, France. [6]INRIA, Paris, France

Abstract

Cultural transmission of reproductive success (CTRS) impacts the evolution of allelic frequencies. Such transmission has been observed in several human populations around the World: people with many siblings tend to have more children, resulting in a higher variance in family size. This transmission also yields a typical signature in population genetics: CTRS impacts the topology of the coalescent trees, making them imbalanced. Unlike positive selection, this imbalance of trees is not associated to specific loci and spans across the whole genome, and thus can be used for CTRS inference.

Two parts of our research will be presented : (1) modeling of bilinear CTRS for understanding its impact on population genetics and demography inference; (2) investigating whether topologies estimated from genomic data are accurate enough for CTRS inference.

First, we modelled CTRS with the forward-in-time simulator SLiM. Based on extensive simulations, we investigate the impact of biological and cultural parameters on the dynamics of population genetics summary statistics. We also highlight that CTRS processes impact demographic inference, as illustrated by the strong bias in the population size history reconstructed by dadi.

Second, we evaluated the ability of two tree reconstruction algorithms (tsinfer and relate) to retrieve the trees imbalance property under varying parameters. We show that a substantial part of the tree imbalance information is captured by both methods although with varying accuracy and widely different behaviors.

Altogether, this novel CTRS simulator and the evaluation of the inferred tree imbalance index reliability will help us build a functional CTRS inference tool.

Room 18-06

The impact of purifying and background selection on the inference of population history: problems and prospects

Parul Johri[1], Kellen Riall[1], Hannes Becher[2], Laurent Excoffier[3], Brian Charlesworth[2], Jeffrey D Jensen[1]
[1]Arizona State University, Tempe, AZ, USA. [2]University of Edinburgh, Edinburgh, United Kingdom. [3]University of Berne, Berne, Switzerland

Abstract

Current procedures for inferring population history generally assume complete neutrality - that is, they neglect both direct selection and the effects of selection on linked sites. We here examine how the presence of direct purifying selection and background selection may bias demographic inference by evaluating two commonly-used methods (MSMC and *fastsimcoal2*), specifically studying how the underlying shape of the distribution of fitness effects (DFE) and the fraction of directly selected sites interact with demographic parameter estimation. The results show that, even after masking functional genomic regions, background selection may cause the mis-inference of population growth under models of both constant population size and decline. This effect is amplified as the strength of purifying selection and the density of directly selected sites increases, as indicated by the distortion of the site frequency spectrum and levels of nucleotide diversity at linked neutral sites. We also show how simulated changes in background selection effects caused by population size changes can be predicted analytically. We propose a potential method for correcting for the mis-inference of population growth caused by selection. By treating the DFE as a nuisance parameter and averaging across all potential realizations, we demonstrate that even directly selected sites can be used to infer demographic histories with reasonable accuracy.

Room 18-07

Overcoming constraints on the detection of recessive selection in human genes from population frequency data

Daniel J Balick[1,2], Daniel M Jordan[3], Shamil Sunyaev[1,2], Ron Do[3]
[1]Harvard Medical School, Boston, MA, USA. [2]Brigham and Women's Hospital, Boston, MA, USA. [3]Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract

Identifying genes that evolve under recessive natural selection is a longstanding goal of population genetics research with important applications to gene discovery for traits and diseases. Unlike model organisms, identification of diploid selection coefficients in humans requires inference from natural populations. We assessed a range of population genetic and conservation-based measures designed to evaluate per-gene selective constraints and found that, while they are highly sensitive to genes under heterozygous selection, they ubiquitously fail to detect genes evolving under recessive selection. Additionally, more sophisticated likelihood-based statistics designed to explicitly infer recessive selection similarly lack power for any human gene of realistic length given the current size of population samples. However, extensive simulations suggested that enrichment of recessive genes may be detectable in aggregate for gene sets, but that this is sensitively dependent on the fraction of the whole genome that is recessive, an unknown quantity in humans. We designed a method to analyze genes in aggregate to identify enrichment for recessive purifying selection that is informed by population genetics simulations with realistic demography. Applying this to empirical gene sets successfully produced validating enrichments for strong recessive selection in genes previously inferred to be under recessive selection in a consanguineous cohort and in genes involved in autosomal recessive monogenic disorders, including in a set as small as 23 genes associated with evolutionarily lethal recessive diseases. We created a publicly accessible tool that visualizes gene set enrichment patterns and allows users to upload their own gene sets for assessment.

Room 18-08

Diving on the genetic basis of the selective juvenile mortality in ***Diplodus puntazzo***

Cinta Pegueroles[1], Héctor Torrado[2], Núria Raventós[2], Carlos Carreras[1], Enrique Macpherson[2], Marta Pascual[1]

[1]Departament de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Barcelona, Spain. [2]Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Blanes, Girona, Spain

Abstract

Diplodus puntazzo is an iconic fish species distributed along the Mediterranean Sea and the Eastern Atlantic. Interestingly, this species has a high growth rate with interest in aquaculture and plays an important ecological role in coastal areas since preys toxic organisms. In an ongoing study, we observed associations between *loci* and early-life traits in settlers and six-months survivors, suggesting that genetics is key for juvenile survival in this species. Here, we aim to characterize the candidate *loci* for settlers' differential survival. First, we investigated whether there is an enrichment for especific genomic locations or gene biotypes by comparing the distributions of candidate loci versus the rest. Second, we identified *loci* associated to several variables in all the populations studied and we characterized those located within exonic regions. Despite the distribution of the candidate *loci* do not seem to differ significantly from the rest of the annotated loci, by integrating tools from population genomics, comparative genomics and structural biology we identified five *loci* containing Single Nucleotide Polimorphisms (SNPs) that may be key in the selective mortality observed in settlers of *Diplodus puntazzo*.

Room 18-09

Repeated selection against parental genetic load in multiple hybrid wood ant populations

Pierre Nouhaud[1], Simon H Martin[2], Beatriz Portinha[1,3], Vitor C Sousa[3], Jonna Kulmuni[1,4]

[1]Organismal & Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland. [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom. [3]cE3c, Centre for Ecology, Evolution and Environmental changes, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal. [4]Tvärminne Zoological Station, University of Helsinki, Hanko, Finland

Abstract

Hybridization combines gene pools with independent evolutionary histories. When the hybridizing species have different effective population sizes (Ne), they should have accumulated slightly deleterious alleles at different rates during divergence. Theory predicts that after hybridization, purifying selection against these slightly deleterious alleles can lead to a deficit of ancestry from the parental species with the smallest Ne in functional genomic regions. However, such examples remain scarce in the literature and it is unclear to what extent the admixture landscape is impacted by selection against deleterious variation that accumulated in the ancestral lineages. The two wood ant species Formica aquilonia and F. polyctena split around 600 kya and differ in long-term Ne estimates (30% reduction for F. polyctena). Here, we took advantage of a natural F. aquilonia × F. polyctena contact zone in Finland to test whether variation in local ancestry along the genome could be explained by selection against deleterious alleles that accumulated in the ancestral F. polyctena lineage. The analysis of whole-genome data from three independent hybrid populations revealed balanced admixture proportions in hybrids. However, after controlling for the ancestry background, coding regions were significantly enriched for F. aquilonia alleles in each hybrid population. This is consistent with the hypothesis that hybridization enables the purging of slightly deleterious alleles that have accumulated in populations with small Ne. Finally, admixture dating suggests that purging occurred in less than 50 generations. Overall, our results support the parental genetic load hypothesis and should contribute to a better understanding of hybrid genome formation.

Room 18-10

By-products of chromosomal fusions: how has freshwater adaptation been triggered and maintained in three-spine stickleback?

Zuyao Liu, Melanie Hiltbrunner, Marius Roesti, Catherine Peichel
Institute of Ecology and Evolution, University of Bern, Bern, 3012, Switzerland

Abstract

Chromosomal rearrangements have been proposed to facilitate adaptation to divergent environments. But, how such rearrangements contribute to adaptation is still not clear. Here, we fill this knowledge gap by studying threespine stickleback in which ancestral marine fish have repeatedly adapted to freshwater across the northern hemisphere. By performing a comparison of the threespine stickleback genome to a de novo assembly of the fourspine stickleback, as well as an outgroup species, we find two chromosomal fusion events have occurred in threespine stickleback. On these fused chromosomes, we find an enrichment of quantitative trait loci (QTL) that underlie traits that contribute to marine-freshwater adaptation. By comparing whole genome sequences of freshwater and marine threespine stickleback populations, we also find an enrichment of regions under divergent selection on these two fused chromosomes. Surprisingly, we find elevated genetic diversity within regions under selection in freshwater. But our demographic models provide evidence for gene flow between the marine and freshwater populations, which is consistent with a recent simulation study showing that gene flow can increase diversity near regions under selection. Finally, we dated the divergence times between threespine stickleback and other stickleback species across the genome and find that the fusions were likely fixed by speciation events. We hypothesize that these fusions created regions of low recombination in the threespine stickleback genome, which allowed adaptive traits to co-evolve, thereby facilitating freshwater adaptation in the face of recurrent gene flow between marine and freshwater threespine sticklebacks.

Room 18-11

Impact of deleterious mutations in genome scans under isolation with migration models

Gonçalo Caneira[1], João C. Frazão[1], Catherine R. Linnen[2], Vitor C. Sousa[1]

[1]CE3C – Centre for Ecology, Evolution and Environmental Changes, Department of Animal Biology, Faculdade de Ciências da Universidade de Lisboa, University of Lisbon, Campo Grande 1749-016, Lisboa, Portugal. [2]Department of Biology, University of Kentucky, Lexington, Kentucky 40506, USA

Abstract

Genomic scans of differentiation are widely used to detect outliers potentially under divergent selection. However, the effect of removing deleterious mutations can confound these analyses. To quantify this effect we model migration and deleterious mutations, using stochastic simulations of 100Kb chromosomes with SLiM. We compare genetic differentiation patterns in sex-limited hemizygous chromosomes (e.g. X/Z sex-chromosome or haplodiploid autosomes) with diploid autosomes. Our results predict that background selection (BGS) increases differentiation, but this effect vanishes with increasing migration rates. For recessive and slightly deleterious mutations, we predict diploid chromosomes to be affected by associative overdominance (AOD) at low recombining regions, creating valleys of differentiation. Interestingly, such valleys of differentiation due to AOD are also seen for hemizygous chromosomes, but at a much narrower range of the parameter space.

We investigate genome-wide patterns in haplodiploids by simulating data according to the inferred demographic history of hybridizing haplodiploids (Neodiprion sawflies). Our results indicate that AOD and BGS can create heterogeneous genomic patterns, potentially biasing genomic scans to detect divergent selection, and that these effects depend on dominance of deleterious mutations and hemizygosity.

Room 18-12

Rapid parallel adaptation at linked, but independent, loci despite gene flow in silent crickets

Xiao Zhang[1], Jack G. Rayner[1], Mark Blaxter[2], Nathan W. Bailey[1]

[1]School of Biology, University of St Andrews, St Andrews, Fife, United Kingdom. [2]Tree of Life, Wellcome Sanger Institute, Cambridge, United Kingdom

Abstract

Gene flow is predicted to impede parallel adaptation via de novo mutation, because it can introduce pre-existing adaptive alleles from population to population. We test this using rapidly evolving field crickets (*Teleogryllus oceanicus*). In this system, a mutant form of silent male cricket lacking male-specific wing structures that produce sound, 'flatwing', arose and spread rapidly and recently in several Hawaiian populations. Silence protects flatwing males from a lethal, acoustically-orienting parasitoid fly. Morphometric comparisons revealed that male crickets from populations on three separate islands have different flatwing phenotypes. Nevertheless, whole genome re-sequencing data revealed considerable recent and ongoing gene flow among the populations. To understand how this counterintuitive pattern arose, we performed independent genome-wide association analyses and localized flatwing to different loci in each population, but affecting shared genomic hotspots, one of which contains the gene *doublesex*. We found strong signatures of selection on these genomic hotspots of adaptation, but the form of selection differs dramatically among islands and corresponds to known flatwing demographics in the wild. Gene expression in developing cricket wings strengthened support for independent mutations on three different islands, which nevertheless affect a shared developmental pathway involving *doublesex* differently in different populations. Our results show how rapid parallel adaptation can occur on contemporary timescales despite the occurrence of gene flow. The rapid, independent origin of functionally identical adaptations may be less constrained than has been previously appreciated, and we identify conditions that enable parallel adaptations to overwhelm the influence of gene flow during episodes of rapid adaptive evolution.

Room 18-13

How selective sweep and linked selection affect genomic variation in *Drosophila serrata*

Yiguan Wang[1,2], Adam Reddiex[1,3], Scott Allen[1], Stephen Chenoweth[1]

[1]School of Biological Sciences, The University of Queensland, Brisbane, Australia. [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom. [3]Biological Data Science Institute, The Australian National University, Canberra, Australia

Abstract

Understanding whether microevolutionary adaptation is dependent on novel beneficial mutations or standing genetic variation, is an important goal of evolutionary genetics. Progress towards this goal has been greatly enhanced in the genomic era through the study of selective sweeps. Selective sweeps fall into two types: hard sweeps via new mutations and soft sweeps via standing genetic variation. By understanding the prevalence of hard versus soft sweeps in nature, we can predict the frequency with which adaptation to new environments is fuelled by new mutations as opposed to pre-existing genetic variation. However, data are currently lacking on the relative frequency of these two types of selective sweep. In this study, we examined 110 genomes of *Drosophila serrata* sampled from eastern Australia and searched for hard- and soft-sweeps using a deep learning algorithm (diploS/HIC). It revealed that approximately 15% of the *D. serrata* genome was directly impacted by soft sweeps, and that 46% of the genome was indirectly influenced via linkage to these soft sweeps. In contrast, hard sweep signatures were very rare, only accounting for 0.1% of the scanned genome. Gene ontology enrichment analysis showed that some evolutionary arms races (e.g. immunity and sperm-competition) might be the targets of selection. Moreover, we also observed an overrepresentation of deleterious variants within soft sweep regions, suggesting that natural selection drags such variants to higher frequency due to their linkage with beneficial loci. This study provides valuable insight into the direct and indirect contributions of positive selection in shaping the genomic variation in nature.

Room 18-14

Genomic patterns at linked sites due to divergent selection with gene flow: diploid vs hemizygous chromosomes

Vitor C Sousa[1], Emily E Bendall[2,3], Robin Bagley[2,4], Catherine R Linnen[2]

[1]1 CE3C – Centre for Ecology, Evolution and Environmental Changes, Department of Animal Biology, Faculdade de Ciências da Universidade de Lisboa, University of Lisbon, Lisbon, Portugal. [2]Department of Biology, University of Kentucky, Lexington, Kentucky, USA. [3]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA. [4]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University at Lima, Lima, Ohio, USA

Abstract

Empirical data from diverse taxa indicate that the hemizygous portions of the genome (X/Z chromosomes) evolve more rapidly than their diploid counterparts. Faster-X theory predicts increased rates of adaptive substitutions between isolated species, yet little is known about species experiencing gene flow. Here we investigate how hemizygosity impacts genome-wide patterns of differentiation during adaptive divergence with gene flow, combining simulations under isolation-with-migration models, and analysis of haplodiploid species. First, using deterministic and stochastic simulations, we show that elevated differentiation at hemizygous loci occurs when there is gene flow, irrespective of dominance. This faster-X adaptive differentiation stems from more efficient selection resulting in reduced probability of losing the beneficial allele, greater migration-selection threshold, greater allele frequency differences at equilibrium, and a faster time to equilibrium. Second, by simulating neutral variation linked to selected loci with SLiM, we show that faster-X differentiation affects linked variation due to reduced opportunities for recombination between locally adaptive and maladaptive immigrant haplotypes. Finally, we propose a novel approach combining demographic modeling and simulations to investigate haplodiploid species. Focusing on hybridizing *Neodiprion* sawflies we found evidence for faster-X differentiation in haplodiploid pine-feeding hymenopteran species adapted to different host plants. Together, our results indicate that divergent selection with gene flow can lead to higher differentiation at selected and linked variation in hemizygous loci (i.e., faster-X adaptive differentiation), both in X/Z-chromosomes and haplodiploid species.

Room 18-15

Is there an evidence of linked selection in *Drosophila melanogaster* sequence data?

Burçin Yıldırm[1,2], Claus Vogl[1,2]
[1]Vienna Graduate School of Population Genetics, Wien, Austria. [2]Department of Biomedical Sciences, Vetmeduni Vienna, Wien, Austria

Abstract

Nucleotide polymorphism in neutrally and nearly neutrally evolving regions (measured by $\theta=4\mu N_e$) correlates with the local recombination rate. This is due to either a direct, mutagenic effect of recombination on μ or an indirect effect, whereby linked selection (hitchhiking or background selection or both) reduces the effective population size N_e. We use sequences from an African *Drosophila melanogaster* population to show that the effect of recombination is mainly direct, likely via a mutagenic effect. Within short introns, the transcript forms a loop between the 5'-splicing signal and the branch point, to which it attaches without base-pairing. Nucleotide sequences in this 5' loop seem to evolve neutrally. Splicing, however, leads to selection on the nucleotide composition in the 'polypyrimidine tract' between the branch-point and the 3'-splice signal, with a scaled selection strength $\gamma=4sN_e$ within the nearly neutral region. We observed that recombination rate correlates with polymorphism in the 5' loop and the polypyrimidine region, which may be a consequence of direct or indirect effect. If the effect were indirect, we would also expect a correlation with selection strength (measured by $\gamma=4sN_e$) in the polypyrimidine region; however, this was not observed. We found no evidence for a strong contribution from the indirect effect of reduced effective population size N_e via the linked selection on sequence evolution in *D. melanogaster* short introns.

Room 18-16

Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population

Margarita Takou, Juliette de Meaux
University of Cologne, Cologne, Germany

Abstract

During range expansion, edge populations are expected to face increased genetic drift, which in turn can alter and potentially compromise adaptive dynamics, preventing the removal of deleterious mutations and slowing down adaptation. Here, we contrast populations of the European subspecies *Arabidopsis lyrata* ssp. *petraea*, which expanded its Northern range after the last glaciation. We document a sharp decline in effective population size in the range-edge population and observe that nonsynonymous variants segregate at higher frequencies. We detect a 4.9% excess of derived nonsynonymous variants per individual in the range-edge population, suggesting an increase of the genomic burden of deleterious mutations. Inference of the fitness effects of mutations and modeling of allele frequencies under the explicit demographic history of each population predicts a depletion of rare deleterious variants in the range-edge population, but an enrichment for fixed ones, consistent with the bottleneck effect. However, the demographic history of the range-edge population predicts a small net decrease in per-individual fitness. Consistent with this prediction, the range-edge population is not impaired in its growth and survival measured in a common garden experiment. We further observe that the allelic diversity at the self-incompatibility locus, which ensures strict outcrossing and evolves under negative frequency-dependent selection, has remained unchanged. Genomic footprints indicative of selective sweeps are broader in the Northern population but not less frequent. We conclude that the outcrossing species *A. lyrata* ssp. *petraea* shows a strong resilience to the effect of range expansion.

Room 18-17

Increased local genetic load as a result of selective sweeps in human populations

Luiz O. Machado[1], Bárbara D. Bitarello[2,3], Jonatas S. Cesar[1], Diogo Meyer[1]
[1]University of São Paulo, São Paulo, São Paulo, Brazil. [2]Bryn Mawr College, Bryn Mawr, Pennsylvania, USA. [3]University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

Do selective sweeps inflate local levels of deleteriousness in the genome? We looked at the proportion of deleterious variants in such regions for human European (EUR) and East Asian (EAS) populations. First, we established the locations that experienced selective sweeps by curating a list of previously detected regions in more than one published genome-wide scan. Next, we defined target regions as those windows (250kb in EUR, 450kb in EAS) around selected regions with significantly less diversity than the genomic average. Finally, we test whether these windows have a higher proportion of deleterious SNPs than the rest of the genome by looking at several deleteriousness annotations.

To test whether these windows have a higher proportion of deleterious SNPs than the rest of the genome, we developed a robust bootstrap-based methodology by simulating several different procedures and evaluating their bias. In the end, we chose an unbiased resampling procedure that ensures that target and control regions have the same number of polymorphisms, similar site frequency spectra, and LD patterns. This is more robust as a test for increased load than simply counting deleterious sites since many factors can affect the absolute number of deleterious variants in a given genomic region.

We found evidence for an excess of deleterious SNPs in target regions in EUR and EAS compared to controls. Moreover, these patterns were restricted to the population harboring the positive selection signatures, indicating that, at least for the genome regions studied, positive selection was important to increase the load in the human genome.

Room 18-18

Evolutionary Forces in The Bengalese Finch Song: Parallels and Implications for the Study of Human Speech

Madza Fraias-Virgens[1], Terrence Deacon[2], Okanoya Kazuo[3], White Stephanie[1], Huerta-Sanchez Emilia[4]
[1]UCLA, Los Angeles, CA, USA. [2]UCB, Berkeley, CA, USA. [3]University of Tokyo and RIKEN Brain Science Institute, Tokyo, Japan. [4]Brown University, Providence, RI, USA

Abstract

In this work, we identify genes and biological pathways of importance for functional aspects of vocal learning in the Bengalese finch (Lonchura striata domestica), a domesticated songbird commonly found in pet shops, but also a popular animal model in the study of learned vocal behaviors. The Bengalese finch has a remarkably complex song, in which transitions between vocal units are not fixed, introducing variability in song sequencing. This vocal complexity evolved during its domestication from the white-backed munia, a wild songbird easily found throughout East Asia. We use whole-genome sequencing data and analytical tools from population genomics to assess the contributions of selection processes and demographic events shaping the Bengalese finch's genetic variation. Using genome-wide Fst scans, we identify several differentiated genomic regions between domesticated and wild songbirds, with the sex chromosome Z showing the greater proportion of highly differentiated genes. We also find that, as many domesticated animals, Bengalese finches are overall less genetically diverse than their wild ancestors, as shown by reduced average heterozygosity per sampled individual. However, genome-wide Tajimas'D scans show that genetic diversity in munias deviates less from expected across the genome, while diversity deviates more from the expected in Bengalese finches, with long stretches of the genome showing either considerable loss or gain of variability. Interestingly, domesticated and wild songbirds differ in multiple components of the dopamine system, a biopathway fundamental to vocal learning. Our results guide comparative efforts toward identifying convergent patterns of evolutionary change leading to vocal learning in our species.

Room 18-19

Spatial structure alters the allele frequency spectrum produced by hitchhiking

Jiseon Min[1], Michael M Desai[1], Daniel B Weissman[2]
[1]Harvard University, Cambridge, MA, USA. [2]Emory University, Atlanta, GA, USA

Abstract

The reduction of genetic diversity due to genetic hitchhiking is widely used to find past selective sweeps from sequencing data, but very little is known about how spatial structure affects hitchhiking. We use mathematical modeling and simulations to find the unfolded allele-frequency-spectrum (AFS) left by hitchhiking in the genomic region of a sweep in a population occupying a one-dimensional range. For such populations, sweeps spread as Fisher waves, rather than logistically. We find that this leaves a characteristic three-part AFS at loci very close to the swept locus. Very low frequencies are dominated by recent mutations that occurred after the sweep and are unaffected by hitchhiking. At moderately low frequencies, there is a transition zone primarily composed of alleles that briefly “surfed” on the wave of the sweep before falling out of the wavefront, leaving a spectrum close to that expected in well-mixed populations. However, for moderate-to-high frequencies, there is a distinctive scaling regime of the AFS produced by alleles that drifted to fixation in the wavefront and then were carried throughout the population. For loci slightly farther away from the swept locus on the genome, recombination limits the lifetime of alleles in the wavefront and introduces a fourth scaling regime. We find that these signatures of space can be strong even in apparently well-mixed populations with negligible spatial genetic differentiation, suggesting that space frequently distorts the signatures of hitchhiking in nature.

Room 18-20

Inference of the distribution of fitness effects in human populations using local genealogical trees

Alan Izarraras-Gomez, Diego Ortega-Del Vecchyo
International Laboratory for Human Genome Research (LIIGH), National Autonomous University of Mexico (UNAM), Juriquilla, Queretaro, Mexico

Abstract

The Distribution of Fitness Effects (DFE) is central to understanding broader questions in evolutionary biology such as: What determines levels of genetic variation and how much of this variation is under the influence of natural selection. Successful estimates of the DFE have mainly used the Site Frequency Spectrum (SFS), a summary statistic that describes the distribution of derived alleles at different allele frequencies on a genome-wide scale. The SFS does not explicitly reflect the genealogical history of a sample, which contains valuable information of the strength of natural selection acting on an allele. The availability of genome-wide local tree inference methods such as *tsinfer* and *Relate* (Kelleher *et al*. 2019 & Speidel *et al*. 2019), allows us to leverage the genealogical information to infer the impact of natural selection acting on a set of new alleles. We propose a new method to infer the DFE using information from genome-wide local trees along with the SFS. We will infer the impact of natural selection from inferred local trees building upon a variation of the *CLUES* framework introduced by Stern *et al*. 2019 which models how the topology of local trees (e.g. number of branches through time) depend on both demography and natural selection. Furthermore, we modify this existing approach to build a composite likelihood method to estimate the DFE using information from the frequency and the local genealogy of a derived allele. We will show how to apply this method to infer a parametric distribution of the DFE via simulations.

Room 19-01

Leveraging Identity-by-Descent in Health Systems to Characterize a Large Effect Variant Conferring Risk for Liver Disease in Puerto Ricans

Gillian M Belbin[1], Stephanie Rutledge[1], Tetyana Dodatko[1], Sinead Cullina[1], Michael C Turchin[1], Sumita Kholi[1], Denis Torre[1], Muh-Ching Yee[2], Christopher R. Gignoux[3], Noura S. Abul-Husn[1], Sander M. Houten[1], Eimear E. Kenny[1]

[1]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Stanford University, San Francisco, CA, USA. [3]University of Colorado, Denver, Colorado, USA

Abstract

Broad-scale adoption of genomic data in health systems offers opportunities for leveraging population genetics to better understand disease risk. We explored patterns of Identity-by-Descent (IBD) sharing in a patient population in New York City and noted elevated levels of sharing in Puerto Rican (PR) ancestry participants. In an extension of population-based linkage mapping, we clustered IBD haplotypes by homology within the PR group and used these clusters to perform a phenome-wide association study, where we systematically explored relationships between shared IBD haplotypes and over 10,000 health outcomes derived from the electronic health records (EHR). In doing so we uncovered one association that achieved study-wide significance between an IBD-haplotype spanning the ABCB4 locus and severe liver disease. We used genome sequencing and in silico approaches to fine-map the signal to a non-coding variant (c.2784-12T>C) in the gene ABCB4. In vitro analysis confirmed the variant disrupted splicing of the ABCB4 pre-mRNA. Four of five homozygotes had evidence of advanced liver disease, and there was a significant association with liver disease among heterozygotes. Examination of local ancestry on the background of the significant IBD haplotype determined the haplotype to be of European origin. Population-level screening revealed the variant to be at a carrier rate of 1.95% in PR individuals, and extremely rare within some European populations, while otherwise absent globally, suggestive of a PR founder effect. This work demonstrates that integrating EHR and genomic data at a population-scale can facilitate novel strategies for understanding the continuum of genomic risk for common diseases, particularly in populations underrepresented in genomic medicine.

Room 19-02

Identification of ancestry-specific health risks in a large cosmopolitan biobank

Christa Caggiano[1,2], Ruhollah Shemirani[3], Alec Chiu[2], Ruth Johnson[4], Valerie Arboleda[5,6], Gillian M. Belbin[7], Noah A. Zaitlen[1,8]

[1]Department of Neurology, UCLA, Los Angeles, California, USA. [2]Interdepartmental Program in Bioinformatics, UCLA, Los Angeles, California, USA. [3]Information Science Institute, University of Southern California, Marina del Rey, California, USA. [4]Department of Computer Science, UCLA, Los Angeles, California, USA. [5]Pathology and Laboratory Medicine, UCLA, Los Angeles, California, USA. [6]Department of Human Genetics, UCLA, Los Angeles, California, USA. [7]Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]Department of Computational Medicine, UCLA, Los Angeles, California, USA

Abstract

Genetic ancestry is a key factor that affects an individual's risk for disease. In biomedicine, however, researchers and clinicians rely on race and ethnicity as noisy measures of shared genetic ancestry and environmental factors. This results in a reduced ability to identify illnesses at elevated prevalence at the more granular level of fine-scale ancestry groups, which could lead to unrecognized health disparities. To address this concern, we studied a large cosmopolitan biobank, consisting of genotype and electronic health record (EHR) data from 28,000 individuals across the Los Angeles area. We calculated identity-by-descent (IBD) of biobank participants with iLASH and detected over 13 million shared DNA segments. Using these IBD segments, we identified 180 genetic communities. Many of these communities are populations not well-studied in clinical human genetics, including groups characterized by the presence of Persian Jews, Armenians, and Ethiopians. We then used EHR data to estimate population-specific disease risks. In this analysis, we observed several well-known disease-ancestry associations, including an elevated risk for Crohn's Disease in Ashkenazi Jews versus non-Jewish Europeans (logistic regression p=3.82 x 10-8, odds-ratio:2.01±1.14), along with novel associations in many non-European communities, such as an increased risk for pruritus in Persians (logistic regression p=8.43 x 10-4; odds-ratio: 2.54±1.72). These results demonstrate the value of using genetic ancestry in precision medicine initiatives, especially for understudied populations.

Room 19-03

Combining Association and Selective Signals to Improve Detection of Causal Variants in Adaptive Traits

Ioanna-Theoni Vourlaki[1], Miguel Pérez-Enciso[1,2,3], Sebastian E. Ramos-Onsins[1]

[1]Centre for Research in Agricultural Genomics (CRAG) Consortium CSIC-IRTA-UAB-UB, Autonomous University of Barcelona (UAB), Bellaterra, Barcelona, Spain. [2]Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain. [3]Autonomous University of Barcelona (UAB), Bellaterra, Barcelona, Spain

Abstract

The detection of the genetical causes of phenotypic adaptation is a fundamental issue in population and quantitative genetics. The number of causal variants and their selective effects are key factors affecting their identification. Selection leaves a footprint in the patterns of variability, such as an increase of homozygosity and linkage disequilibrium, in the neighborhood of causative variants, which are absent in the non-adaptive traits. Our proposed method combines the association between the phenotype and the genotype and the selective signal. This is performed via an extended genotype homozygosity matrix. This matrix contains the contribution of each individual to genotype homozygosity at each position. It is expected association analyses using this extended matrix will increase power of detecting adaptive loci. Simulated data are generated under different genetic architecture scenarios in order to evaluate the sensibility and the specificity of this methodology for traits having different correlation with fitness.

Room 19-04

Genomic signatures of divergence and environmental adaptation in low recombining regions

Andrés J. Cortés[1], Matthew W. Blair[2]

[1]Colombian Agricultural Research Corporation, Rionegro, ANT, Colombia. [2]Tennessee State University, Nashville, TN, USA

Abstract

Recombination rate and linkage disequilibrium analyses are the basis for exploring the genomic architecture of species' divergence and adaptation. Yet, the genomic signatures associated with divergence and selection are still poorly traced, leading to a knowledge gap on whether divergence and adaptation to heterogeneous landscapes ultimately differ in their genetic architecture. Hence, our goal here was to assess if genomic regions of suppressed recombination, reduced effective population size, and increased drift are more prone to harbor signatures of within-species differentiation and adaptive variation. We used gemomic information gathered in more than 200 common bean accessions across the Americas to quantify genome-wide divergence and allelic associations with *in situ* bioclimatic variables. Highly differentiated genomic regions, according to various summary statistics (F_{ST} and Δ_{Div}), coincided with regions of low SNP density. Divergence in chromosome Pv10 further overlapped with a pericentromeric inversion. Meanwhile, 1-Mb genomic windows indicative of environmental associations were clustered towards the more diverse arms of the chromosomes, where LD values dropped, exhibiting more positive Tajima's D scores. These results indicate that: (1) genomic divergence in *Phaseolus* is mainly shaped by genomic features that limit recombination rate, while (2) adaptation footprints imprint highly recombined regions without any particular signal of reduced recombination enhancing local adaptive genetic variation. This exemplifies that genomic signatures of divergence and adaptation may necessarily concur due to genomic constrains such as suppressed recombination. We close prospecting genomic prediction and machine learning approaches to better integrate conflicting signatures of divergence and local adaptation given pervasive genomic features.

Room 19-05

Predictability and parallelism in the evolution of recent hybrid genomes

Quinn K Langdon[1,2], Daniel L Powell[1,2], Bernard Kim[1], Shreya M Banerjee[1,3], Tris Dodge[1,3], Ben Moran[1,3], Paola Fascinetto-Zago[4,5], Molly Schumer[1,2]

[1]Stanford University, Stanford, CA, USA. [2]Centro de Investigaciones Científicas de las Huastecas “Aguazarca”, Calnali, Mexico. [3]Centro de Investigaciones Científicas de las Huastecas “Aguazarca”, Canali, Mexico. [4]Cornell Univeristy, Ithaca, NY, USA. [5]Texas A&M University, College Station, TX, USA

Abstract

Hybridization between species is widespread across the tree of life. As a result, many species, including our own, harbor regions of their genome derived from hybridization. Despite the recognition that this process is widespread, we understand little about how the genome stabilizes following hybridization, and whether the principles driving this stabilization tend to be shared across species. Here, we dissect the drivers of variation in local ancestry across the genome in replicated recent hybridization events between two species pairs: *X. birchmanni* x *X. cortezi* and *X. birchmanni* x *X. malinche*. We show that broad scale ancestry patterns in hybrids of both species pairs are in part predictable from the recombination landscape and locations of functionally important elements. In addition, we identify dozens of regions of the genome where minor parent ancestry is unusually low or high across species pairs, pointing to shared regions under selection. One such region includes a newly identified hybrid incompatibility that is shared across *X. birchmanni* x *X. cortezi* and *X. birchmanni* x *X. malinche* hybrid populations.

Room 19-06

Coestimation of recombination and substitution rates in protein sequences with approximate Bayesian computation

Miguel Arenas
CINBIO, Vigo, Spain. Universidade de Vigo, Vigo, Spain

Abstract

Unlike nucleotide sequences, methods to estimate the recombination rate in protein sequences are scarce. In order to accommodate this need, here I present a computational framework, called *ProteinEvolverABC*, to jointly estimate recombination and substitution rates from alignments of protein sequences. The framework implements the approximate Bayesian computation approach, with and without regression adjustments, and includes a variety of substitution models of protein evolution, demographics and longitudinal sampling. Its evaluation produced accurate coestimation of recombination and substitution rates under diverse evolutionary scenarios. As illustrative practical examples, I applied it to some viral protein families, including coronaviruses, showing heterogeneous substitution and recombination rates. *ProteinEvolverABC* is freely available from https://github.com/miguelarenas/proteinevolverabc, includes a graphical user interface for helping the specification of the input settings, extensive documentation and ready-to-use examples. Conveniently, the simulations can run in parallel on multicore computers.

Room 19-07

The dependence of homologous recombination rate on the level of heterozygosity in hypervariable fungus *Schizophyllum commune*

Aleksandra V Bezmenova[1], Elena A Zvyagina[2], Tatiana V Neretina[2], Anna V Fedotova[1], Georgii A Bazykin[1], Alexey S Kondrashov[3]

[1]Skoltech, Moscow, Moscow, Russian Federation. [2]N. A. Pertsov White Sea Biological Station, Primorskiy, Republic Karelia, Russian Federation. [3]University of Michigan, Ann Arbor, Michigan, USA

Abstract

Basidiomycete *Schizophyllum commune* is a unique organism with the highest known genetic diversity, such that per-nucleotide heterozygosity level can reach 20% within a single population. Polymorphism level is instrumental in many genomic applications, ranging from studies of recombination to GWASes. Together with small genome length (38.5 Mb), this makes *S. commune* a promising system to study a range of evolutionary processes with resolution that was previously unachievable. In this project, we aim to address how the homologous recombination rate depends on the level of heterozygosity in *S. commune.* It was previously shown that recombination in this fungus tends to occur within more conserved regions – in particular, genes. Following this observation, we assumed that in homozygous regions homologous recombination rate might be higher than that in heterozygous regions. We developed an experimental system in which we can directly measure the recombination rate in completely homozygous regions, and compare it to the normal recombination rate in regions with high heterozygosity. We saw no significant evidence that the recombination rate in completely homozygous regions is higher than that in heterozygous regions.

Room 19-08

Recombination detection for SARS-CoV-2

Anastasia Ignatieva[1], Jotun Hein[2,3], Paul A Jenkins[1,3]
[1]University of Warwick, Coventry, United Kingdom. [2]University of Oxford, Oxford, United Kingdom. [3]The Alan Turing Institute, London, United Kingdom

Abstract

The processes of genetic mutation and recombination are fundamental drivers of viral evolution. While the effects of mutation are generally visible in sequencing data, detecting the presence of ongoing recombination can be a very challenging problem. Recombination events may be undetectable if mutations are not present on the correct branches of the genealogy, and in some cases the effects of recombination can be impossible to distinguish from those of recurrent mutation.

The problem is further complicated for SARS-CoV-2 by its relatively slow accumulation of genetic diversity. While coronaviruses in general are known to have relatively high recombination rates, the extent of ongoing recombination of SARS-CoV-2 within human hosts has remained unclear. We present a method for detecting recombination, which aims to reconstruct genealogical histories while disentangling the effects of recombination and recurrent mutation, and to quantify the probability that recombination has generated the patterns of incompatibilities observed in a given sample.

By analysing samples of SARS-CoV-2 sequencing data, we show that our method can detect sequences carrying patterns of mutations that are consistent with recombination, offering easily interpretable results that provide insights into the evolutionary events that may have generated the given sequences.

Room 19-09

The recombination landscape and genome stability in the Blackcap (*Sylvia atricapilla*)

Karen Bascón Cardozo, Linda Odenthal-Hesse, Miriam Liedvogel
Max Planck Institute for Evolutionary Biology, Plön, Germany

Abstract

Recombination is responsible for reshuffling alleles and breaking up haplotypes, thus influencing genetic variability and the efficacy of selection. Bird genomes lack the protein PRDM9, responsible for the recombination landscape dynamics in most metazoans, and therefore possess unique properties compared to other vertebrates. Birds show an apparent stasis in the positioning of recombination hotspots, yet recombination rates differ widely across the genome and between different (sub-)species of birds. The causes of natural variation in recombination rates and the impact of recombination on evolutionary processes remain poorly understood, particularly in wild species of birds. We studied genome resequencing data across a large number of European blackcap (*Sylvia atricapilla*) individuals with different migratory phenotypes. The characterization of the recombination landscape of this species at a broad and file scale revealed variation in recombination rate along the genome, which was associated with several genetic features. Recombination rates were negatively associated with chromosome size and positively associated with GC content and gene density. Promoters and exons showed higher recombination rates than intergenic regions, and recombination rates were heavily associated with CpG islands. However, this association appears to be further influenced by their position, the length of CpG islands, and local DNA methylation patterns.

Room 19-10

Muller's ratchet in bacteria

Elise Tourrette[1], Daniel Falush[2]
[1]none, Orsay, France. [2]Institut Pasteur Shanghai, Shanghai, China

Abstract

Bacteria are usually subject to low recombination and to gene conversion (exchange of short track of DNA) rather than crossovers. In such species, we could expect phenomena such as Muller’s ratchet, i.e. the loss of the fittest bacteria due to the accumulation of deleterious mutations in asexual populations, in particular if the population size were to be decreased, for example due to a bottleneck. If admixture were to happen between two populations, one subject to a bottleneck and the other not, how easily would the migrated strain from the non-bottleneck population invade the bottleneck population? Although Helicobacter pylori has a higher recombination rate than most bacterial species, we observe a higher mutational load in the populations subjected to an out-of-Africa bottleneck while the admixed populations show a deficit in out-of-Africa populations’ ancestry. By using forward-in-time simulations, we tried to replicate the patterns of diversity seen in H. pylori and we investigate the importance of recombination for bacterial evolution due to demographic events, with a focus on H. pylori parameters. From our results, we show that Muller’s ratchet can happen in bacteria, even under high recombination, during and after a bottleneck. Moreover, due to the difference in fitness, the invasion of mutations from the non-bottleneck population to the bottleneck population is possible and decreases the effect of the Muller’s ratchet.

Room 19-11

Do all guppy chromosomes have similar patterns of male versus female recombination?

Deborah Charlesworth, Suo Qiu, Jim Gardner
University of Edinburgh, Edinburgh, Scotland, United Kingdom

Abstract

Crossovers in male guppies appear to be strongly localized to the termini of the acrocentric chromosomes, whereas in females they occur randomly along the lengths of each chromosome (with perhaps lower rates in physically limited regions near the centromeres). It is thought that recombination rates in males may have changed recently in response to the presence of sexually antagonistic polymorphisms maintained in linkage disequilibrium with the male-determining locus, including male coloration factors that are detrimental in females, as they increase predation. If such evolutionary changes in recombination have occurred, the sex chromosome pair might have evolved greater terminal localization compared with the autosomes. It is also predicted that populations with high predation rates may evolve lower recombination rates than ones with lower predation, where the fitness penalty affecting recombinant females is lower. Direct evidence concerning these predictions is scanty, because recombination rate estimates are scarce from guppy males. I describe some estimates using molecular markers in males from two populations.

Room 19-12

How do human polymorphic inversions affect local recombination rates?

Ruth Gómez-Graciani[1], Jon Lerga-Jaso[1], Marta Puig[1], Alejandra Delprat[1], Antonio Barbadilla[1,2], 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]ICREA, Barcelona, Spain

Abstract

Chromosomal inversions are structural variants that suppress recombination in heterozygosis and that have been involved in multiple evolutionary processes. In most organisms, this recombination inhibition could be caused by a physical limitation to synapse during meiosis, or by the early loss of zygotes carrying aberrant chromosomes resulting from a crossover between opposite orientations. Thus, inversions could impact fertility and be intrinsically deleterious, which means that to become polymorphic they should have positive evolutionary effects that overcome the fitness loss. Here, we investigate how inversions affect recombination in humans by doing a large-scale analysis of a published set of 813,122 crossovers and 787 aneuploid chromosomes from 20 sperm donors and our set of 93 well-characterized autosomal polymorphic inversions, ranging from 0.1 kb to 4.4 Mb. First, following a probabilistic approach, we improved the original single-individual recombination rate maps by increasing the resolution from 500 kb to 150 kb. Next, inversions were imputed using a combination of methods and 61 of them with good quality imputation results in more than 5 donors including both homozygotes and heterozygotes were used for downstream analysis. We detected a decrease in crossover rates within the inverted region in heterozygous compared to homozygous individuals, which is especially noticeable in inversions >25 kb. Moreover, we found a positive correlation between the genetic length affected by inversions in heterozygosis and the number of chromosomal aberrations detected on each individual. These results suggest that the inhibition of recombination between alleles is due to the unviability of recombinant chromosomes.

Room 19-13

Adaptive divergence of meiotic recombination rate in ecological speciation

Swatantra Neupane, Sen Xu
University of Texas at Arlington, Arlington, Texas, USA

Abstract

Theories predict that directional selection during adaptation to a novel habitat results in elevated meiotic recombination rate. Yet the lack of population-level recombination rate data leaves this hypothesis untested in natural populations. Here, we examined the population-level recombination rate variation in two incipient ecological species, the microcrustacean *Daphnia pulex* (an ephemeral-pond species) and *Daphnia pulicaria* (a permanent-lake species). The divergence of *D. pulicaria* from *D. pulex* involved habitat shifts from pond to lake habitats as well as strong local adaptation due to directional selection. Using a novel single-sperm genotyping approach, we estimated the male-specific recombination rate of two linkage groups in multiple populations of each species in common garden experiments and identified a significantly elevated recombination rate in *D. pulicaria*. Most importantly, population genetic analyses show that the divergence in recombination rate between these two species is most likely due to divergent selection in distinct ecological habitats rather than neutral evolution.

Room 19-14

Differential loss of recombination genes after whole genome duplications in vertebrates

Federico G. Hoffmann, Jean-François Gout, Amy L Dapper
Mississippi State University, Mississippi State, MS, USA

Abstract

Whole-genome duplications (WGDs) are an important source of material for molecular innovation. However, WGDs also represent a unique challenge to the recombination pathway because the sudden presence of a second set of nearly homologous chromosomes can result in errors in segregation during meiosis. Reducing the rate of recombination is a simple mechanism to prevent these errors. The positive relationship between copy number of recombination genes and recombination rate suggests that selection may favor the rapid loss of duplicates after WGDs for these genes. In support of this hypothesis, duplicates of genes involved in meiotic recombination are among those most rapidly lost following WGDs in plants. However, evidence for this outside of plants is lacking. Vertebrates provide an ideal system to test this hypothesis because they have undergone 8 WGDs that can be assessed using publicly available data in a robust phylogenetic framework. We integrated bioinformatic, synteny and phylogenetic analyses to infer patterns of retention after WGD for recombination gene families in vertebrates. Our results indicate that: 1) meiotic recombination genes belong to relatively small gene families, ranging in size from 1 to 9 paralogs; 2) most of the duplications in these gene families predate the divergence of animals and fungi; and 3) these genes have reverted to single-copy state after the majority of vertebrate WGDs. Our analyses indicate that after WGDs, recombination genes follow shared evolutionary fates in plants and vertebrates.

Room 19-15

Inversions largely avert the accumulation of deleterious mutations despite suppressing recombination between large blocks of genes

Kaichi Huang[1], Marco Todesco[1], Natalia Bercovich[1], Gregory L. Owens[2], Kate L. Ostevik[3,4], Loren H. Rieseberg[1]

[1]University of British Columbia, Vancouver, BC, Canada. [2]University of Victoria, Victoria, BC, Canada. [3]Duke University, Durham, NC, USA. [4]University of California Riverside, Riverside, CA, USA

Abstract

Chromosomal inversions can act as recombination modifiers that suppress recombination between alleles contributing to local adaptation. However, recombination is also critical for the purging of deleterious variants, such as transposable elements (TEs) and maladaptive changes in proteins. Thus, inversions are predicted to accumulate such mutations, as has been reported for other genomic regions of low recombination. Exploring patterns of molecular evolution for potentially negative consequences of inversions will deepen our understanding of this important genomic architecture. In this study, we investigate the distribution of TEs and coding sequence evolution within a number of inversions across different sunflower species. We find that inversions do not show an enrichment of TEs or elevated deleterious loads relative to background levels. However, there is a minor increase in load associated with inversions in populations that are polymorphic for a given inversion compared to populations that are monomorphic for one of the inversion orientations. These results suggest that inversions are ideal recombination modifiers from an evolutionary standpoint. By suppressing recombination between, but not within orientations, inversions permit independent evolution of large blocks of genes, while largely avoiding the penalties typically associated with low recombination rates.

Room 19-16

Chromosome fusion affects genetic diversity and evolutionary turnover of functional loci, but consistently depends on chromosome size

Francesco Cicconardi[1], James J Lewis[2], Simon H Martin[3], Robert D Reed[4], Charles G Danko[2], Stephen H Montgomery[5]

[1]School of Biological Sciences, University of Bristol, Bristol, United Kingdom. [2]Baker Institute for Animal Health, Cornell University, Ithaca, NY, USA. [3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom. [4]Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA. [5]School of Biological Sciences, University of Bristol Bristol, Bristol, United Kingdom

Abstract

Major changes in chromosome number and structure are linked to a series of evolutionary phenomena, including intrinsic barriers to gene flow or suppression of recombination due to chromosomal rearrangements. However, chromosome rearrangements can also affect the fundamental dynamics of molecular evolution within populations by changing relationships between linked loci and altering rates of recombination. Here, we build chromosome-level assembly *Eueides isabella* and, together with a recent chromosome-level assembly of *Dryas iulia*, examine the evolutionary consequences of multiple chromosome fusions in Heliconius butterflies. These assemblies pinpoint fusion points on 10 of the 20 autosomal chromosomes and reveal striking differences in the characteristics of fused and unfused chromosomes. The ten smallest autosomes in *D. iulia* and *E. isabella*, which have each fused to a longer chromosome in *Heliconius*, have higher repeat and GC content, and longer introns than predicted by their chromosome length. When fused, these characteristics change to become more in line with chromosome length. The fusions also led to reduced diversity, which likely reflects increased background selection and selection against introgression between diverging populations, following a reduction in per-base recombination rate. We further show that chromosome size and fusion impact turnover rates of functional loci at a macroevolutionary scale. Together these results provide further evidence that chromosome fusion in Heliconius likely had dramatic effects.

Room 19-17

Human inbreeding has decreased in time through the Holocene

Francisco C. Ceballos[1,2], Kanat Gürün[1], Ezgi Altınışık[3], Hasan Can Gemici[1], Cansu Karamurat[1], Dilek Koptekin[1], Kıvılcım Başak Vural[4], Igor Mapelli[1], Ekin Sağlıcan[1], Elif Sürer[1], Yılmaz Selim Erdal[3], Anders Götherström[5], Füsun Özer[3], Çiğdem Atakuman[1], Mehmet Somel[1]

[1]Middle East Technical University, Ankara, Turkey. [2]Health Institute Carlos III, Madrid, Spain. [3]Hacettepe University, Ankara, Turkey. [4]Middle East Technical University, Ankara, Spain. [5]Stockholm University, Stockholm, Sweden

Abstract

The history of human inbreeding is controversial. In particular, how the development of sedentary and/or agricultural societies may have influenced overall inbreeding levels is unclear. Here we present an approach for reliable estimation of runs of homozygosity (ROH) in genomes with ≥3x mean sequence coverage across >1 million SNPs, and apply this to 411 ancient Eurasian genomes from the last 15,000 years. We show that the frequency of inbreeding, as measured by ROH, has decreased over time. The strongest effect is associated with the Neolithic transition, but the trend has since continued, indicating a population size effect on inbreeding prevalence. We further show that most inbreeding in our historical sample can be attributed to small population size instead of consanguinity. Although singular cases of high consanguinity can also be identified in the archaeogenomic record, these appear limited to members of farming societies.

Room 19-18

Genome-wide signals of selection in experimental populations of Trinidadian guppies (*Poecilia reticulata*)

Mijke J van der Zee[1], James R Whiting[1], Josephine R Paris[1], Ron D Bassar[2], Joseph Travis[3], David N Reznick[4], Detlef Weigel[5], Bonnie A Fraser[1]

[1]University of Exeter, Exeter, United Kingdom. [2]Williams College, Williamstown, USA. [3]Florida State University, Tallahassee, USA. [4]University of California Riverside, Riverside, USA. [5]Max Planck Institute for Developmental Biology, Tuebingen, Germany

Abstract

It is now accepted that phenotypic evolution can occur quickly but understanding the genetic basis of rapid adaptation is largely unknown. Population genomic studies of experimental populations of guppies provide a unique opportunity to study this phenomenon.

Guppy populations that were transplanted from high predation (HP) to low predation (LP) environments have been shown to evolve phenotypes similar to naturally-colonised LP populations in as few as 8 generations and these phenotypes persist in subsequent generations in laboratory environments. Here, we present results from whole genome sequencing of four experimental LP populations and their high-predation source. An unique advantage of this experiment is that we can compare our genetic signals to monthly mark-recapture census data. We first used site frequency and haplotype statistics (e.g. ROHs and IBD segments) to infer past demography. We find that haplotype statistics are more sensitive to founding and known bottleneck events. Next, we used genome-wide scans (including a novel multivariate scan method) to uncover signatures of selection and we identified strong candidate loci for convergent evolution in all of our experimental sites. These results help us understand how guppies are able to adapt so quickly to new environments.

Room 19-19

Recombination, selection and population history effects on the distribution of runs of homozygosity in two populations of wild red deer

Anna Hewett
University of Edinburgh, Edinburgh, United Kingdom

Abstract

Inbreeding results in genomic segments that are IBD coming together as a run of homozygosity (ROH). In addition to ROH indicating the level of inbreeding in a population, the genomic distribution of ROH can reveal information about the mechanisms driving their own distribution. Previous studies have shown multiple factors can affect ROH distributions, such as: recombination, population history and selection. Here we use two wild populations of red deer, one from the Scottish island of Rum (3046 individuals) and one from the Scottish mainland (157 individuals) to search for ROH >2.5Mb across 35,132 autosomal SNPs. ROH were searched for using both physical and genetic map positions to investigate the effect of recombination. Using the physical map, inter-population similarities in ROH distribution were apparent. However, using the genetic map there is a more uniform distribution of ROH and less correlation between the populations. This suggests that both the inter-population similarities and variation in ROH distribution are partially driven by variation in recombination rate and that by using the genetic map one can account for recombination effects on ROH distribution. Following this result, using the genetic map we found 5 ROH hotspots (where ROH were unusually common in the population) in Rum deer, with >15% of the population having a ROH at the same location. As these hotspots persist after accounting for recombination we suggest they may mark regions of positive selection. Consistent with this, haplotype diversity in ROH hotspots was reduced in comparison to the rest of the genome.

Room 19-20

Using IBD segments to infer population structure and demographic history of wild boars

Angeles de Cara, Lucia Perez-Pardal, Leili Khalatbari, Hosein Yusefi, Albano Beja-Pereira
CIBIO, Vairão, Portugal

Abstract

The development of single nucleotide polymorphism arrays has been crucial in our understanding of human and livestock population diversity. SNP arrays are thus a great resource, considerably cheaper than whole genome sequencing.

It is however important to carefully describe how the array has been developed, which individuals have been used, or which markers are included. This is because these arrays are developed to provide the most information about those markers, which may be related to genes of particular interest.

Thus, they are not random markers from the genome, and their site frequency spectrum is not necessarily neutral. In humans, in order to make demographic inferences from many individuals, the Human Origins array was developed using 13 panels of SNPs. This has allowed other researchers to develop corrections by taking into account the design. However, when using other panels like those developed for domestic species into their wild counterpart, we need to take great care.

Here, taking into account this issue, we explore the genetic diversity and population structure of wild boars, using data from the porcine 60k SNP array. We use haplotype-based methods to infer population structure and genetic diversity, and shared segments of identity-by-descent to infer demographic history. Our goal is to describe the most likely scenario of demographic history for Iranian wild boars, which have previously only been studied using mtDNA. We conclude that the Iranian population is fairly isolated. We observe two genetically distinct clusters, while field observations have described four distinct phenotypes.

Room 20-01

Comparative Analysis of SINE DNA variants impacting Mobilization

Madison N Altieri, Julia V Halo, Abigail S Jarosz, Jordyn M Hutchison
Bowling Green State University, Bowling Green, OH, USA

Abstract

SINEs and LINEs are conserved repetitive elements in higher eukaryotic genomes. The maintenance of their conservation serves as a mechanism to mobilize in the genome. Throughout evolution the host genome has tolerated, but also repressed, LINE and SINE mobilization. This dichotomy represents the ‘arms race’ between SINEs and LINEs and the host genome allowing amplification within the host’s genome without becoming lethal to the organism. Over time, SINEs and LINEs gain genotypic variations to their sequence that render their mobilization at the virtue of the variations accumulated. The SINE relies on LINE-encoded machinery to mobilize, which is able to recognize and mobilize the SINE intermediate within the genome. Furthermore, trends suggest that divergence from the consensus sequence typically affects the SINE’s ability to mobilize in the genome. However, there are exceptions to this trend that lead us to believe that variants in the SINE sequence contribute to their mobilization. In this research we hope to find i.) structural variants of 7SL and tRNA derived SINEs that aid in the mobilization of SINEs, and ii.) assess the activities of tRNA derived SINEs to hypothesize their accession to the ribosome. By linking the secondary structure of tRNA derived SINEs to their mobilization, the mechanism of association with the ribosome to hijack LINE’s enzymatic machinery may be revealed.

Room 20-02

Lineage-specific genes and duplications are enriched among genes displaying expression bias during development of slime molds

Frederic J.J. Chain[1], Saara K Luna[2]
[1]University of Massachusetts Lowell, Lowell, MA, USA. [2]Harvard University, Cambridge, MA, USA

Abstract

Dictyostelid amoebae are slime molds that transition from unicellular growth to multicellular development under starvation conditions via a process that emerged over 600 million years ago. Building on previous studies investigating the genetics underlying this development, we sought to determine the contributions of lineage-specific genes and duplications. Using published gene expression datasets across five distinct developmental stages, we found that species-specific singleton genes tend to have biased expression as measured by the statistic of tissue-specificity tau, indicating predominant expression of these genes in a single developmental stage. This enrichment of expression bias is observed across species and throughout development, such that each developmental stage is overrepresented with species-specific singleton genes. In contrast, singleton genes that are shared across species are underrepresented among biased genes, as they tend to be broadly expressed throughout development. Subsets of duplicate genes are also enriched among genes with biased expression, but different gene families are involved in the different species. In one species (*Dictyostelium discoideum*), lineage-specific expansions were enriched among genes with stage-specific expression, whereas their orthologs in other slime molds were not. These results suggest that specific stages of the developmental process leading to multicellularity in slime molds has undergone lineage-specific recruitment of genes that may play a role in the evolution of species-specific innovations, potentially via *de novo* gene birth in addition to gene duplication and divergence.

Room 20-03

Reconstruction of a phylogenetic tree of cells by using single-cell human transcriptome data

Satoshi Oota[1], Kazuho Ikeo[2]
[1]RIKEN, WAKO, SAITAMA, Japan. [2]NIG, Mishima, Shizuoka, Japan

Abstract

Cell lineage tracing is one of the fundamental topics in biology. While prospective lineage tracing is notoriously complex and costly, retrospective lineage tracing is more feasible and less invasive. For the retrospective approach, however, we still needed reliable somatic mutation data by deep sequencing. Thanks to various technologies that flourished to acquire single-cell transcriptome data, a massive amount of single-cell transcriptome data is now available on public databases.

Here we reconstructed a phylogenetic tree of cells in the human placenta using single-cell transcriptome data. The tree potentially represents the cell lineages of the sampled cells. While values of the transcriptome coverage were relatively low, we found that our ‘tree’ consistently represents the known cell lineage of the human placenta. We also inferred ancestral somatic mutations with the parsimony principle and estimated the selective pressure on each branch.

Overall, our approach was reasonably effective in detecting phylogenetic signals of the cell lineage tree. In the meantime, we still need further evaluation of our methods with additional data: e.g., somatic mutation data obtained by a prospective approach.

Room 20-04

Quantifying the Evolution of Fluconazole Resistance in *S. Cerevisiae* Using Molecular Barcodes

Sam Apodaca[1,2,3], Kara Schmidlin[1], Kerry Geiler-Samerotte[1,2]

[1]Arizona State University Biodesign Center for Mechanisms of Evolution, Tempe, Arizona, USA. [2]Arizona State University School of Life Sciences, Tempe, Arizona, USA. [3]Arizona State University School of Molecular Sciences, Temp, Arizona, USA

Abstract

A major problem facing modern medicine is the evolution of drug resistance. Drugs can be rendered ineffective if their target is able to acquire beneficial mutations so multi-drug treatment strategies are being explored as one way to hinder the evolution of resistance. In order to understand the viability of these strategies, we must understand the full spectrum of resistance mutations that an organism can develop, not just the most common ones. Because rare mutations exist that are resistant to both drugs can exist, knowing the frequency of such mutations is important for making predictions about how problematic they will be. More generally speaking, understanding the full spectrum of adaptive mutations available in a given environment is a key goal in evolutionary biology. This experiment aimed to expand on previous research on the evolution of drug resistance in *S. cerevisiae* by using molecular barcodes to track ~100,000 evolving lineages. The barcoded cells were evolved with serial transfers for 200 generations in different concentrations of fluconazole, radicicol, or combinations of both drugs. Sequencing data was used to track barcode frequencies over the course of the evolution, allowing us to track resistant lineages as they arose and quantify differences in resistance evolution across environments. We observed many adaptive mutants, even in the multi-drug environments, and saw that these mutants rose to high frequency more quickly in environments representing stronger selective pressures. Next, we will study gene-by-environment interactions (GxE) to explore whether adaptive mutants remain adaptive or suffer tradeoffs across environments.

Room 20-05

Epigenetic change as a gene-independent evolutionary force, or not?

Hanne De Kort[1], Tuomas Toivainen[2], Filip Van Nieuwerburgh[3], Bart Panis[1], Timo P. Hytönen[2], Olivier Honnay[1]

[1]KU Leuven, Leuven, Belgium. [2]University of Helsinki, Helsinki, Finland. [3]Ghent university, Ghent, Belgium

Abstract

Epigenetic change has frequently been demonstrated to co-occur with environmental and phenotypic change, pointing to the evolutionary significance of epigenetic variation. Epigenetic change may, however, rarely really drive phenotypic evolution, but rather constitute a stepping stone between genetic and phenotypic change. Most genetic variants (with transposable elements as prime examples) most likely trigger epigenetic change somewhere along the molecular routes to phenotypic change. We hypothesize that most epigenetic variants are tightly linked to genetic variants, and that this linkage is not simply due to genomic proximity. Based on comparative whole (epi)genome analysis of 20 woodland strawberry plants (Fragaria vesca) originating from distinct elevations, and raised under distinct soil moisture settings, we show high co-variation between epigenetic DMCs (differentially methylated cytosines) and genetic SNPs (single nucleotide polymorphisms). This co-variation is not associated with genomic proximity, and is highest between neutral SNPs (not currently targetted by natural selection) and drought-responsive DMCs. We also found that drought-responsive DMCs typically co-vary with hundreds of SNPs, suggesting a broad genetic basis for natural selection on epigenetic drought regulation. Although further research is required to fully rule out SNP-independence of genome-wide DMCs, our findings provide a conceptual framework for polygenic processes shaping genome-wide methylation patterns.

Room 20-06

Distinct epigenetic and transcriptional responses of two mangrove species under exposure to high-dose UV-B

Yushuai Wang[1], Chenglong Huang[1], Tianyuan Zhang[1], Cairong Zhong[2], Shulin Deng[3,4], Tian Tang[1]
[1]State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, Guangdong, People's Republic of China, Guangzhou, China. [2]Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou 571100, Hainan, People's Republic of China, Haikou, China. [3]CAS Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement & Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, People's Republic of China, Guangzhou, China. [4]Xiaoliang Research Station for Tropical Coastal Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, People's Republic of China, Guangzhou, China

Abstract

As an integral part of sunlight, ultraviolet-B (UV-B) affects diverse biological processes in all phototrophic organisms. Tropical plants such as mangroves are recurrently exposed to strong solar UV-B radiation. However, the mechanism by which they adapted to UV-B radiation remains unclear. Here we report that two mangrove species *Avecennia marina* and *Rhizophora apiculata* exhibited distinct epigenetic and transcriptional responses under chronic high-dose UV-B. *A. marina* showed almost equal hyper or hypo-methylation in all three sequence contexts after UV-B exposure. In contrast, *R. apiculata* mainly exhibited hypo-methylation in the CHG context, accompanied with the relaxation of transposable elements (TEs) suppression by small RNAs. Up-regulation of TE-nearby genes in *R. apiculata* contributed to more mis-expression than *A. marina*. While both species showed similar transcriptional responses including up-regulation of flavonoid biosynthesis genes and down-regulation of photosynthesis genes, *A. marina* specifically up-regulated ABC transpoter genes and ubiquinone biosynthesis genes that are known to be protective against harmful effects of UV-B exposure. Taken together, differences in plant responses to UV-B radiation can occur at both the epigenetic and transcriptional levels. Our results suggest that maintaining genome integrity against TE de-repression may be a common theme in plant UV-B defence while transcpitonal changes underlying physiological defence mechanisms can be the same or different between species.

Room 20-07

Environmental contexts and metabolic tissues influence allele-specific expression patterns in mice

Celine L St. Pierre, Heather A Lawson
Washington University School of Medicine, St. Louis, MO, USA

Abstract

Allele-specific expression (**ASE**) is a phenomenon where one allele is preferentially expressed over the other allele. Genetic and epigenetic differences can cause ASE, such as a haplotype-specific variant in a promoter or DNA methylation silencing one parent's allele (genomic imprinting). Depending on a gene's function, these expression imbalances can lead to phenotypic variation. 30-56% of genes show evidence of ASE, indicating that allele-specific effects have widespread impacts on gene regulation. However, whether and how ASE patterns are sensitive to environmental signals remains underexplored. Here, we use a simple yet powerful F_1 reciprocal cross mouse model to explore how external (diet) and internal (sex) environmental factors can influence ASE patterns. Male and female mice from a F_1 reciprocal cross of the LG/J and SM/J strains were fed a high-fat or low-fat diet. We obtained RNA-Seq transcriptomes from three metabolically-relevant tissues: hypothalamus, white adipose, and liver. We harnessed the strain-specific genetic variants to distinguish between two classes of ASE: parent-of-origin dependent (unequal expression based on an allele's parental origin) and sequence dependent (unequal expression based on an allele's nucleotide identity). We found that ASE patterns of both classes are highly tissue- and context-dependent. They vary across metabolic tissues, between males and females, and in response to dietary environments, thus providing a systems biology perspective on the gene-by-environment architecture underlying complex metabolic traits. Untangling the genetic, epigenetic, and environmental mechanisms contributing to allele-specific gene regulation is essential for understanding the relationship between DNA sequence and phenotypic variation.

Room 20-08

GxE for Hypoxia Tolerance and Hypoxia-induced Differential Gene Expression Reveals the Effects of Genotypes Origin in Daphnia

Millicent N Ekwudo, Morad Malek, Cora E Anderson, Lev Y Yampolsky
East Tennessee State University, Johnson City, Tennessee, USA

Abstract

Hypoxia is often a limiting factor determining aquatic organisms' survival and distribution. Although metabolic and transcriptional responses to hypoxia are exceptionally well studied, little is known about intraspecific variation in the ability to tolerate hypoxia and in plastic response to either chronic mild or acute severe hypoxia. Here we report differences in chronic and acute hypoxia tolerances in four geographically distinct genotypes of *Daphniamagna*, two originating from large permanent ponds and two from small intermittent hypoxia-prone habitats. Genotypes from permanent habitats showed a lower life expectancy in chronic hypoxia and lower survival in 24-hour severe hypoxia tests than their counterparts originated from permanent habitats. To pinpoint transcriptional responses that may be associated with this GxE interaction, we singled out transcripts that responded to both chronic mild and acute severe hypoxia relative to normoxic control in hypoxia-tolerant, but not in hypoxia-sensitive genotypes, distinguishing between adaptive transcriptional plasticity associated with hypoxia tolerance and constrained plasticity observed in hypoxia-sensitive genotypes. This approach revealed several candidate hypoxia-tolerance genes, including genes playing a role in respiratory metabolism, such as lactate (Ldh), pyruvate (Pdh), and NADH dehydrogenases, in oxygen transport and retention, (cytoglobin ortholog) and in reactive oxygen species metabolism such as superoxide dismutase (SOD), the latter showing ecological subfunctionalization among paralogs. Additionally, we reveal potential roles in hypoxia tolerance of juvenile hormone regulation pathway, as well as *Daphnia* orthologs of mammalian cytokines and angiotensin-converting enzyme.

Room 20-09

Population structure and selection analysis in Cameroonian next-generation sequencing data

Sonia Olaechea-Lázaro[1], Neskuts Izagirre[1], Saioa López[2], Óscar García[3], José Miguel Lorenzo-Salazar[4], Carlos Flores[4,5,6,7], Krishna Veeramah[8], Garrett Hellenthal[9,10], Mark Thomas[9,10], Santos Alonso[1]
[1]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country, Leioa, Vizcaya, Spain. [2]Department of Strategy and Data Insights, Wellcome Trust, London, United Kingdom. [3]Forensic Science Unit, Forensic Genetics Section, Basque Country Police, Erandio, Vizcaya, Spain. [4]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Islas Canarias, Spain. [5]Instituto de Tecnologías Biomédicas (ITB), Universidad de La Laguna, Santa Cruz de Tenerife, Islas Canarias, Spain. [6]Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Islas Canarias, Spain. [7]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain. [8]Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, USA. [9]UCL Genetics Institute, University College London (UCL), London, United Kingdom. [10]Research Department of Genetics, Evolution & Environment, UCL, London, United Kingdom

Abstract

Cameroon is considered an “Africa-in-miniature” not only due to its high genetic, ecological, and linguistic diversity, but also due to the wide variety of subsistence strategies adopted by its inhabitants. Here, we analyse whole-exome sequencing (WES) data to understand how environmental factors mould human genetic diversity in this context.

We analysed 100 whole-exomes from Cameroon sampling 30 ethnic groups (including Fulani, Bassa, Kotoko, Mambila). We evaluated population structure and diversity (PCA, Fst) and signatures of selection (Tajima’s D, Fu and Li’s D, PBS). Given that buccal swabs were the DNA source and a proportion of the reads were unmapped (~1%), these have been used to identify the oral microbiome.

Our analysis suggest that Cameroonians might be genetically subdivided into three main population clusters that locate to the North, West, and coast regions of the country. Moreover, we identify putative region-specific selection signals associated to environmental factors. Of note, our initial results suggest that *SLC39A4*, a gene involved in zinc transport, has been target of positive selection. Moreover, *Saccharomyces cerevisiae*, *Streptococcus pneumoniae* and *Neisseria elongata* appear in our sample as the most frequent fungi, virus and bacteria, respectively.

As a conclusion, we have identified population structure and signatures of natural selection using WES, which has proved an adequate tool to assess this phenomena, as well as for metagenomic analysis.

Room 20-10

Cultural and genetic structures along the Silk Road: a cross-discipline comparison

Serena Aneli[1], Massimo Mezzavilla[2], Paolo Gasparini[2,3], Luca Pagani[1,4]
[1]Department of Biology, University of Padova, Padova, Italy. [2]Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy. [3]Department of Medical, Surgical and Health Sciences, Trieste, Italy. [4]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia

Abstract

The complex interplay among genetics, culture and environment forms an individual's identity, influencing their behaviour, choices and health. However, to what extent this intertwined network could quantitatively describe an individual is still difficult to say.

Here, we focused on dietary habits and genome-wide data of 538 individuals coming from six countries along the Silk Road: Georgia, Armenia, Azerbaijan, Uzbekistan, Kazakhstan and Tajikistan. These lands saw the spreading of innovations, foods - as well as DNA - halfway across Eurasia, thus representing an ideal subject to explore different dietary patterns possibly related to cultural factors and ancestry.

Taking food preferences as proxies for culture, we used Discriminant Analysis of Principal Components (DAPC) to probabilistically infer "cultural clusters", where mixed memberships are allowed. We performed five DAPC analyses, iteratively removing the predominant foods. The first two layers, driven by alcoholic beverages and pork, clustered the individuals along the East-West direction mirroring the genetic admixture patterns, except for Azerbaijan, which shares the Islamic culture with Eastern countries. Vegetables and sweets guided the following two layers, by clustering the individuals according to their age (p-value=1.11e-18). Interestingly, the last layer was driven by protein-rich foods with higher consumption significantly associated with increased Steppe pastoralists ancestry (Yamnaya-like and Mongolian).

Our soft clustering approach enabled us to model and compress the individual's dietary information in vectors of probabilities. Encoding other cultural variables would quantitatively describe even better individuals' culture, thus ultimately supporting genetic association studies, regardless of their subjects' ancestry and culture.

Room 20-11

Divergent evolution along coastal-highland gradients in the wild tomato species *Solanum chilense*

Gustavo A. Silva-Arias, Edgardo M. Ortiz, Aurelien Tellier
Technichal University of Munich, Freising, Bayern, Germany

Abstract

Organisms distributed along environmental gradients constitute optimal models for understanding genotype-environment interactions and genetic mechanisms involved in adaptation related to recent range expansion processes. The wild tomato species *Solanum chilense* is an ideal model to study mechanisms of abiotic-stress resistance because it is adapted to extremely stressful environments along valleys covering coastal-to-highland habitats next to the Atacama Desert. Plants from coastal and highland populations show well-differentiated phenotypes, and common-garden experiments support specialization in abiotic-stress resistance traits. We studied a recent expansion process in *S. chilense* combining landscape genetics and niche modeling frameworks. Using a high-throughput target sequencing, we genotyped 200 individuals from populations arranged in valleys representing four replicates of altitude gradients arranged latitudinally and additional two well-differentiated lineages at the edge of the species' distribution. We sequenced 500 intergenic markers and ca. 1500 candidate abiotic-stress resistance genes previously identified in *S. chilense* using gene expression analyses. Genetic structure analyses performed on putatively neutral markers showed latitudinal differentiation supporting higher similarity among populations from the same valley and clear isolation-by-distance signal. Past projections of niche models and demographic reconstructions support a recent (post-glacial) recolonization of highland regions. Further divergence of a new lineage in southern highland localities likely diverged from populations from a single valley. Based on correlations between allele frequencies and climatic variables, we identified candidate adaptive genes and transcription factors that support the inference of gene-network evolution for the colonization of new habitats.

Room 20-12

Local adaptation may cause pseudogenization in the evolution of Sulawesi macaque species

Nami Arakawa[1], Kanthi Arum Widayati[2], Laurentia Henrieta Permita Sari Purba[2], Xiaochan Yan[3], Hiroo Imai[3], Bambang Suryobroto[2], Yohey Terai[1]
[1]SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan. [2]IPB University, Bogor, Indonesia. [3]Kyoto University, Inuyama, Japan

Abstract

When an ancestral generalist species adapts to a local environment and becomes a specialist, gene usage is expected to change. Genes that are no longer needed for life generally accumulate mutations in the specialist. As a consequence, they become pseudogenes specific to the specialist. In this study, we tested this hypothesis using Sulawesi macaque species as a model. The Sulawesi macaques, *Macaca* species endemic to island of Sulawesi (Central Indonesia), have differentiated into seven morphologically distinct species in their respective allopatric habitats. If they adapt to their respective local areas, genes that are not vital in the area are likely found as pseudogenes. We thus determined exome sequences from ten individuals each for four Sulawesi macaque species, *M. nigra*, *M. nigrescens*, *M. hecki*, and *M. tonkeana*, and extracted genes possessing homozygous premature stop codons as pseudogenes from the exome data. As a result, we found many pseudogenes related to olfaction, taste, detoxification, and hair components in the species. Olfactory receptor gene family presented the largest number of pseudogenes specific to each species. In this gene family, eight genes were pseudogenized in all four species, whereas 33 genes were species-specific pseudogenes. Since genes related to olfaction, taste, and detoxification are important for foraging, Sulawesi macaque species may have adapted to food resources specific to their respective habitat areas, which may cause pseudogenization of the genes that are no longer used for foraging for local foods.

Room 20-13

Genetic and environmental rewiring of genetic interaction networks

Po-Hsiang Hung[1], Kieran Dennehy Collins[1], Sasha F. Levy[2,3,4], Gavin Sherlock[1]
[1]Stanford University, Stanford, CA, USA. [2]Joint Initiative for Metrology un Biology, Stanford, CA, USA. [3]Stanford University, Palo Alto, CA, USA. [4]SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Abstract

A genetic interaction (GI) between two genes results in a double mutant phenotype that deviates from expectations based on phenotypic effects of the two single mutants. GI networks are a critical tool for describing gene functions, functional modules, crosstalk between pathways, and how the network topology influenced by environments/genotypes results in phenotypic changes. Comprehensive GI networks have been described in only a few model organisms, and in a single experimental condition for each. The long evolutionary distances between model organisms make it hard to distinguish whether the driving force(s) of network rewiring is from mutation, gene duplication, or gene loss. Moreover, these “static views” of GI networks likely miss many GIs that are only present in other environments or genotypes, and are unlikely to elucidate the mechanisms by which GI networks are modified across environments or evolutionary time. To this end, systematic measures of the interactions between a large set of gene pairs in multiple environments and closely related genotypes are required. However, state-of-the-art methods are too low throughput or need to overcome technical challenges to perform such studies at a sufficient scale. To fill this knowledge gap we are developing an approach that combines CRISPRi with saturating transposon mutagenesis that will allow us to quantitatively study GI dynamics at a large scale. We will use this approach to measure GI network dynamics in three closely related yeasts across diverse environments. These results will characterize how GI networks change under both genetic and environmental perturbations.

Room 20-14

Salp in bloom: Genome dynamics provide insight into *Salpa thompsoni*'s reproductive success

Kate R Castellano, Jill Wegrzyn, Paola Batta-Lona, Ann Bucklin, Rachel J O'Neill
University of Connecticut, Storrs, CT, USA

Abstract

Warming trends in the Southern Ocean have altered trophic dynamics in favor of the pelagic tunicate, *Salpa thompsoni*, which is showing rapid population expansion (blooms), that displace other key species. Salp blooms are facilitated by their reproductive life history, which consists of seasonal alternations between sexual and asexual stages. However, we lack a foundational understanding of the genomic features that define this species and support salp bloom formation. Both paired-end short read and linked-read sequencing approaches have failed to generate a quality genome assembly for any salp species, likely the result of the high repeat content of the salp genome. Herein, using Oxford Nanopore long read sequencing, de novo assembly and comprehensive transcriptomics, we have derived a new assembly for *S. thompsoni*, consisting of 8,815 contigs, an N50 of 188 kb, and genome coverage of 78%. Through this work, we have discovered strong secondary structures within the *S. thompsoni* genome that dramatically affect sequencing efficiency. Our analyses of these secondary structures led to the discovery of abundant G-quadruplex sequences distributed throughout the *S. thompsoni* genome at a significantly higher frequency compared to other tunicate species, suggesting such structures are a defining feature of this salp genome. The link between these G-quadruplex sequences, de novo gene and repeat annotations and transcriptional profiles across life stages will be discussed. Collectively, out results provide novel insights into the function of unique genomic features in the regulation of genome stability between asexual to sexual reproduction.

Room 20-15

Diving mammals lose Paraoxonase 1 function in multiple different ways

Wynn K Meyer[1,2], Jerrica Jamison[3,2], Allie M Graham[4], Charlotte Cournoyer[5], Irene Kaplow[6], Alexis Michaels[7,2], Jiaxuan Yang[8,2], Rebecca Richter[9], Clement Furlong[9], Nathan Clark[4,2]
[1]Lehigh University, Bethlehem, PA, USA. [2]University of Pittsburgh, Pittsburgh, PA, USA. [3]University of Toronto Scarborough, Scarborough, ON, Canada. [4]University of Utah, Salt Lake City, UT, USA. [5]University of Florida, Gainesville, FL, USA. [6]Carnegie Mellon University, Pittsburgh, PA, USA. [7]Duquesne University, Pittsburgh, PA, USA. [8]Tsinghua University, Beijing, China. [9]University of Washington, Seattle, WA, USA

Abstract

Multiple clades of mammals have adapted to semi-aquatic lifestyles, with accompanying changes in anatomy, physiology, and behavior. In addition to changes that help these animals navigate and survive in aquatic environments, these lineages also share an unexpected feature: the functional loss of the enzyme Paraoxonase 1 (PON1), a lipid antioxidant that additionally protects against neurotoxicity from the oxon byproducts of organophosphate pesticides. We here demonstrate, using both predicted protein sequence data and direct biochemical testing for the PON1 enzyme, that at least three lineages of semi-aquatic mammals have lost PON1 function. Within pinnipeds (seals, sea lions, and walruses), two separate lineages appear to have lost PON1 function independently, and one such loss most likely involved changes to regulatory elements, rather than to the coding sequence. In this case, we highlight several candidate regions that may underlie the loss, based on analysis of sequence evolution across carnivores and chromatin accessibility in dog liver. The observed repeated functional losses of PON1 may imply a benefit to the loss of this enzyme in a semi-aquatic environment. We find a weak trend of association between longer predicted dive capacity and PON1 functional loss, and we do not see a positive association between PON1 functional loss and shift to semi-aquatic diet. This fuller picture of PON1 evolution in carnivores suggests features of the ancestral environment that may have led to this protein’s loss, and it may inform prioritization of species to monitor for organophosphate exposure from pesticide runoff.

Room 20-16

Laying the genomic foundation for foundation species: early insights from a deep-sea primnoid coral genome

Michelle L. Neitzey[1], Kate R. Castellano[1], Timothy M. Shank[2], Rachel J. O'Neill[1]
[1]University of Connecticut, Storrs, CT, USA. [2]Woods Hole Oceanographic Institute, Woods Hole, MA, USA

Abstract

Cold-water corals, defined as living below 50 meters, represent 65% of coral species. They contribute to the diversity of deep-sea life by providing vertical substrate for other organisms to collect food and shelter. Because of their slow-growing nature and slow rates of recruitment, they are highly impacted by anthropogenic activities. Deep trawling bycatch can often include corals and have long-lasting effects on benthic communities. Oil spills can devastate coral communities for years after an initial spill. Increases in ocean temperature and ocean acidification are shrinking the available habitat for these corals. Genomic information on deep-sea corals would provide valuable biological information on these foundation species, and inform conservation efforts, such as marine protected areas. However, the genomic adaptations of corals to the deep-sea environment have yet to be investigated. To enable comparative genomics studies, we used short and long-read sequencing technologies to develop the first deep-sea coral genome from the Primnoidae family. Primnoid species exhibit one of the greatest depth ranges for coral families, making them an excellent candidate for comparing genomic adaptations across pressure, temperature, and alkalinity with depth. Assembly pipeline, draft genome statistics, and correlative transcriptome data analyses will be presented. Future research will focus on genome improvement and annotation to investigate potential adaptations.

Room 20-17

Antibiotic treatment increases fitness effects of spontaneous mutations

Wei-Chin Ho, Jadon Gonzales, Michael Lynch
Center of Mechanisms of Evolution, The Biodesign Institute, Arizona State University, Tempe, AZ, USA

Abstract

The nature of spontaneous mutations, including rates and effects, largely determines the evolutionary dynamics of genotypes and phenotypes. Therefore, studying the interaction between mutations and environments is critical for understanding organismal evolution in new environments. While mutation rates can depend on environmental conditions, whether the effect sizes of mutations are also variable remains controversial. In principle, plastic effect sizes of mutations can result from an environment-altered physiology and/or from historical selection for environment canalization/robustness. To study these issues, using wild-type *Escherichia coli* exposed to the antibiotic norfloxacin as a model system, we focused on two sets of mutation-accumulation (MA) lines of *E. coli* where the mutations had accumulated in a nearly neutral trend with or without norfloxacin and evaluated their growth curves with or without norfloxacin in the plate reader. The distributions of fitness effects for four treatments (2 MA environments x 2 evaluating environments) were then inferred using a maximum-likelihood framework. The mean effects for spontaneous mutations accumulated without norfloxacin are significantly larger when evaluated with norfloxacin. In addition, the mean effects for the spontaneous mutations accumulated with norfloxacin present are significantly larger when evaluated with norfloxacin. Therefore, norfloxacin generally increases mutational effects, whether or not the mutation-accumulating environment and the growth-evaluating environments are matched. Our data suggest that mutational effects are primarily a function of the physiological background, while the impact from the historical selection for environment canalization is likely negligible. These results enhance our understanding of genotype-environment interactions, context-dependent fitness landscapes, and microbial evolution to antibiotics.

Room 20-18

Social processes shaping variation in admixed Cabo Verdeans

Katharine L Korunes[1], Giordano B Soares-Souza[2], Katherine Bobrek[3], Hua Tang[4], Isabel I Araujo[5], Amy Goldberg[1], Sandra Beleza[2]

[1]Duke University, Durham, NC, USA. [2]University of Leicester, Leicester, United Kingdom. [3]Emory University, Atlanta, GA, USA. [4]Stanford University, Stanford, CA, USA. [5]Universidade de Cabo Verde, Praia, Cape Verde

Abstract

Complex social and demographic processes shape patterns of genetic variation in populations. The mosaic ancestry of admixed populations provides an extra line of genomic information to study these processes. However, common summaries of variation used to study recent population history, such as runs of homozygosity (ROH), are often difficult to interpret in admixed populations. Here, we consider the admixed human population of Cabo Verde to understand how recent mating patterns influence genetic variation. First settled in the late 1400s, Cabo Verdeans are descendants of Portuguese colonizers and enslaved West African people. We consider Cabo Verde's well-studied historical record alongside genome-wide SNP data from 563 individuals across islands. We find that spouses are more correlated in their global ancestry proportion than expected in a randomly mating population, consistent with positive ancestry-assortative mating. These nonrandom mating patterns, along with predicted bottlenecks during founding of the islands, are expected to increase ROH in the Cabo Verdean population. However, we observe overall low level of ROH across islands. To understand this discrepancy, we partition ROH by length classes and analyze the local ancestry switches within ROH. We find that shorter ROH are indeed low, likely representing admixture interrupting homozygous segments. Yet, long tracts that are likely formed post-admixture, are indeed higher in the Cabo Verdean population. We also find evidence of sex-biased admixture through comparisons of X and autosomal ancestry. This case study provides insight into the processes that generate genetic variation and the interpretation of ROH in admixed populations.

Room 20-19

Runs of homozygosity & mutation load uncovers taxon-specific demographic history in endemic island ungulates

Sabhrina Gita Aninta[1], Rosie Drinkwater[1], Dwi Sendi Priyono[2], Athena Syarifa[2], Selina Brace[3], Stephen Rossiter[1], Nurul Winarni[4], Jatna Supriatna[4], Laurent Frantz[1,5]

[1]Queen Mary University of London, London, United Kingdom. [2]Wildlife Conservation Society, Bogor, Indonesia. [3]Natural History Museum, London, United Kingdom. [4]Research Center for Climate Change - Universitas Indonesia, Depok, Indonesia. [5]Ludwig-Maximilians-Universität, München, Germany

Abstract

Insular tropical biodiversity is under increasing pressure as anthropogenic disturbances can lead to genomic erosion, especially for large bodied species which often already possess low genetic diversity. Disentangling whether a species possesses low genetic diversity due to long term small population size (i.e. on a small island) or due to recent decline (i.e. due to anthropogenic disturbances), however, can be challenging. To address this issue, we assess the usefulness of genomic analyses, including runs-of-homozygosity (ROH), through analysing the whole genome data of both modern and museum samples of two large ungulates, anoa (the “dwarf-buffalo”, Bubalus spp.) and babirusa (“deer-pig”, Babyrousa spp.), both endemic to the Wallacea archipelago in Indonesia. Genomic analyses of 67 anoa and 46 babirusa indicated that 70% and 80% of the samples possess signs of runs of homozygosity. We found that populations coming from small islands showed similar levels of runs of homozygosity as found on the largest island. Interestingly, however, anoa that is thought to have colonised Wallacea more recently (~2 millions years as opposed to ~13 for babirusa) possess on average less but longer RoHs. Although the genomic load was overall similar in both species, babirusa possessed lower load in shorter ROHs than anoa. This suggests that anoa underwent more recent population decline, potentially because of stronger hunting pressure, and did not have enough time for purging deleterious mutations within its longer RoH. Overall, our study highlights the huge potential of ROH for disentangling long term versus recent population processes in conservation studies.

Room 20-20

Estimating the time since admixture from phased and unphased molecular data

Veronica Miro Pina[1], Thijs Janzen[2]
[1]Centre for Genomic Regulation, Barcelona, Spain. [2]University of Groningen, Groningen, Netherlands

Abstract

After admixture, recombination breaks down genomic blocks of contiguous ancestry. The breakdown of these blocks forms a new ‘molecular clock’, that ticks at a much faster rate than the mutation clock, enabling accurate dating of admixture events in the recent past. However, existing theory on the break down of these blocks, or the accumulation of delineations between blocks, so called ‘junctions’, has been limited to using regularly spaced markers on phased data. Here, we present an extension to the theory of junctions using the Ancestral Recombination Graph that describes the expected number of junctions for any distribution of markers along the genome. Furthermore, we provide a new framework to infer the time since admixture using unphased data. We demonstrate both the phased and unphased methods on simulated data and show that our new extensions perform much better than previous methods, especially for more ancient admixture times. Lastly, we demonstrate the applicability of our method on an empirical dataset of labcrosses of yeast (Saccharomyces cerevisae) and on two case studies of hybridization in swordtail fish and Populus trees.

Room 21-01

Heteroplasmy and the dynamics of cytoplasmic DNA transmission

Amanda K Broz, Lexi Keene, Mychaela Hodus, Daniel B Sloan
Colorado State University, Fort Collins, CO, USA

Abstract

Heteroplasmy is the presence of more than one type of mitochondrial or plastid genome in an individual, which can arise from mutations in organelle genomes. The extent to which heteroplasmies are vertically transmitted across generations or eliminated through successive bottlenecking is not well understood in plants. Here, we used ddPCR to quantitatively track the distribution and inheritance of heteroplasmies associated with de novo single nucleotide variants (SNV) in both mitochondria and plastid genomes of *Arabidopsis thaliana*. Patterns of SNV inheritance revealed striking differences between the two organelles. In progeny of heteroplasmic mothers, we found that plastid SNVs rapidly reached fixation or went extinct, whereas mitochondrial heteroplasmies were retained over generations. Within individual plants, we found that rates of plastid heteroplasmy varied substantially within an individual, indicative of vegetative segregation. Conversely, rates of mitochondrial heteroplasmy were relatively stable between tissues and throughout development. Our results suggest that the genetic bottlenecks associated with organelle DNA transmission both within and across generations are tighter in plastids than mitochondria, leading to the rapid loss of heteroplasmy in plastids. These patterns likely reflect differences in organelle dynamics, as plastids undergo little inter-organelle genetic exchange, while mitochondria participate in fusion events that facilitate homologous recombination between genome copies. To our knowledge, this study is the first to quantitatively assess the transmission dynamics of heteroplasmic variants arising from de novo mutation and their associated patterns of inheritance.

Room 21-02

Genome-wide local ancestry and evidence for cytonuclear disequilibria in African hybrid cattle populations (*Bos taurus*/*indicus*)

James A Ward[1], Gillian P McHugo[1], Michael J Dover[1], Tad S Sonstegard[2], Daniel G Bradley[3], Michael Salter-Townshend[4], David E MacHugh[1,5]

[1]Animal Genomics Laboratory, School of Agriculture and Food Science, University College Dublin, Dublin, Ireland. [2]Acceligen, 3388 Mike Collins Drive, Eagan, Minnesota, USA. [3]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [4]UCD School of Mathematics and Statistics, University College Dublin, Dublin, Ireland. [5]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland

Abstract

Cattle play an important role in African economies and society. Today, most African cattle are hybrids of the humpless *Bos taurus* (taurine) and the humped *Bos indicus* (indicine). These subspecies originated from domestications of *Bos primigenius primigenius* and *Bos primigenius namadicus*, respectively. The ancestral divergence time between these is estimated to be between 200,000 to 500,000 years ago, and as such there are significant differences between the genomes of the two. Despite the significant amount of indicine admixture in African cattle, they only carry taurine mitochondrial DNA. The efficient function of the mitochondria relies on the fine-tuned interactions that exist between the products of over 1000 nuclear genes and 37 mitochondrial genes. This bi-genomic system presents a challenge when there may be discordance between the nuclear and mitochondrial DNA, as may be the case in hybrid populations. Using high-density SNP data from over 500 cattle, representing 10 hybrid African cattle breeds, 4 pure taurine breeds and 4 pure indicine breeds we investigated the hypothesis that there has been adaptive introgression of taurine ancestry at regions of the genome that code for nuclear encoded mitochondrial genes, indicating that cytonuclear disequilibria may exist in hybrid African cattle populations. Using local ancestry analysis to infer the ancestry at individual SNPs and then using a bootstrapping approach, we generated distributions of mean ancestry deviation at groups of nuclear-encoded mitochondrial genes. Our results suggest that there is a significant deviation towards more taurine ancestry at nuclear-encoded mitochondrial genes in hybrid African cattle.

Room 21-03

Genetic bottleneck of mitochondrial DNA shaping the somatic mutations in lymphocyte

Zhongjie Tang
School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong, China

Abstract

Mitochondria are essential cellular organelles that play critical roles in hematopoiesis. Somatic mutations in mitochondrial DNA(mtDNA) randomly accumulate , drift and lead to heteroplasmy in individual cell during cell generations. How hematopoietic cells tackle mtDNA mutation load to function properly are not fully understood. In this study, we decipher somatic mtDNA mutation profiles in all PBMC lineages by analyzing mtscATAC-seq and related RNA sequencing data of PBMCs.

We observed a programmed reduction of mtDNA copy number and unexpected accumulation of high frequency mutations during lymphocyte development, which may be related to decrease of POLG2 gene expression in CLP , pro-B and pre-B cells. Our dilution model indicates the elevation of the high frequency site in lymphocytes can’t happened without mtDNA genetic bottleneck. we inferred that the mtDNA bottleneck was designed to facility the purify selection to strength protecting roles of lymphocyte during the immunization process. Our study provides new cognition about how somatic mtDNA mutation pass and accumulate during hematopoiesis.

Room 21-04

The dynamics of the mitochondrial genome throughout metazoan evolution unearth signatures of selection over gene organization: Insights into mitochondrial DNA regulation

Noam Shtolz, Dan Mishmar
Ben Gurion University, Beer Sheva, Israel

Abstract

It is widely perceived that mitochondrial DNA (mtDNA) gene content remained largely consistent among metazoans; While true for most vertebrates, the extent of such phenomena and the dynamics of mtDNA organization across all metazoans remain unknown. To this end, we analyzed mtDNA variability across all available metazoan mtDNAs (N = ~9000), intending to identify the evolutionary forces that shaped mitochondrial gene order and content. Our analysis was performed using an in-house pipeline designed to construct a curated metazoan mtDNA database. Cluster analysis based on mtDNA gene content and order revealed strong conservation of mtDNA organization within certain phyla. Accordingly, Chordata and Arthropoda formed a single cluster each, encompassing 99% and 95% of organisms, respectively. Unexpectedly, and despite the polycistronic nature of mtDNA transcription, we discovered conservation of long gene clusters (up to the entire mtDNA's length) that retained the same gene order per mtDNA strand in chordates. Contrarily, mollusks show low conservation even of short gene clusters. While considering shorter gene orders, we found that certain gene pairs, such as ND4L-ND4 and ATP8-ATP6, remained neighbors in nearly all chordates and arthropods (~99% in both), reflecting strong negative selection preventing their separation. Finally, while analyzing RNA-seq data from of metazoans, we found that mtDNA rearrangements affected transcriptional patterns, suggesting altered regulation of mtDNA gene expression during metazoan evolution. Taken together, our work reveals that selection shapes not only mtDNA gene content but also its organization across metazoan evolution, thus paving the path towards understanding mtDNA regulation across metazoan evolution.

Room 21-05

Mitochondrial mutational spectrum in human cancers is sensitive to cellular hypoxia.

Alina G. Mikhailova[1,2], Polina Lisitsa[1], Alina A. Mikhailova[3], Kristina Ushakova[1], Evgenii Tretyakov[4], Andrey Yurchenko[5], Vsevolod Makeev[2], Dmitrii Knorre[6], Ilia Mazunin[7], Sergey Nikolaev[5], Jacques Fellay[8], Konstantin Khrapko[9], Konstantin Gunbin[1], Konstantin Popadin[1,8]
[1]Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation. [2]Vavilov Institute of General Genetics RAS, Moscow, Russian Federation. [3]University of Münster, Münster, Germany. [4]Medical University of Vienna, Vienna, Austria. [5]Université Paris Saclay, Villejuif, France. [6]Moscow State University, Moscow, Russian Federation. [7]Skolkovo Institute of Science and Technology, Moscow, Russian Federation. [8]École polytechnique fédérale de Lausanne, Lausanne, Switzerland. [9]Northeastern University, Boston, USA

Abstract

The mutational spectrum of the mitochondrial genome (mtDNA) may be sensitive to the oxidative damage since mitochondria maintain the oxidative metabolism. Recently we have shown that the frequency of AH>GH (heavy strand notation) substitutions in mtDNA is positively correlated with cellular and organismal longevity (https://doi.org/10.1101/589168). We have shown that somatic AH>GH substitutions are more frequent at earlier stages of tumorigenesis and in cancers derived from slow-replicating tissues. The logic behind this finding was that long lived and slow-dividing cells have a rich aerobic environment, permitting a high oxidative metabolism, while short-lived fastly-dividing cells can run out of oxygen ending up in hypoxic conditions. To validate our hypothesis that mtDNA mutational spectrum is sensitive to hypoxia we tested mtDNA mutation rate and spectra in cancer samples with different levels of aerobic metabolism, ranging from normoxia to hypoxia. Using a collection of somatic mtDNA mutations and hypoxia scores derived for thousands individual cancer samples in the framework of the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium we observed that indeed mtDNA mutations depend on the level of hypoxia. Firstly, the fraction of AH>GH is decreased in highly hypoxic cancers. Secondly, the total mtDNA mutation rate is lower in hypoxic cancers (mainly due to drop in AH>GH). Altogether, we suggest that AH>GH substitutions, are sensitive to oxidative damage and thus can be a new marker of the redox stress in mtDNA.

Room 21-06

Mitonuclear DNA discordance is associated with decreased levels of mitochondrial DNA gene expression

Edmundo Torres-Gonzalez, Kateryna D Makova
Penn State University, University Park, PA, USA

Abstract

Most mitochondrial protein complexes include both nuclear and mitochondrial gene products, which coevolved to work in concert. This coevolution can be disrupted due to disparity in genetic ancestry between the nuclear and mitochondrial genomes in recently admixed populations, whose ancestors evolved separately. Such mitonuclear discordance might result in phenotypic effects. A number of nuclear-encoded mitochondrial genes and their products enable, and regulate levels of, expression of mitochondrial DNA (mtDNA) genes. We hypothesized that mitonuclear discordance results in impaired expression of genes encoded by mtDNA. To test this hypothesis, we utilized the data from the GTEx project, which contains expression levels for >100 African Americans and >700 European Americans. The varying proportion of African and European ancestry in recently admixed African Americans provides a range of mitonuclear discordance values, which can be correlated with mtDNA gene expression levels. We demonstrate that, for most mtDNA genes, expression in energetically-demanding tissues is significantly lower in African Americans with higher mitonuclear discordance. Moreover, we found a negative correlation between mtDNA gene expression and mitonuclear discordance. These results suggest that, consistent with our hypothesis, mitonuclear discordance affects levels of gene expression in admixed populations. However, some mtDNA genes (e.g. ND6, which is encoded by a separate transcript) do not follow this trend, suggesting a more nuanced relationship between expression and discordance. Nevertheless, these results represent the second example of a phenotypic effect of mitonuclear discordance on human admixed populations, with the first being its effects on mtDNA copy number.

Room 21-07

Mitonuclear interactions produce diverging responses to mild stress in Drosophila larvae

Enrique Rodríguez, Finley Grover Thomas, Florencia Camus, Nick Lane
University College London, London, United Kingdom

Abstract

Mitochondrial function depends on direct interactions between respiratory proteins encoded by genes in two genomes, mitochondrial and nuclear, which evolve in very different ways. Serious incompatibilities between these genomes can have severe effects on development, fitness and viability. The effect of subtle mitonuclear mismatches has received less attention, especially when subject to mild physiological stress. Here, we investigate how two distinct physiological stresses, a high protein diet and the glutathione precursor N-acetyl cysteine (NAC), affect development time, egg-to-adult viability, and the mitochondrial physiology of *Drosophila* larvae with an isogenic nuclear background set against three mitochondrial DNA haplotypes: one coevolved (WT) and two slightly mismatched (COX and BAR). Larvae fed the high-protein diet developed faster and had greater viability in all haplotypes. The opposite was true of NAC-fed flies, especially those with the COX haplotype. Unexpectedly, the slightly mismatched BAR larvae developed fastest and were the most viable on both treatments. These changes in larval development were linked to a shift to complex-I driven mitochondrial respiration in all haplotypes on the high-protein diet. In contrast, NAC increased respiration in COX larvae but drove a shift towards oxidation of proline and succinate. The flux of reactive oxygen species was increased in COX larvae treated with NAC, and was associated with an increase in mitochondrial DNA copy number. Our results support the notion that subtle mitonuclear mismatches can lead to diverging responses to mild physiological stress, undermining fitness in some cases, but surprisingly improving outcomes in other mismatched fly lines.

Room 21-08

Mother's curse is pervasive across a large mitonuclear Drosophila panel

Lorcan Carnegie, Max Reuter, Kevin Fowler, Nick Lane, M. Florencia Camus
University College London, London, United Kingdom

Abstract

The maternal inheritance of mitochondrial genomes entails a sex-specific selective sieve, whereby mutations in mitochondrial DNA can only respond to selection acting on females. In theory, this enables male-harming mutations to accumulate in mitochondrial genomes as long as they are neutral, beneficial, or only slightly deleterious to females. Ultimately, this bias could drive the evolution of male-specific mitochondrial mutation loads, an idea known as mother's curse. Earlier work on this hypothesis has mainly used small Drosophila panels, in which naturally sourced mitochondrial genomes were coupled to an isogenic nuclear background. The lack of nuclear genetic variation in these designs has precluded robust generalization. Here, we test the predictions of mother's curse using a large Drosophila mitonuclear genetic panel, comprising nine isogenic nuclear genomes coupled to nine mitochondrial haplotypes, giving a total of 81 different mitonuclear genotypes. Following a predictive framework, we tested the mother's curse hypothesis by screening our panel for wing size. This trait is tightly correlated with overall body size and is sexually dimorphic in Drosophila. Moreover, growth is heavily reliant on metabolism and mitochondrial function, making wing size an ideal trait for the study of the impact of mitochondrial variation. We detect high levels of mitonuclear epistasis, and more importantly, we report that mitochondrial genetic variance is larger in male than female Drosophila for eight out of the nine nuclear genetic backgrounds used. These results demonstrate that the mitochondrial DNA does modulate male life history traits in a more generalisable way than previously demonstrated.

Room 21-09

Somatic deletions in the human mitochondrial genome: the global secondary structure, G-quadruplexes and direct nucleotide repeats explain majority of breakpoints

Victor Shamanskiy[1], Alina A. Mikhailova[2,1], Kristina Ushakova[1], Alina G. Mikhaylova[3,1], Sergei Oreshkov[1], Dmitry Knorre[4,5], Evgenii O. Tretiakov[6], Ilia Mazunin[7,8], Konstantin Gunbin[9,1], Konstantin Khrapko[10], Konstantin Popadin[11,12,13,14]

[1]Immanuel Kant Baltic Federal University, Center for Mitochondrial Functional Genomics, Kaliningrad, Russian Federation. [2]University of Münster, Institute for Evolution and Biodiversity, Munster, Denmark. [3]Vavilov Institute of General Genetics RAS, Laboratory of Systems Biology and Computational Genetics, Moscow, Russian Federation. [4]Lomonosov Moscow State University, Belozersky Institute of Physico-Chemical Biology, Moscow, Russian Federation. [5]I M Sechenov First Moscow State Medical University, Institute of Molecular Medicine, Moscow, Russian Federation. [6]Medical University of Vienna Center for Brain Research, Department of Molecular Neurosciences, Vienna, Austria. [7]Skolkovo Institute of Science and Technology, Center of Life Sciences, Skolkovo, Moskovskaââ, Russian Federation. [8]Fomin Women's Health Clinic, Department of Genomic Medicine, Moscow, Russian Federation. [9]Institute of Molecular and Cellular Biology SB RAS, Theoretical, Novosibirsk, Russian Federation. [10]Northeastern University, Biology Department, Boston, USA. [11]IKBFU, Center for Mitochondrial Functional Genomics, Kaliningrad, Russian Federation. [12]University of Lausanne, The Center for Integrative Genomics, Lausanne, Lausanne, Switzerland. [13]Ecole Polytechnique Federale de Lausanne, School of Life Sciences, Lausanne, Lausanne, Switzerland. [14]Swiss Institute of Bioinformatics, Department of Life Sciences, Lausanne, Switzerland

Abstract

Ageing is often associated with clonal expansion of somatic mitochondrial (mtDNA) deletions, while their origin is still poorly known. Deletions are often flanked by direct nucleotide repeats, however, repeats solely do not provide an exhaustive explanation of deletion distribution. Here, we aim to decipher additional factors affecting formation of mtDNA deletions.

1) We hypothesized that repeats have higher chances to be realized into deletions in case of their spatial proximity. Analyzing distribution of human deletions we observed a hotspot (6-9kb and 13-16kb), which is not explained by direct repeats and might be driven by close contacts of these two regions during mtDNA replication. Using several approaches we reconstructed the secondary structure of the major arc and proposed that it is organized as a large-scale hairpin-like loop with a center close to 11 kb and stem between 6-9 kb and 13-16 kb.

2) We analysed the nested distribution of inverted and direct nucleotide repeats and observed that DIID combinations are the most fragile regions of the human mtDNA. These results are best compatible with the replication slippage mechanism where the nested pattern of direct and inverted repeats leads to the formation of deletion.

3) Drawing analogy with the recent finding in the nuclear genome we hypothesized that non-B DNA local structures such as G-quadruplexes may facilitate deletion formation in mtDNA.

Altogether we propose a multi-factorial model which explains distribution of somatic deletion in the human mtDNA and can be used to predict burden of somatic mtDNA deletion in different human haplogroups.

Room 21-10

Causes and consequences of mitochondrial mutation rate variation

Justin C Havird, Kendra Maclaine, Lydia Tressel, Ryan J Weaver
University of Texas at Austin, Austin, TX, USA

Abstract

Eukaryotes rely on mitochondria and mitochondrial genomes to generate the majority of their cellular ATP. Mitochondrial mutation rates are generally high in bilaterian animals, often exceeding those in the nuclear genome by an order of magnitude or more. The opposite pattern is found in most angiosperms as mitochondrial rates are generally lower than nuclear ones. However, much variation exists in both groups. Of particular interest are plant lineages that have experienced recent, rapid accelerations in mitochondrial mutation rates, putting them on par with those of bilaterians such as mammals. Here we investigated whether mitochondrial genome copy number predicts mitochondrial mutation rates in plants. We find that across multiple plant lineages, groups with accelerated mitochondrial mutation rates also show reduced genome copy numbers, sometimes estimated as only one to two mitochondrial genomes per nuclear genome. This suggests reduced recombination-mediated repair efficiency may drive mitochondrial mutation rate variation in plants. Lineages with fast-evolving mitochondrial genomes are also predicted to show strong signatures of mitonuclear coevolution, the hypothesis that changes in mitochondrial genomes should select for complementary changes in interacting nuclear-encoded genes (N-mt genes). We use evolutionary rate covariation (ERC) analyses in mammals to show that taxa with fast-evolving mitochondrial genomes also have fast-evolving N-mt genes, but relatively normal rates in nuclear-encoded genes that lack mitochondrial interactions. While mitonuclear coevolution seems to be a common feature of eukaryotes, nuclear compensation (a specific form of mitonuclear coevolution) seems to only characterize lineages with particularly fast mitochondrial mutation rates.

Room 21-11

New insight on the organization and evolution of Palaeognathae mitogenomes and implications on the ancestral gene rearrangement in Aves

Adam Dawid Urantowka[1], Aleksandra Kroczak[1,2], Paweł Mackiewicz[2]

[1]Department of Genetics, Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland.

[2]Department of Bioinformatics and Genomics, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland

Abstract

The most common bird gene order, found first in *Gallus gallus*, is regarded ancestral for all Aves. However, other rearrangements including a duplicated control region and neighboring genes have been reported in many representatives of avian orders. This raises a question about the actual prevalence of mitogenomic duplications and the validity of the current view on the avian mitogenome evolution. In this context, Palaeognathae is especially interesting because is sister to all other living birds, i.e. Neognathae. We applied an appropriate PCR strategy to look for omitted duplications in mitogenomes. The analyses showed the duplicated control regions with adjacent genes in *Crypturellus, Rhea* and *Struthio* as well as *ND6* pseudogene in three moas. The copies are very similar and were subjected to concerted evolution. Mapping the presence and absence of duplication onto the Palaeognathae phylogeny indicates that the duplication was an ancestral state for this avian group. This feature was inherited by early diverged lineages and lost two times in others. Data collected for other avian mitogenomes revealed that the last common ancestor of all birds and early diverging lineages of Neoaves could also possess the mitogenomic duplication. The duplicated control regions could increase effectiveness of replication and transcription as well as the number of replicating mitogenomes per organelle. In consequence, energy production by mitochondria may be also more efficient. This work was supported financially by the National Science Centre Poland (Narodowe Centrum Nauki, Polska) under Grant no. 2017/25/N/NZ8/01197.

Room 21-12

A billion-year trend of amino acid substitutions in the mitochondrial genome

Alina A Mikhailova[1], Alina G Mikhailova[2], Victor Shamanskiy[2], Kristina Ushakova[2], Alima Galieva[2], Konstantin Khrapko[3], Konstantin Gunbin[2], Jacques Fellay[4], Masashi Tanaka[5], Konstantin Popadin[4]
[1]Institute for Evolution and Biodiversity, Münster, Germany. [2]Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation. [3]Northeastern University, Boston, USA. [4]Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [5]National Institutes of Biomedical Innovation, Tokyo, Japan

Abstract

It has been shown that the rates of reciprocal amino acid substitutions in prokaryotic and eukaryotic organisms are not balanced, leading to the long-term increase (i.e. ‘gainers’) or decrease (i.e. ‘losers’) in the frequency of some amino acids. However, the evolutionary driving forces establishing this trend are still unknown. Here, focusing on the strongly asymmetrical mutational spectrum of the mitochondrial genome (an excess of G to A and T to C, light chain notation), we predicted the preferential direction of amino acid substitutions from losers (LeuTT, Phe, Cys, Trp, Gly, and Val) to gainers (Pro, His, Gln, Asn, Lys, and Thr). Analysing collections of nonsynonymous mtDNA mutations from human cancers (PCAWG), human pathogenic mutations (MitoMap database), human population polymorphisms, and mtDNA polymorphism from hundreds of vertebrate species we observed that the vast majority of substitutions are indeed in the expected direction: from losers to gainers. Moreover, the observed bias is the most pronounced in datasets where mutagenesis is stronger than selection (cancer and human pathogenic mutations for example). Comparing the amino acid composition of mtDNA genes between orthologs of mitochondrial genes in alpha-proteobacteria, fungi, plants, invertebrates, and five classes of vertebrates, we observed a global billion-year trend: losers become rarer while gainers become more frequent among these taxa. These results are in line with the accumulation of slightly-deleterious variants (i.e. from losers to gainers) in mtDNA from the moment of endosymbiosis emergence till the current days due to genetic drift, which becomes stronger from bacteria to vertebrates.

Room 21-13

Dissecting the Sequential Evolution of a Selfish Mitochondrial Genome in *Caenorhabditis elegans*

Joseph J Dubie, Ulfar Bergthorsson, Vaishali Katju
Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA

Abstract

Despite 1.4 billion years of endosymbiosis, selfishly acting mitochondrial genomes occasionally arise, outcompeting other mitotypes and increasing their intracellular frequency despite deleterious effects on their hosts. Unfortunately, mitochondrial genomes are challenging to engineer, preventing experiments aimed at distinguishing among theoretical predictions about what factors contribute to the advantage of one mitotype over another. We discovered a selfishly acting mitochondrial genome that arose spontaneously in a line of *Caenorhabditis elegans* following 346 generations of experimental evolution, comprising a large deletion, two indels, and a missense mutation. The regular cryopreservation of this line throughout experimental evolution provided an opportunity to dissect the sequential origin and fitness effects of the mutations comprising its selfish mitochondrial genome. We investigated whether subsequent mitochondrial mutations compensate for the deleterious effects of preceding ones, through dosage compensation or complex interactions, and whether a particular class of mutations confers selfish behavior in mitochondria. Through assaying four life-history traits in backcrossed lines with wild-type nuclear genomes and heteroplasmic mutation-bearing mitochondrial genomes, we found that the addition of each subsequent mitochondrial mutation reduced overall fitness in backcrossed lines. By tracking intraindiviual heteroplasmic frequency over 10 generations in bottlenecked populations, we found a mean increase in heteroplasmic frequency indicative of selfish drive across the mitochondrial mutation classes. We were unable to disentangle the contribution of the original deletion from the selfish behavior of each subsequent mutation, and it is possible that subsequent mutations lack a selfish drive of their own but serve to enhance that of the original deletion.

Room 21-14

Compensatory Evolution of Disease Associated Residues in the Oxidative Phosphorylation (OXPHOS) pathway

Abhilesh S Dhawanjewar, Colin D Meiklejohn, Kristi L Montooth
University of Nebraska-Lincoln, Lincoln, Nebraska, USA

Abstract

A substitution that is pathogenic in one genetic background may be neutral or even beneficial in the presence of other substitutions that compensate for its deleterious effects. Mutations in the mitochondrial genome are the cause of multisystemic disorders in humans, yet many human disease-associated residues (DARs) are observed as wild-type in non-human species. This phenomenon is commonly explained by compensatory epistatic evolution wherein interacting residues present in the non-human species ameliorate the deleterious effects of the DAR. We mapped 289 and 414 DARs in human mitochondrial (mtDNA) and nuclear (nucDNA) OXPHOS proteins onto their orthologs in 1062 and ~120 mammalian species, respectively, to identify instances of compensatory epistatic evolution. We found different levels of compensation for the two genomes with 145 (50.17%) of mtDNA DARs and 57 (13.76%) of nucDNA DARs present as wild-type in non-human mammals. The significantly higher proportion of compensated DARs (c-DARs) in the mtDNA cannot be completely explained by the larger number of species in our mtDNA mammalian dataset suggesting compensation to be a genomic quality rather than being a functional one. We also find that, on average, compensated mutations tend to be similar in physico-chemical properties and have relatively smaller effects on protein structure than uncompensated mutations. The compensated DARs also had a more variable structural neighborhood and higher levels of covariation with their structural neighbors, suggesting that molecular compensation is largely the result of structurally local mutations and the likelihood of compensation depends on the amino acid residue’s location in the protein.

Room 21-15

The species-specific burden of slightly deleterious mutations in mammals

Anastasia D Sokol[1], Alina Mikhailova[2,1], Konstantin Gunbin[3], Konstantin Popadin[4]
[1]Center for Mitochondrial Functional Genomics Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation. [2]The University of Münster, Institute for Evolution and Biodiversity, Münster, Germany. [3]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation. [4]School of Life Sciences Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

Abstract

Small-sized populations accumulate numerous slightly deleterious variants which can decrease genome quality and predispose species to extinction. Thus, approximation of the effective population size (Ne) through the species-specific ecological and genetic traits is of great importance for conservation biology. For example, it has been shown that low-Ne-species are characterized by large body mass and a high ratio of accumulation of nonsynonymous to synonymous substitutions (Dn/Ds). Here to clarify the genetic signature of low-Ne-species we analyze properties of mtDNA polymorphisms for 403 mammalian species with known generation length and IUCN Red List status.

We observed an increase in Dn/Ds and Grantham distance with generation length; moreover, this effect was strong not only on the species-specific level but also on the level of orders.

Also, we found significant correlations between the probability of being listed in one of the categories of IUCN and Grantham distance as well as Dn/Ds. We conclude that species prone to extinction are characterized by more numerous (high Dn/Ds) and more severe (increased Grantham) slightly deleterious variants.

Taking into account that long-lived species have low Ne and thus are more prone to extinction, the two above-mentioned results corroborate each other, emphasizing an association between genetic and ecological proxies of Ne. We conclude that both Dn/Ds and Grantham distance are highly sensitive to the effective population size and thus they can be useful for the assessment of the species-specific conservation status.

Room 21-16

Natural history and phyletic distribution of accessory subunits e and g that participate in the formation of dimeric/oligomeric F_1F_O ATP synthase.

José A. Hernández-Zúñiga[1], Oscar Flores-Herrera[1], Enrique García-Hernández[2], Héctor Riveros-Rosas[1]
[1]Depto. Bioquímica, Fac. Medicina, Universidad Nacional Autónoma de México, Mexico City, CDMX, Mexico. [2]Inst. Química, Universidad Nacional Autónoma de México, Mexico City, CDMX, Mexico

Abstract

F_1F_O ATP synthase complex is largely conserved across bacteria, mitochondria and chloroplast. In mitochondria, this complex can be found as a dimer/oligomer formed through accessory subunits. The dimers of mitochondrial ATP synthase are unique because they organize into long rows that induce membrane invaginations named cristae. However, a wide diversity at the morphology of mitochondrial cristae can be found among Eukaryotes.

On other hand, different dimeric/oligomeric F_1F_O ATP synthase arrangements have been reported in eukaryotes: ovine, bovine, *Saccharomyces cerevisiae*, *Chlamydomonas reinhardtii*, *Tetrahymena thermophila*, *Trypanosoma Brucei*, and *Euglena gracilis*, and different subunits seems to be involved in dimer formation.

To get insights about how these subunits were recruited into the ATP synthase complex, we performed a phylogenetic analysis of reported accessory subunits e and g involved in dimer/oligomer formation.

Amino acid sequences were retrieved either from Pfam (http://pfam.xfam.org/), or using BlastP searches against the complete genomes available at NCBI's RefSeq genome database (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq). Progressive multiple amino acid sequence alignments were performed with ClustalX (http://www.clustal.org/clustal2/). Phylogenetic analyses were conducted using the MEGA_X (http://www.megasoftware.net).

Accessory subunit e and g were found only in eukaryotes; therefore, these proteins originated after the last eukaryotic common ancestor. Subunit g was found in animals, fungi, and plants. On other hand, subunit e was found only in animals and fungi. In both cases conserved the motive GXXXG. This suggest a broad distribution of dimeric/oligomeric ATP synthase complexes, but probably evolved independently several times through the evolution of eukaryotes.

Room 21-17

Convergent adaptation in mitochondria of phylogenetically distant birds: does it exist?

Valentina Burskaia[1,2], Ilja Artyushin[3], Nadezhda A Potapova[1], Kirill Konovalov[4], Georgii A Bazykin[1,2]
[1]Institute for Information Transmission Problems, Moscow, Moscow, Russian Federation. [2]Skolkovo Institute of Science and Technology, Moscow, Moscow, Russian Federation. [3]Lomonosov Moscow State University, Moscow, Moscow, Russian Federation. [4]Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Abstract

In a wide range of taxa, proteins encoded by mitochondrial genomes are involved in adaptation to lifestyle that requires oxygen starvation or elevation of metabolism rate. It remains poorly understood to what extent adaptation to similar conditions is associated with parallel changes in these proteins. We search for genetic signal of parallel or convergent evolution in recurrent molecular adaptation to high altitude, migration, diving, wintering, unusual flight abilities, or loss of flight in mitochondrial genomes of birds. Developing on previous work, we design an approach for detection of recurrent coincident changes in genotype and phenotype, indicative of an association between the two. We describe a number of candidate sites involved in recurrent adaptation in *ND* genes. However, we find that the majority of convergence events can be explained by random coincidences without invoking adaptation.

Room 21-18

The alternative human mitochondrial proteome

Laura Kienzle[1], Stefano Bettinazzi[1], Marie Brunet[2], Xavier Roucou[2], Christian Landry[3], Thierry Choquette[1], Annie Angers[1], Sophie Breton[1]

[1]Université de Montréal, Montréal, Québec, Canada. [2]Université de Sherbrooke, Sherbrooke, Québec, Canada. [3]Université de Laval, Laval, Québec, Canada

Abstract

Small open reading frames (smORFs) and alternative open reading frames (altORFs) are emerging as a new class of important genes. Recent research revealed the presence of several small functional ORFs inside the human mitochondrial genes 12S and 16S. In this study, we aimed to investigate the existence of an alternative proteome coded by the human mitochondrial genome. From the 227 altORFs with a minimum size of 60 nucleotides predicted in the human genome, one alternative protein of 99 amino acids, MTALTND4, was identified by immunoblotting in HeLa and HEK293T cell lysates. MTALTND4 is encoded in the mitochondrial *nd4* gene, translated inside mitochondria and found in the cytoplasm as well as in plasma. Preliminary physiological analyses revealed a dose-dependent impact of this peptide on mitochondrial respiration, suggesting a role in energy metabolism regulation and response to hypoxic events. MTALTND4 is the first discovered human mitochondrial alternative protein whose coding sequence is nested in the coding sequence of an annotated protein. This discovery confirms that the coding potential of the human mitochondrial genome is larger than anticipated, as potentially the role and functions of mitochondria themselves. Future research on these peptides could help to deepen our knowledge on mitochondria’s implications in cellular functions, diseases and evolution.

Room 21-19

Comparison of transcription, polymorphism, and synonymous codon usage in mito-nuclear OXPHOS genes of a DUI species

Ran Xu[1], Mariangela Iannello[1], Justin C Havird[2], Liliana Milani[1], Fabrizio Ghiselli[1]
[1]University of Bologna, Bologna, Emilia-Romagna, Italy. [2]The University of Texas at Austin, Austin, Texas, USA

Abstract

In most eukaryotes, oxidative phosphorylation (OXPHOS) is the main energy production process and involves both mitochondrial and nuclear genomes. The close interaction between the two genomes is critical for the coordinated function of the OXPHOS process. Some bivalves show a doubly uniparental inheritance (DUI) of mitochondria, where two highly divergent mitochondrial genomes, one inherited through eggs (F-type) and the other through sperm (M-type) coexist in the same individual, thus nuclear OXPHOS genes need to cooperate with two different mitochondrial genomes. In this study, we compared the mitochondrial and nuclear OXPHOS genes in terms of transcription, polymorphism, and synonymous codon usage in the DUI species Ruditapes philippinarum. Mitochondrial and nuclear OXPHOS genes showed remarkably different transcription profiles and strong co-transcription signals were observed within mitochondrial (separate for F- and M-type) and within nuclear OXPHOS genes but not between mitochondrial and nuclear OXPHOS genes, suggesting that the coordination between mitochondrial and nuclear OXPHOS might involve post-transcriptional and/or translational regulation. McDonald-Kreitman and other tests indicated that F-type and M-type OXPHOS genes might undergo different selection forces. Polymorphism and divergence in M-type genes deviate significantly from neutrality and the forces driving this deviation might be complex, involving both demographic events and selection. Besides, other forces acting on OXPHOS genes were also detected: both mitochondrial and nuclear OXPHOS genes were subject to context-dependent mutations, while translational selection was only detected in nuclear OXPHOS genes.

Room 21-20

The mtDNA-encoded COX2 protein: bivalves have the longest

Mélanie Tassé[1], Thierry Choquette[1], Annie Angers[1], Eric Pante[2], Sophie Breton[1]
[1]Université de Montréal, Montréal, Québec, Canada. [2]LIENSs Laboratory, CNRS - La Rochelle University, La Rochelle, France

Abstract

Several bivalve species (>100) possess a unique mode of mitochondrial transmission named doubly uniparental inheritance (DUI). In DUI, the maternal mitochondria (and their corresponding female or F mtDNA) are transmitted to offspring; exceptional in metazoan, the paternal mitochondria (and their male or M mtDNA) is also transmitted to male offspring. *Scrobicularia plana*, a bivalve of the Semelidae family, was recently found to have an M genome much longer than the F genome with a difference of > 10,000 bp. This difference is in part due to the exceptionally long size of the male *cox2* gene (*Mcox2*), which possesses an insertion of ~ 4,800 bp (compared to the Mcox2 typically encoded in the M mtDNA of DUI bivalves) making it the longest Metazoan *cox2* gene known to date. This insertion is conserved among individuals within our study population (Fouras, France, NE Altantic) and in reading frame with MCOX2 ORF. Western blot analysis and RT-PCR confirm that the long MCOX2 transcript and protein (~ 220 kDa) are expressed in *S. plana* male gametes. Analysis of synonymous and non-synonymous substitution rates suggests that the insertion in *Mcox2* evolves under purifying selection. The implications of these results will be discussed.

Room 22-01

Assessing fitness effects of transposable elements in S. cerevisiae

Michael Tene, Gregory Lang
Lehigh University, Bethlehem, Pensylvania, USA

Abstract

Transposable elements make up a large portion of many genomes, including 45% of the human genome and 80% of the corn genome. While many transposons are inactive due to a disruption in the mobilization genes or due to a second transposon insertion within the first, they play an important role in genome evolution and may impose a fitness cost to the host. Host fitness could be affected in two ways: due to the increased burden of a large number selfish transposons or through interference with neighboring genes. We have characterized the proximity of transposable element sequences to coding sequences in the Saccharomyces cerevisiae reference strain. We plan to use CRIPSPR/Cas9 to construct ~100 barcoded strains carrying single-transposon excisions. We will quantify fitness effects of individual transposable elements across a panel of environments using bulk fitness assays.

Room 22-02

Modelling the Population Genetics of Non-Equilibrium Molecular Evolution in Large and Hyper-Diverse Populations: A Finite Element Method Approach

Roman Frolov[1], Bianca DeSanctis[2], Ivan Krukov[3], Erin Brintnell[1], A.P. Jason de Koning[1]
[1]University of Calgary, Calgary, AB, Canada. [2]University of Cambridge, Cambridge, United Kingdom. [3]McGill University, Montreal, QC, Canada

Abstract

Our lab recently characterized the conditions under which the widely applied rate of molecular evolution popularized by Kimura, k=2Nv PFix, breaks down in a range of bi-allelic models. This led to the identification of an intermediate range of population mutation rates ($\theta \geq 0.05$) where mutation is neither weak nor strong by conventional standards, and where: the rate of neutral evolution is slower than the mutation rate, and the rate of non-neutral evolution is reduced by a factor that depends on the rate of mutation and strength of selection. Thus, in populations and for mutation types where the population mutation rate is high-normal (e.g., in RNA viruses, hyperdiverse eukaryotes, for microsatellites, etc.), a variety of the rules of molecular evolution are predicted to be violated. Here, we extend this work to arbitrary multi-allelic diffusion models. Accurate numerical analysis of multi-allelic models including general mutation and selection is notoriously difficult. Most notably, when mutation rates are high, the “convection” (i.e., associated with mutation and selection) part of the system becomes dominant, which leads to the formation of boundary layers, numerical instabilities, and poor performance of conventional linear solvers. We address these challenges and develop an efficient high-performance numerical algorithm based on the Finite Element Method (FEM) with adaptive mesh refinement. We show it has good performance and that it significantly expands the scope and range of models that can be studied. This approach is used to examine molecular evolution in large and hyper-diverse populations under non-equilibrium conditions caused by episodic fitness shift.

Room 22-03

Fitness equilibrium and the role of population size in determining long-term evolvability

Jason A Tarkington
Stanford, Stanford, CA, USA

Abstract

Predicting evolutionary fitness trajectories is difficult because it requires understanding how biological processes at different scales (mutation within the genome and selection in populations) interact to produce changes in population fitness. This problem is particularly interesting because mutational and selective processes are dependent on populations size, fitness, and relevant genetic variation, which therefore may hold the key to predicting how the fitness of a population will evolve. To understand how population size and genetic variation impact fitness trajectories replicate populations of *Sacchromyces cerevisiae* founded from genotypes across a large range of fitness will be allowed to evolve across a range of populations sizes (N_e). Previous work has shown that fitness tends to increase in large populations and decrease in small populations, and that fitness equilibrium exists when the power of natural selection is balanced by the power of drift and Muller's ratchet and fitness is stable as evolution continues. Here, we ask whether multiple fitness equilibrium exist corresponding to different populations sizes and hyperbolic fitness trajectories or, corresponding with power law fitness trajectories, whether a single fitness unstable fitness equilibrium exists at an intermediate population size above which fitness increases indefinitely and below which the population is bound for extinction.

Room 22-04

Quantitative genetics for studying life history evolution in preindustrial human populations

Walid Mawass, Emmanuel Milot
Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

Abstract

Detection in nature of microevolutionary change by natural selection has been a constant aim for many biologists, requiring proper estimation of genetic changes occurring within a population. Quantitative genetics (QG) has provided both theory and methodology to extract information about the raw material on which selection can work. These types of analyses, mainly exploiting pedigreed populations in nature, have helped provide evidence of contemporary microevolution change in response to selection. One example of such a population is a preindustrial human population from a small, isolated island in the St-Lawrence River, île aux Coudres (IAC). Using the reconstructed genealogies of this population in QG analyses, the phenotypic change in the age at first reproduction (AFR) of married women, which went from 26 y to 22 y during 140 years, was shown to be partially due to a genetic change in response to selection (younger mothers had higher reproductive success). According to theory, fluctuations in environmental conditions can induce changes in genetic parameters and the rate of evolutionary change. Using this population, we show that genotype-by-environment interactions (GxE) not only affected the genetic covariance between AFR and reproductive success but also the expected genetic change from one generation to the next. Also, we have access to the reconstructed pedigrees of populations sharing a genetic background with the population of IAC. We show results of power and precision analysis conducted on these pedigrees for the first time, revealing the role of sample size and depth on genetic parameter estimations (ex. additive genetic variance).

Room 22-05

Collateral fitness effects of mutations

Jacob D. Mehlhoff, Marc Ostermeier
Johns Hopkins University, Baltimore, Maryland, USA

Abstract

The underlying mechanisms by which mutations lead to fitness effects are typically attributed to changes in protein specific activity or abundance. Here, we reveal the importance of a mutation's collateral fitness effects, which we define as effects that do not derive from changes in the protein's ability to perform its physiological function. We constructed libraries of all single-codon missense mutations in four antibiotic resistance genes native to *Escherichia coli*. We used deep mutational scanning to comprehensively measure the collateral fitness effects of these mutations during growth competition experiments in the absence of antibiotic. We found that over 42% of all missense mutations in TEM-1 beta-lactamase were deleterious to growth rate in the absence of antibiotic, indicating that for some proteins, collateral fitness effects occur as frequently as effects on protein activity and abundance. Deleterious mutations caused improper post-translational processing, incorrect disulfide-bond formation, protein aggregation, changes in gene expression, and pleiotropic changes in cell phenotype. We identified deleterious collateral fitness effects, but at a lower frequency, in the CAT-I, NDM-1, and aadB antibiotic resistance genes. The prevalence of deleterious collateral fitness effects suggests they may play a role in constraining protein evolution, particularly for highly-expressed proteins, proteins under intermittent selection, and for proteins whose contribution to fitness is buffered against deleterious effects on protein activity and abundance. By identifying the frequency, magnitude, and mechanisms of collateral fitness effects and how they differ across genes, we can understand their role in genetic selection and the shaping of protein evolutionary pathways.

Room 22-06

Environmental dependence of collateral fitness effects

Erh-Yeh Tsou, Jacob Mehlhoff, Jonah Facciolli, Jacob Fetterolf, Dahlia Rohm, Marc Ostermeier
Johns Hopkins University, Baltimore, MD, USA

Abstract

Genomes are often exposed to complex and ever-changing environments. Genes which are essential in some conditions, may be superfluous or even deleterious to fitness in others. Similarly, mutational effects may be environment-dependent. Uncovering the mechanisms by which fitness effects arise and their dependence on growth environment is key in understanding how cells adapt to selective pressures. In previous work, we found that over 42% of all missense mutations within the *Escherichia coli* TEM-1 antibiotic resistance protein were deleterious to growth rate in the absence of antibiotic. We define such fitness effects that arise independent of changes to the protein's ability to perform its physiological function as collateral fitness effects. Notably, mutations which led to deleterious collateral fitness effects caused improper post-translational processing, incorrect disulfide-bond formation, protein aggregation, and changes in gene expression. Here, we used deep mutational scanning to comprehensively measure collateral fitness effects during growth at 30°C, 37°C, and 42°C. We reveal how the frequency, magnitude, and mechanisms of collateral fitness effects differ with changes in growth conditions.

Room 22-07

Population and subspecies diversity at mouse centromere satellites and its influence on CENP-A association

Uma P. Arora[1,2], Beth L. Dumont[1,2]
[1]The Jackson Laboratory, Bar Harbor, ME, USA. [2]Tufts University, Boston, MA, USA

Abstract

Mammalian centromeres are satellite-rich chromatin domains that execute conserved roles in kinetochore assembly and chromosome segregation. Centromere satellites evolve rapidly between species, but little is known about population-level diversity across these loci. We developed a *k*-mer based method to quantify centromere copy number and sequence variation from whole-genome sequencing data. We applied this method to diverse inbred and wild house mouse (*Mus musculus*) genomes to profile diversity across the core centromere (minor) satellite and the pericentromeric (major) satellite repeat. We show that minor satellite copy number varies more than 10-fold among inbred mouse strains, whereas major satellite copy numbers span a 3-fold range. In contrast to widely held assumptions about the homogeneity of mouse centromere repeats, we uncover marked minor satellite sequence heterogeneity within single genomes. Intriguingly, we also find that wild-caught mice harbor dramatically reduced minor satellite copy number and elevated satellite sequence heterogeneity compared to inbred strains, suggesting that inbreeding may reshape centromere architecture in pronounced ways. Using chromatin immunoprecipitation of the centromere-specific histone H3 variant CENP-A in a subset of diverse inbred strains, we show that CENP-A associates with a narrow repertoire of centromere satellite sequences in each strain. Taken together, our results provide an initial portrait of centromere variation across *Mus musculus,* including the role of centromere variation on CENP-A association.

Room 22-08

Repeated out-of-Africa expansions of Helicobacter pylori driven by replacement of deleterious mutations

Daniel Falush[1], Harry Thorpe[2], Koji Yahara[3], Sebastian Suerbaum[4], Kaisa Thorell[5]
[1]Institute Pasteur Shanghai, Shanghai, China. [2]Institute of Genomics, Oslo, Norway. [3]National Institute of Infectious Diseases, Tokyo, Japan. [4]Hannover Medical School, Hannover, Germany. [5]University of Goteborg, Gothenberg, Sweden

Abstract

All genomes mutate but the consequences of the resulting deleterious mutational load are poorly understood. *Helicobacter pylori* lives in the human stomach, has a higher mutation rate than most bacteria and has accompanied anatomically modern humans in migrations including the out-of-Africa expansion more than 50,000 years ago. *H. pylori* from East Asia have accumulated at least 500 more non-synonymous mutations than African strains, which we propose is due to reduced efficacy of selection during the out-of-Africa bottleneck. *H. pylori* from Europe and the middle East trace a substantially higher fraction of ancestry from modern African populations than the humans that carry them, which we find is due to at least three separate admixture events. African ancestry is elevated at positions in the genome where non-synonymous mutations are at high frequency in Asia. We propose that this is due to replacement of deleterious mutations that accumulated during the bottleneck, with the high overall African ancestry proportion due clonal expansion of strains of African origin. We use simulations to show that a Muller's ratchet like effect can lead to long-term segregation of deleterious mutations within bacterial populations after a bottleneck, despite high rates of homologous recombination, but that population fitness can be restored by migration of small numbers of bacteria from non-bottlenecked populations. Our results demonstrate that population bottlenecks can have long-term genomic and demographic consequences, even in very numerous species.

Room 22-09

Gene surfing of underdominant alleles promotes formation of hybrid zones

Kimberly J. Gilbert[1], Antoine Moinet[2], Stephan Peischl[2]

[1]Institute of Plant Sciences, University of Bern, Bern, Bern, Switzerland. [2]Institute of Ecology and Evolution, University of Bern, Bern, Bern, Switzerland

Abstract

The distribution of genetic diversity over geographic space has long been investigated in population genetics and serves as a useful tool to understand evolution and history of populations. Within some species or across regions of contact between two species, there are instances where there is no apparent ecological determinant of sharp changes in allele frequencies or divergence. To further understand these patterns of spatial genetic structure and potential species divergence, we develop a model for the establishment of clines in non-neutral genetic diversity which occur via heterozygote disadvantage due to the surfing of underdominant alleles during range expansions. We provide analytical approximations for the fixation probability of underdominant alleles and demonstrate that such alleles can readily establish via gene surfing in 1D range expansions. We extend these results to multiple loci via a mixture of analytical theory and individual-based simulations. We study the interaction between the strength of selection against heterozygotes, migration rates, and local recombination rates on the formation of stable hybrid zones. A key result of our study is that clines created by surfing at different loci can attract each other and align after expansion, if they are sufficiently close in space and in terms of recombination distance. Our findings shed new light on how range expansions interact with selection and recombination, and suggest that range expansions can set the stage for parapatric speciation due to the alignment of multiple selective clines at one geographic location, even in the absence of ecologically divergent selection.

Room 22-10

Natural selection and landscapes of diversity in the great apes

Murillo F Rodrigues, Andrew D Kern, Peter L Ralph
University of Oregon, Eugene, OR, USA

Abstract

Patterns of genetic diversity along chromosomes – or landscapes of diversity – are highly correlated between related groups of species, such as flycatchers, monkeyflowers and aspens. Even though these correlations should decay with divergence time under neutrality, empirical data demonstrate strong correlations persisting over long evolutionary timescales. What could be maintaining correlations in diversity across pairs of related taxa? Natural selection is known to couple genomic features, such as recombination rate and functional density, with diversity. These features are expected to be largely shared between related taxa, so selection and its linked effects could in principle maintain correlations over longer periods of time. We used great apes' genomic data to describe the landscapes of diversity and the correlations between pairs of species. Using forward-in-time simulations of the great apes' evolutionary history under different selective regimes (neutral, background selection, sweeps), we explore how different evolutionary processes shape correlations over time. We found high correlations among the great apes' landscapes of diversity, even between pairs of species that diverged over 10Mya. Forward-in-time simulations of the great apes' history show that the observed correlations are far in excess of what can be generated under a neutral null model. In simulations with selection and a realistic distribution of functional sites, we found that correlations between landscapes of diversity might be maintained over longer periods of time. Future work includes simulation-based inference of selection parameters in our model.

Room 22-11

Sex-specific phenotypic effects and evolutionary history of an ancient polymorphic deletion of the human growth hormone receptor

Marie Saitou[1], Skyler Resendez[1], A.J. Pradhan[1], F. Wu[1], N.C. Lie[2], N.J. Hall[3], Qihui Zhu[4], L. Reinholdt[5], Y. Satta[6], S. Nakagome[7], N. Hanchard[8], G. Churchill[5], C. Lee[4], G.E. Atilla-Gokcumen[1], X. Mu[1], Omer Gokcumen[1]

[1]University at Buffalo, Buffalo, NY, USA. [2]Baylor College of Medicine, Houston, TX, USA. [3]Baylor College of Medicine, Houston, NY, USA. [4]The Jackson Laboratory of Genomic Medicine, Farmington, CT, USA. [5]The Jackson Laboratory, Bar Harbor, ME, USA. [6]SOKENDAI, Kanagawa Province, Japan. [7]Trinity College, Dublin, Ireland. [8]NHGRI, Bethesta, MD, USA

Abstract

The deletion of the third exon of the growth hormone receptor (GHR) is one of the most common genomic structural variants in the human genome. This deletion (GHRd3) has been linked to response to growth hormone, placenta size, birth weight, growth after birth, time of puberty, adult height, and longevity. However, its evolutionary history and the mechanisms through which it affects phenotypes remain unresolved. We analyzed thousands of genomes and provide evidence that this deletion was nearly fixed in the ancestral population of anatomically modern humans and Neanderthals. However, it underwent a paradoxical adaptive reduction in frequency approximately 30 thousand years ago in East Asia that roughly corresponds with the emergence of archaeological evidence for multiple modern human behaviors, dramatic changes in climate, and a concurrent population expansion. We provide evidence that GHRd3 is associated with protection from edematous severe acute malnutrition primarily in males. Using a mouse line engineered to contain the deletion, we found Ghrd3's effect on the liver transcriptome of male mice grown without any calorie restriction mimics response to calorie restriction through regulation of circadian pathways. In contrast, under calorie restriction, Ghrd3 leads to the female-like gene expression in male livers. As a likely consequence, the dramatic weight difference between male and female mice disappears among GHRd3 mice under calorie restriction. Our data provide evidence for sex- and environment-dependent effects of GHRd3and are consistent with a model in which the allele frequency of GHRd3 varies throughout human evolution as a response to fluctuations in resource availability.

Room 22-12

Ancient genomics reveal positive selection on immune genes during the Black Plague

Tauras P Vilgalys[1], Jennifer Klunk[2], Mari Cobb[1], Rebecca Redfern[3], Julia Gamble[4], Jesper L Boldsen[5], Sharon DeWitte[6], Anne Dumaine[1], Christian Demure[7], Javier Pizarro-Cerdá[7], Guy Rouleau[8], Hendrik Poinar[9], Luis Barreiro[1]

[1]University of Chicago, Chicago, Illinois, USA. [2]Daicel Arbor Sciences, Anne Arbor, Michigan, USA. [3]Museum of London, London, United Kingdom. [4]University of Manitoba, Winnipeg, Manitoba, Canada. [5]University of Southern Denmark, Odense, Denmark. [6]University of South Carolina, Columbia, South Carolina, USA. [7]Institut Pasteur, Paris, France. [8]McGill University, Montreal, Quebec, Canada. [9]McMaster University, Hamilton, Quebec, Canada

Abstract

Infectious diseases have exerted strong selective pressures throughout human evolution. However, it remains difficult to link instances of positive selection to specific infectious agents, especially when those agents are no longer prevalent in modern populations or have low mortality rates. The Black Death, the greatest mortality event in recorded history, offers such an opportunity, as it killed an estimated 30-50% of the European population during the 14th century. We enriched and sequenced neutral and immune-related loci across the genome from 321 individuals from London, UK and various localities across Denmark who died shortly before, during, or after the Black Death. We found that genetic variants located near immune genes were more differentiated between pre- and post-plague samples than expected under neutrality. Eight variants show particularly high differentiation in both the English and Danish cohorts, representing the strongest candidates for positive selection. Many of these loci still show signatures of positive selection in contemporary populations. We further tested these variants using a combination of bulk and single-cell sequencing, and identified six genes near candidate loci which are differentially regulated following infection of macrophages and PBMCs with *Y. pestis*, as well as candidate loci that are associated with variation in gene expression levels (eQTL) in infected cells. These variants have been associated with modern human diseases including psoriasis, Crohn’s disease, and rheumatoid arthritis. Collectively, our analyses identify a number of loci which likely experienced positive selection during the Black Death, some of which continue to influence disease risk in modern individuals.

Room 22-13

Pervasive incomplete lineage sorting in primates and its genomic and functional determinants

Iker Rivas-González[1], Marjolaine Rousselle[1], Fang Li[2,3], Long Zhou[3], Josefin Stiller[2], Dongdong Wu[4,5,6], Kasper Munch[1], Mikkel H Schierup[1], Guojie Zhang[2,4,6,7]

[1]Bioinformatics Research Center, Aarhus University, Aarhus, Denmark. [2]Villum Center for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [3]BGI-Research, BGI-Shenzhen, Shenzhen, China. [4]State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China. [5]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan, China. [6]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan, China. [7]China National Genebank, BGI-Shenzhen, Shenzhen, China

Abstract

For closely related species, individual genes might be discordant to the phylogeny due to incomplete lineage sorting (ILS). To investigate the genomic and functional basis of ILS, we analyzed 50 de novo sequenced primate species using CoalHMM. Our results show that ILS is widespread across the primate phylogeny and that it is the main source of gene tree incongruence, typically ranging 10-40% in old-world monkeys, and reaching over 50% in gibbons and new-world monkeys. We correlated ILS and recombination, revealing that ILS proportions in highly recombining regions are up to 10 percentage points higher than in lowly recombining ones, implying that there is up to a 2-fold variation in the local ancestral population size along the genome. Moreover, even though we expect decreased ILS on the X chromosome due to its effective population size being 25% lower than in autosomes, we find that this reduction is much stronger than predicted, suggesting that other factors such as the mating system are at play. We also explored both the specific ILS landscape in each individual node, as well as the general patterns of consistently high- or low-ILS tracts across the whole primate phylogeny. Additionally, we functionally characterize the ILS in genes and gene categories. This reveals that exons have lower ILS proportions than intergenic regions, and that immunity genes are consistently high in ILS along the primate phylogeny, while housekeeping genes generally show lower levels of ILS.

Room 22-14

Site level factors that affect the rate of adaptive evolution in humans and chimpanzees; the effect of contracting population size.

Vivak Soni, Adam Eyre-Walker
University of Sussex, Brighton, United Kingdom

Abstract

It has previously been shown in Drosophila species that the rate of adaptive evolution is lower between amino acid pairs that are more dissimilar. We investigated whether this pattern is found in the divergence between humans and chimpanzees using an extension of the MacDonald-Kreitman test. We find that the rate of adaptive evolution, relative to the rate of mutation, is lower for amino acid pairs that are more dissimilar in terms of their polarity, volume and a more generalised measure of amino acid dissimilarity, the ratio of the numbers of non-synonymous to synonymous polymorphisms (pN/pS). However, the strength of this latter correlation is significantly shallower than in Drosophila. We suggest that this is due to the population contraction that has occurred since humans and chimpanzees diverged. We demonstrate theoretically that population size reduction can generate an artefactual positive correlation between the rate of adaptive evolution and any factor that is correlated to the mean strength of selection acting against deleterious mutations, even if there has been no adaptive evolution (the converse is also expected). Our measure of amino acid dissimilarity, pN/pS, is negatively correlated to the mean strength of selection, and hence we would expect the correlation between the rate of adaptive evolution to also be negatively correlated to pN/pS, if there is no adaptive evolution. The fact that our rate of adaptive evolution is positively correlated to pN/pS suggests that the correlation does genuinely exist, but that it has been attenuated by population size contraction.

Room 22-15

Neural ADMIXTURE Clustering

Albert Domingez[1], Daniel Mas[2], Carlos Bustamante[2], Xavier Giro-i-Nieto[1], Alex Ioannidis[2]
[1]Polytechnic University of Catalonia, Barcelona, Catalonia, Spain. [2]Stanford University, Stanford, California, USA

Abstract

ADMIXTURE and Structure are widely used unsupervised clustering methods for characterizing the ancestry composition of genomic sequences. ADMIXTURE and Structure decompose sequences of SNPs into a fractional cluster assignment for each sequence and a set of centroids, based on the frequencies of SNPs, for each cluster. This provides an interpretable representation that allows geneticists to characterize the population's structure. However, with the increasing number of whole-genome data available with a high-density of variants sequenced, such methods become computationally infeasible. Furthermore, many runs with different hyperparameters are required to properly depict the population clusters, leading to days of compute time. In this work we present Neural ADMIXTURE, a neural network autoencoder that follows the same modeling assumptions as ADMIXTURE, providing similar (or even better) results while reducing the compute time by several orders of magnitude. Furthermore, the network can include several outputs providing the equivalent result as running the original admixture with different numbers of clusters and the models can be stored to perform cluster assignment with fast computational time.

SYMP1-1

The symbiont's symbiont: investigating the role of temperate phages in rhizobia

Ellie Harrison
university of Sheffield, Sheffield, United Kingdom

Abstract

Rhizobia - nitrogen fixing symbionts of legumes - are one of the most important and best studied symbionts on the planet. Like many bacteria, rhizobial genomes themselves are home to a diverse menagerie of genetic symbionts - mobile genetic elements - such as temperate phages. Like any symbiotic partnership, these relationships can alter the ecology, and ultimately, the evolution of the hosts they inhabit. Here I will present our work to understand the role of temperate phages in the rhizobia-legume symbiosis. We find extensive evidence of temperate phage communities in Rhizobium leguminosarum populations and use experimental evolution to examine how these genetic symbionts can alter the bacteria-plant interaction.

SYMP1-2

What drives phage susceptibility in encapsulated bacteria? A study of environmental phages and *Klebsiella pneumoniae*

Beatriz Beamud[1,2], Neris García-González[1,2], Lourdes Tordera[1], Mar Gómez-Ortega[1], Fernando González-Candelas[1,2,3], Pilar Domingo-Calap[2], Rafael Sanjuan[2]

[1]FISABIO-Public Health/ Univ. Valencia. Joint Research Unit Infection and Public Health, Valencia, Spain. [2]Institute for Integrative Systems Biology, I2SysBio, Universitat de València-CSIC, Paterna, Spain. [3]CIBER in Epidemiology and Public Health, Genomics and Health, Valencia, Spain

Abstract

The bacterial capsule is one of the main virulence factors of bacteria and the first barrier encountered by phages. We have used Klebsiella pneumoniae, a bacteria included in the ESKAPE pathogen group, as a model to understand phage susceptibility in an encapsulated host. For this purpose, we isolated 70 phages from 13 diverse groups and established a collection of 139 K pneumoniae strains representative of the capsule and genomic diversity of the species. We tested all possible combinations and found that phage-bacteria interactions are rare (<0.05%), being largely modular due to the capsule of the host. These interactions were variable depending on the technique used for host-range determination. Phages are able to strip capsules due to the specificity conferred by depolymerase enzymes but there are further post-adsorptive mechanisms which largely limit phage infection. We found homologous tail fiber domains in distinct phage groups with overlapping host range, suggesting a frequent gene-for-gene model of coevolutionary adaptation. Finally, we challenged ~150 spontaneous acapsular mutants to phages and found that capsule is a widespread requirement but some phages are able to infect acapsular mutants. This character had low phylogenetic inertia with the exception of G8 phages which escaped from capsular modularity and were more frequently associated with successful acapsular infections. The latter varied greatly among bacterial backgrounds, suggesting a polyvalent mode of action for some phages. These results reveal the complex interactions between bacteria and their phages and have implications for the biomedical and technological applications of phages.

SYMP1-3

Interplay between the cell envelope and mobile genetic elements shapes gene flow in populations of a nosocomial pathogen.

Matthieu Haudiquet, Amandine Buffet, Olaya Rendueles, Eduardo P.C. Rocha
Institut Pasteur, Paris, France

Abstract

Mobile genetic elements (MGEs) drive genetic transfers between bacteria using mechanisms that require a physical interaction with the cellular envelope. In ESKAPE species, the high-priority multidrug-resistant nosocomial pathogens, the membrane is surrounded by chemically diverse capsules, the first point of contact of virions and conjugative pili. While the capsule can be a barrier to MGEs, it also evolves rapidly by horizontal gene transfer. Here, we aim at understanding this apparent contradiction by studying the co-variation between the repertoire of capsule genes and MGEs in ~4000 genomes of Klebsiella pneumoniae and experimentally in natural isolates. We show that capsules drive phage-mediated gene flow between closely related serotypes. Such serotype-specific phage predation also explains the frequent inactivation of capsule genes, observed in more than 3% of the genomes. Inactivation is strongly epistatic, recapitulating the capsule biosynthetic pathway. We show that conjugative plasmids are acquired at higher rates in isolates lacking a functional capsular locus and in capsule mutants. This suggests that capsule inactivation by phage pressure facilitates its subsequent re-acquisition by conjugation. Accordingly, capsule re-acquisition leaves long recombination tracts around the capsular locus. The loss and re-gain process re-wires gene flow towards other lineages whenever it leads to serotype changes. Such changes happen preferentially between chemically related serotypes, hinting that the fitness of serotype-swapped strains depends on the host background. These results reveal trade-offs between the evolution of virulence and multidrug resistance and caution that some alternatives to antibiotics by selecting for capsule inactivation may facilitate the acquisition of antibiotic-resistance genes.

SYMP1-4

Ecological drivers of CRISPR immune systems

Wei Xiao[1], Philip LF Johnson[1], Jake Weissman[2]
[1]University of Maryland, College Park, MD, USA. [2]University of Southern California, Los Angeles, CA, USA

Abstract

CRISPR-Cas is the only known adaptive immune system of prokaryotes. It is a powerful defense system against mobile genetic elements such as bacteriophages. While CRISPR-Cas systems can be found throughout the prokaryotic tree of life, they are distributed unevenly across taxa and environments. Since adaptive immunity is more useful in environments where pathogens persist or reoccur, ecological drivers of uneven CRISPR distribution likely involve the density or diversity of the host/pathogen community. We directly tested hypotheses connecting CRISPR incidence with prokaryotic density/diversity by analyzing 16S rRNA and metagenomic data from publicly available environmental sequencing projects. In terms of density, we found that CRISPR systems are significantly favored in lower abundance (less dense) taxa and disfavored in higher abundance taxa, at least in saltwater environments. When we extended this work to compare taxonomic diversity between samples, we found CRISPR system incidence significantly correlated with diversity in human oral environments. Together, these observations confirm that, at least in certain types of environments, the prokaryotic ecological context indeed plays a key role in selecting for CRISPR immunity, potentially due to correlations with pathogen dynamics.

SYMP1-5

Molecular interplay between CRISPR-Cas and viral satellites in *Staphylococcus aureus*

Dalton V Banh, Luciano A Marraffini
The Rockefeller University, New York, NY, USA

Abstract

The flow of mobile genetic elements (MGEs) between cells is a major driver for the evolution of bacteria, especially *Staphylococcus aureus*, which is a leading cause of life-threatening infections. MGEs like *S. aureus* Pathogenicity Islands (SaPIs) are sources for antibiotic resistance and virulence determinants. In addition to endowing their hosts with beneficial traits, SaPIs parasitize bacteriophages for their own selfish dissemination and interfere with the phage life cycle, providing bacteria with a form of population-level phage defense. SaPIs are ubiquitous—with clinical *S. aureus* isolates carrying at least one (and often multiple) elements. Although gene transfer can often be advantageous, such as in the case of SaPIs, invading MGEs can also impose major fitness costs. To maintain genome integrity, many bacteria use an adaptive immune system called CRISPR-Cas to defend themselves against foreign genetic elements, including plasmids and phages. Comparative genomic analyses, however, reveal that CRISPR-Cas is extraordinarily rare in staphylococci. Although the evolutionary drivers are likely multifactorial, we propose that the scarcity of staphylococcal CRISPR-Cas systems is in part due to CRISPR immunity restricting the mobility of SaPIs. To test this hypothesis, we employed genetic approaches to elucidate the molecular mechanisms by which CRISPR-Cas systems interact with invading SaPI genomes, as well as during their mobilization by phages. By utilizing an animal model, we also characterized how CRISPR-Cas influences the dynamics of SaPI transfer between bacteria *in vivo*. Our work reveals unexpected consequences for CRISPR immunity and highlights potential evolutionary tradeoffs between CRISPR-Cas and horizontal gene transfer.

SYMP1-6

Large scale genome reconstructions from human gut metagenomes reveal weak coevolution between phages and their bacterial hosts

Andrea Silverj, Moreno Zolfo, Fabio Cumbo, Kun D. Huang, Federica Pinto, Omar Rota-Stabelli, Nicola Segata
University of Trento, Trento, Autonomous Province of Trento, Italy

Abstract

Bacteriophages, or simply phages, are viruses that infect bacteria. They are the most abundant biological entities in the human gut, but their diversity, evolution and ecological relationships with their bacterial hosts are still underexplored. In particular, it is not clear if there is a correspondence between the phylogenies of phages and their hosts.
To increase the number of phages that can be phylogenetically modelled and to test coevolution with their hosts, here we mapped all known viral reference genomes from RefSeq against a large set of previously assembled and binned contigs from more than 9,000 human gut metagenomic samples. A filtering strategy was devised to identify viral sequences from the unbinned fraction of the contigs, in order to select the phages which are not integrated in bacterial genomes. By using this approach, we retrieved more than 5,000 genomes belonging to nearly 70 different phage species and reconstructed molecular phylogenies of the newly characterised genomes. We then used the annotated information available on the host specificity, together with host predictions generated by CRISPR spacer matches, to link the retrieved phages to their hosts, building cophylogenies for each of the phage-host pairs that co-occurred in the same samples.
Our results show that only a few phages follow their corresponding host phylogenies, suggesting that in most cases there’s no correspondence between the evolutionary histories of bacteria and their associated known phages. Our findings indicate that phage-bacteria relations and eco-evolution in the human gut are more complex and intricate than previously thought.

SYMP1-7

Yeast defense against killer virus toxin

Ilya Andreev, Meru Sadhu
National Institutes of Health, Bethesda, Maryland, USA

Abstract

Killer viruses in fungi do not lyse infected cells; instead, they continually reside in the host cell cytoplasm, propagating through cell divisions and meioses. Killer viruses encode proteinaceous killer toxins that are secreted from the host cell to inhibit the growth of uninfected cells nearby, providing a local fitness advantage to infected cells and their resident viruses. Although the modes of toxicity of many killer toxins have been well studied, little is known about whether non-killer yeast have defense mechanisms against these toxins. We observed a wide spectrum of resistance to the killer toxin K28 among 16 diverse natural isolates of *Saccharomyces cerevisiae*. Using QTL mapping between a resistant and sensitive strain, we discovered that YAR028W, a previously uncharacterized gene, plays a vital role in resistance to K28 toxin. YAR028W is a member of the enigmatic DUP240 gene family, which is recently arisen in the Saccharomycetaceae fungi. DUP240 genes have no previously ascribed function and are highly variable between and within species, suggesting the entire gene family may be engaged in an evolutionary arms race with killer toxins.

SYMP2-1

Unpredictable outcomes: How mitonuclear variation modulates stress

Nick Lane, Enrique Rodriguez, Finley Grover Thomas, Hugh Carter, Vassilios Kotiadis, Florencia Camus
Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

Abstract

Incompatibilities between mitochondrial and nuclear genes can perturb cell respiration, biosynthesis, signalling and gene expression. Apparently trivial incompatibilities, with limited phenotypic effects under normal conditions, can be altered by stress in unpredictable ways. We have explored responses to redox stress and diet in *Drosophila melanogaster* lines with coevolved (WT) or mildly mismatched mitochondrial DNA (COX, BAR) against an isogenic nuclear background. As expected, high-protein diet increased fertility and decreased lifespan in all lines, but there were big differences between lines, with BAR females being more fertile, more active and longer lived than other lines. But BAR females were extremely vulnerable to redox stress induced by the glutathione precursor N-acetyl cysteine (NAC). High doses of NAC had limited effect on male lines, or on WT and COX females, but ~80% of BAR females died within 12 days. NAC increased tissue glutathione levels and suppressed normal respiration, especially at complex I (as measured by high-resolution tissue-specific fluorespirometry), most strikingly in BAR females. Intriguingly, H_2O_2 flux remained stable, suggesting that tissue redox state can be maintained through suppression of respiration, to the point of death. I will conclude with some thoughts on how mitonuclear incompatibilities perturb Krebs cycle flux and redox homeostasis, with downstream effects on gene expression and stress response in a changing environment.

SYMP2-2

A lethal genetic incompatibility in mitochondrial Complex I of swordtail fish hybrids

Benjamin M Moran[1,2], Cheyenne Y Payne[1,2], Daniel L Powell[1,2], Russell Corbett-Detig[3], Shreya M Banerjee[1], Quinn Langdon[1], Erik Iverson[4], Justin Havird[4], Manfred Schartl[5,6], Molly Schumer[1,7]
[1]Stanford University, Stanford, CA, USA. [2]Centro de Investigaciones Científicas de las Huastecas "Aguazarca", A.C., Calnali, Hidalgo, Mexico. [3]UC Santa Cruz, Santa Cruz, CA, USA. [4]University of Texas at Austin, Austin, TX, USA. [5]University of Würzburg, Würzburg, Bayern, Germany. [6]The Xiphophorus Stock Center, Texas State University, San Marcos, TX, USA. [7]Howard Huges Medical Institute, Stanford, CA, USA

Abstract

The genetic interactions that cause reproductive barriers are fundamental to the formation of new species and can help us understand the diversification of life on Earth. These genetic interactions often take the form of “hybrid incompatibilities,” where genes derived from two different species no longer interact properly. Theory predicts that hybrid incompatibilities involving more than two genes should be common between diverging species, but empirical evidence has lagged this prediction. Here, we describe a mitonuclear incompatibility involving two nuclear genes within respiratory Complex I and the mitochondrial genome of naturally hybridizing swordtail fishes. We show that individuals with homozygous mismatched protein combinations fail to complete embryonic development, and explore the mechanisms by which heterozygous individuals avoid these fitness consequences. Using in silico structure models, we localize the protein-protein interactions that likely underlie lethality, and document a history of accelerated evolution and introgression in the genes involved. Together, this work provides the most comprehensive portrait to date of the architecture of a complex lethal incompatibility found in naturally hybridizing species and the evolutionary mechanisms that led to its formation.

SYMP2-3

A case of reverse Lansing effect: mitochondria of older mothers in their daughters' brains

Cora E Anderson, Millicent N Ekwudo, Morad Malek, Lev Y Yampolsky
East Tennessee State University, Johnson City, TN, USA

Abstract

Offspring of older mothers have been shown to have lower life expectancy in a variety of organisms (Lansing effect). Accumulation of heritable and non-heritable proteome- or membrane-related changes in germ-line mitochondria have been often hypothesized as a causative factor of Lansing effect. Here we report a genotype of Daphnia magna in which a reversed maternal effect is observed: daughters of older mothers have an extended lifespan and produce significantly more male offspring than daughters of younger mothers. This effect is half-eliminated by a single generation of maternal age reversal treatment. We demonstrate that in this genotype daughters of older mothers inherited mitochondria that show higher membrane potential (measured by rhodamine123 assay) in the brain tissues, but not in muscles (where there is effect of maternal age) and not in excretory organs (where the effect is reversed). We discuss possible germline mitochondria rejuvenation mechanism that might allow for the observed unusual maternal age effect and hypothesize that a broader screen of Daphnia genotypes may reveal a more widespread occurrence of the observed effects.

SYMP2-4

Experimental evolution yields key mechanistic insights on the population dynamics of selfish mitochondrial genomes

Bryan L Gitschlag, Claudia V Pereira, Ann T Tate, Maulik R Patel
Vanderbilt University, Nashville, TN, USA

Abstract

An exceptionally successful symbiosis—between a bacterial ancestor of mitochondria and a host—facilitated the eukaryotic revolution, a key evolutionary transition. Mitochondria lie at the center of many vital physiological processes; likewise, mitochondrial defects underlie a range of deleterious phenotypes and hereditary diseases. As endosymbionts, mitochondria harbor their own genomes, which exist at high copy number and undergo relaxed replication that occurs throughout the cell cycle. Accordingly, mitochondrial DNA (mtDNA) mutations do not conform to Mendelian patterns of inheritance, making mitochondrial genetics a unique challenge with broad evolutionary and biomedical implications. Due to relaxed replication, selection can favor “selfish” mutations that promote mutant genome proliferation at the expense of host fitness. To gain fundamental insights into mitochondrial genetics, I have established mitochondrial mutant strains of *Caenorhabditis elegans* as experimentally tractable models for investigating selfish genome dynamics. By designing transmission-bias and organismal competition experiments to isolate and quantitatively measure changes in mutant mtDNA frequency at different levels of selection—within and between hosts, respectively—I have developed an approach to discover mechanisms underlying both the proliferation and host fitness consequences of selfish mtDNA. Focusing on the mutant genome *uaDf5*, I found that selfish mtDNA can proliferate by exploiting key stress-resistance mechanisms, namely the mitochondrial unfolded protein response and the metabolic regulator FoxO/DAF-16. I am now conducting experiments to measure within- and between-host selection forces across a wider collection of mtDNA mutant variants. I expect this work will uncover general themes and important differences in the inheritance patterns of selfish mitochondrial genomes.

SYMP2-5

mtDNA heteroplasmy analysis of green sea turtles reveals population-specific signatures and sheds light on potential heteroplasmy evolutionary processes and advantages

Yaron Tikochinski[1], Carlos Carreras[2], Gili Tikochinski[1], Sibelle Vilaça[3]

[1]Ruppin Academic Center, Michmoret, Israel. [2]Universitat de Barcelona, Barcelona, Spain. [3]Università di Ferrara, Ferrara, Italy

Abstract

Population genetic analysis of green turtles have been challenging, especially when mtDNA is the analysis tool. In the Mediterranean population, it is almost completely impossible to detect SNPs. Recently we have developed a new haplotyping method based on the D-loop short tandem repeats (mtSTRs), We used Illumina high throughput sequencing to study the population of mtDNA molecules on ~200 green sea turtles from Israel. We found heteroplasmy levels in the mtSTR ranging from 5% and up to 45% in all of the samples with a mean of more than 10 haplotypes per individual. Intra nest variation was an indication of the changes in haplotype frequencies from maternal haplotype composition to the offspring thus showing the effect of the cellular bottleneck. To evaluate the implications of heteroplasmy in mtDNA evolution we constructed a Python-based model simulating the accumulation of mutations and establishment of new haplotypes in a homogeneous population of green turtles. Our model of mtDNA inheritance indicated that heteroplasmy favored the increase of population diversity through time and buffered against population bottlenecks. Individuals with recent haplotypes showed higher levels of heteroplasmy than the individuals with ancient haplotypes, suggesting a potential advantage of maintaining established copies when new mutations arise.

SYMP2-6

Mitochondrial mutations in *Caenorhabditis elegans* show signatures of oxidative damage and an AT-bias

Gus Waneka[1], Joshua M. Svendsen[1], Justin C. Havird[2], Daniel B. Sloan[1]
[1]Colorado State University, Fort Collins, CO, USA. [2]University of Texas at Austin, Austin, TX, USA

Abstract

Rapid mutation rates are typical of mitochondrial genomes (mtDNAs) in animals, but it is not clear why. The difficulty of obtaining measurements of mtDNA mutation that are not biased by natural selection has stymied efforts to distinguish between competing hypotheses about the causes of high mtDNA mutation rates. Several studies which have measured mtDNA mutations in nematodes have yielded small datasets with conflicting conclusions about the relative abundance of different substitution classes (i.e. the mutation spectrum). We therefore leveraged Duplex Sequencing, a high-fidelity DNA sequencing technique, to characterize *de novo* mtDNA mutations in *Caenorhabditis elegans.* This approach detected nearly an order of magnitude more mtDNA mutations than documented in any previous nematode mutation study. Despite an existing AT bias in the *C. elegans* mtDNA (75.6% AT), we found that a significant majority of mutations increase genomic AT content. Compared to some prior studies in nematodes and other animals, the mutation spectrum reported here contains an abundance of CGAT transversions, supporting the hypothesis that oxidative damage may be a driver of mtDNA mutations in nematodes. Further, we found an excess of GT and CT changes on the coding DNA strand relative to the template strand, consistent with increased exposure to oxidative damage. Analysis of the distribution of mutations across the mtDNA revealed significant variation among protein-coding genes and as well as among neighboring nucleotides. This high-resolution view of mitochondrial mutations in *C. elegans* highlights the value of this system for understanding relationships among oxidative damage, replication error, and mtDNA mutation.

SYMP2-7

Dynamic evolution of the MutS family of DNA repair proteins in animals

Viraj Muthye, Dennis V Lavrov
Iowa State University, Ames, Iowa, USA

Abstract

The DNA Mismatch Repair System (MMR) corrects mismatch errors that arise during DNA replication. Members of the MutS protein family carry out the identification of mismatches. MutS proteins are found in prokaryotes, eukaryotes, and viruses. Six MutS homologs are present in yeast (MSH1-6), five of which function in the nuclear genome (MSH2-6). MSH1 is involved in DNA repair in the mitochondrial genome of yeast and is believed to be lost in animals. This potentially suggests a loss of MMR in animal mitochondria, explaining the higher rates of mitochondrial sequence evolution in animals. Interestingly, octocorals (phylum Cnidaria) contain a gene encoding a MutS-like protein (mtMutS) in their mitochondrial genome. Octocorals show some of the lowest rates of mitochondrial sequence evolution in animals, suggesting that mtMutS is involved in MMR. Here, we analyzed the evolution of the MutS protein family in animals. Surprisingly, we identified MSH1 in several animal groups. We saw that animal groups that lacked either of the two mitochondrial MutS proteins (mtMutS/MSH1) had higher rates of mitochondrial sequence evolution than groups with either one. Our analysis confirmed a previous observation that mtMutS was acquired by octocoral mitochondrial via horizontal gene transfer from a virus. We identified MSH2 and MSH6 in nearly all animal species we analyzed. However, MSH3, MSH4, and MSH5 were missing in multiple animals. Overall, our analysis reveals the dynamic evolution of the MSH family in animals, with multiple losses of MSH1, MSH3, some losses of MSH4 and MSH5, and a gain of octocoral mtMutS.

SYMP4-1

Gene network simulations provide testable predictions for the molecular domestication syndrome

Ewen BURBAN[1], Maud Tenaillon[2], Arnaud Le Rouzic[3]
[1]ECOBIO, Rennes1, Bretagne, France. [2]GQE le Moulon, Gif-sur-Yvette, Île-de-France, France. [3]EGCE, Gif-sur-Yvette, Île-de-France, France

Abstract

The domestication of plant species lead to repeatable morphological evolution, often referred to as the phenotypic domestication syndrome. Domestication is also associated with important genomic changes, such as the loss of genetic diversity and modifications of gene expression patterns. Here, we explored theoretically the effect of domestication at the genomic level by characterizing the impact of a domestication-like scenario on gene regulatory networks. We ran population genetics simulations in which individuals were featured by their genotype (an interaction matrix encoding a gene regulatory network) and their gene expressions, representing the phenotypic level. Our domestication scenario included a population bottleneck and a selection switch mimicking canalizing selection, i.e. evolution towards more stable expression to parallel enhanced environmental stability in man-made habitat. We showed that domestication profoundly alters genetic architectures. Based on the well-documented example of the maize domestication, our simulations predicted (i) a drop in neutral allelic diversity, (ii) a change in gene expression variance that depended upon the domestication scenario, (iii) transient maladaptive plasticity, (iv) a deep rewiring of the gene regulatory networks, with a trend towards gain of regulatory interactions between genes, and (v) a global increase in the genetic correlations among gene expressions, with a loss of modularity in the resulting coexpression patterns and in the underlying networks. Here, we provide empirically testable predictions on the differences of genetic architectures between wild and domesticated forms. The characterization of such systematic evolutionary changes in the genetic architecture of traits contributes to define a molecular domestication syndrome.

SYMP4-2

phyloConverge: a likelihood-based approach to identify genomic regions underlying phenotypic convergence at nucleotide resolution

Elysia Saputra[1,2], Nathan Clark[3,2,4], Maria Chikina[2]

[1]Joint Carnegie Mellon University-University of Pittsburgh Program in Computational Biology, Pittsburgh, PA, USA. [2]University of Pittsburgh, Pittsburgh, PA, USA. [3]University of Utah, Salt Lake City, UT, USA. [4]Pittsburgh Center for Evolutionary Biology and Medicine, Pittsburgh, PA, USA

Abstract

How nucleotide-level alterations underlie organism-level phenotype changes is a critical question in molecular evolution. Convergent phenotypes, whereby multiple species independently develop similar characteristics, are useful for inferring such genotype-phenotype associations. Several methods have been developed to associate convergent phenotypes with genetic elements. However, for non-coding sequences, the notion of a "genetic element" is not well-defined but is a prerequisite for all existing methods.

To address this challenge, we propose a new method, phyloConverge, which is built on the well-known *phyloP* framework to test for convergent evolutionary rate shifts at high resolution. PhyloConverge uses a maximum likelihood phylogenetic model and phylogeny-aware bias correction to scan entire multiple sequence alignments and compute nucleotide-resolution conservation scores for the convergence hypothesis.

We benchmark phyloConverge using a dataset previously analysed with an alternative method to identify conserved non-coding elements (CNEs) convergently diverged in 4 independent subterranean mammal lineages. Because these lineages have degenerate eyes, we expect to observe the loss of constraint on eye-related regions. We computed rate acceleration scores for 491,576 CNEs and assessed how top-scoring CNEs overlapped eye-related open chromatin regions. PhyloConverge produced stronger enrichment across eye-related datasets than competing methods. Gene ontology and pathway analyses also showed enrichment for relevant functions.

Finally, we computed high-resolution acceleration scores across nucleotides within diverged CNEs. We discovered genome-wide divergence of 179 transcription factor binding site motifs, with top-ranking motifs previously implicated in ocular development. PhyloConverge provides a means to identify new regulatory regions underlying convergent traits at the nucleotide level, without defining elements *a priori*.

SYMP4-3

Position effect and its externality

Qian Gui, Shuyun Deng, Waifang Cao, Xiaoshu Chen
Sun Yat-sen University, Guangzhou, China

Abstract

Position effect, the influence of the chromosomal location of a gene on its activity, is a fundamental property of the genome. By placing a GFP gene cassette at 482 different locations across all chromosomes in budding yeast, we quantified the position effects on protein expression level and noise at the genomic scale. DNA replication timing, 3D chromosomal conformation, and several histone modifications are major covariates of position effects. Essential genes are enriched in genomic regions with inherently low expression noise, supporting the hypothesis that chromosomal clustering of essential genes results from selection against their expressional stochasticity. Our results suggest that position effects have shaped the evolution of chromosome organization.
The other side of the coin, how position effects affect the local chromosomal environment, has remained largely unexplored, as have the mechanism and phenotypic consequences of this “externality” of the position effect. For this reasonwe further examined the transcriptome profiles of the above constructed strains, we found that in genomic regions enriched in essential genes, GFP expression tended to be lower, and the genes near the integration site tended to show greater expression reduction. More importantly, we found that changes in the expression of neighboring genes, but not GFP expression, significantly altered the cellular growth rate. Our results consistent with competition for transcriptional resources among neighboring genes and highlight the impact of position effects on the fate of exogenous gene integration and has significant implications for biological engineering and the pathology of viral integration.

SYMP4-4

Selection on accessible chromatin regions in *Capsella grandiflora*

Robert Horvath[1], Emily B Josephs[2], Edouard Pesquet[1], John R Stinchcombe[3], Stephen I Wright[3], Douglas Scofield[4], Tanja Slotte[1]

[1]Stockholm University, Stockholm, Sweden. [2]Michigan State University, Lansing, Michigan, USA. [3]University of Toronto, Toronto, Canada. [4]Uppsala University, Uppsala, Sweden

Abstract

Accurate estimates of genome-wide rates and fitness effects of new mutations are essential for an improved understanding of molecular evolutionary processes. Although eukaryotic genomes generally contain a large non-coding fraction, functional non-coding regions and fitness effects of mutations in such regions are still incompletely characterized. A particularly promising approach to characterize functional non-coding regions relies on identifying accessible chromatin regions (ACRs) tightly associated with regulatory DNA. Here, we applied this approach to identify and estimate selection on ACRs in *Capsella grandiflora*, a crucifer species that is ideal for population genomic quantification of selection due to its favourable population demography. We describe a population-wide ACR distribution based on ATAC-seq data for leaf samples of 16 individuals from a natural population. We then use population genomic methods to estimate fitness effects and proportions of positively selected fixations (α) in ACRs. We find that intergenic ACRs harbor a considerable fraction of weakly deleterious new mutations, as well as a significantly higher proportion of strongly deleterious mutations than comparable inaccessible intergenic regions. ACRs are further enriched for expression quantitative trait loci (eQTL) and depleted of transposable element (TE) insertions, as expected if intergenic ACRs are under selection because they harbor regulatory regions. Our results are important for an improved understanding of selection on non-coding regions and the role of nearly neutral mutations for evolutionary processes in outcrossing Brassicaceae species.

SYMP4-5

The importance of cis-regulatory variation in regulatory and gene-expression divergence of stickleback ecotypes.

Stanley Neufeld[1], Jukka-Pekka Verta[1,2], Muhua Wang[1], Jonas Schwickert[1], Malina John[1], Michael Ritter[1], Domenico Scionti[1], Sebastian Kick[1], Melanie Kirch[1], Elena Avdievich[1], Vrinda Venu[1,3], Felicity C Jones[1]
[1]Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Baden-Württemberg, Germany. [2]University of Helsinki, Helsinki, Finland. [3]Los Alamos National Laboratory, Los Álamos, New Mexico, USA

Abstract

The adaptation of natural populations to changing environments is often driven by numerous genetic loci predominantly found in non-coding regions with likely gene regulatory roles. Using adaptively diverging marine and freshwater stickleback fish ecotypes as a model, we quantified the degree to which cis- and trans- acting factors underlie expression divergence with allele-specific gene expression assays. Finding a major role for cis-regulatory differences compared to trans-regulatory differences, we next performed comparative epigenomics, chromatin profiling, transcriptomics and genetics to identify thousands of cis-regulatory elements with divergent epigenomic profiles between the ecotypes. Divergent elements are enriched at the promoter and 5'UTR of genes, are proximal to genes showing differential expression, and vary across tissues, with the liver showing considerably higher regulatory divergence than kidney or gills. Allele-specific analyses in F1 hybrids reveals that divergence in chromatin accessibility is itself mostly cis-regulated and these elements show molecular signatures of natural selection. Additionally, forward genetic mapping of chromatin variation identifies QTLs located on the same chromosome as each of the divergent chromatin peaks, and finds little evidence for trans-acting factors controlling differences in chromatin accessibility and regulatory genome function. The high resolution maps of the chromatin and epigenomic landscape in diverging stickleback ecotypes provides functional annotation of regulatory elements within adaptive loci. Combined, our studies show overwhelming evidence for the major role of cis-regulation of the regulatory genome and gene expression in the adaptive divergence of marine and freshwater stickleback in the early stages of speciation.

SYMP4-6

A shared regulatory allele of *Agouti* contributes to parallel evolution of cryptically colored beach mice

Brock Wooldridge[1,2], Andreas F Kautt[1,2], Sade McFadden[1,2], Hopi E Hoekstra[1,2]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]Howard Hughes Medical Institute, Harvard University, Cam, MA, USA

Abstract

Despite the many examples of parallel evolution, in only few cases are both the genetic basis and evolutionary history of repeatedly evolved traits well-understood. The oldfield mouse (*Peromyscus polionotus*) is widespread across the southeastern United States and exhibits considerable coat-color variation: dorsal coats range from dark brown in mainland habitat to near white, which has evolved independently on the white sand beaches of Florida's Gulf and Atlantic coasts as an adaptation to visual predators. Here, we focus on a phenotypically-variable population of oldfield mice, *P. p. albifrons*, which is closely related to dark-colored mainland mice and a likely source population for the pale-colored Gulf Coast beach mice. performing genome-wide association mapping of 15 pigmentation traits in wild *albifrons* mice, we find that pigmentation is associated with a previously undescribed ~2kb intronic region of the vertebrate pigmentation gene *Agouti*. LacZ reporter assays demonstrate that this candidate region drives expression in the epidermis of developing mouse embryos – when pigment patterns are established – supporting a role in regulating *Agouti* expression and color pattern evolution. Furthermore, extended tracts of homozygosity indicate that the light haplotype has experienced recent and strong positive selection in the polymorphic *albifrons* populations. Notably, this same light haplotype is fixed in both Gulf Coast and Atlantic beach mice, although these populations live >3,000km apart. Given their evolutionary history, these results suggest that this newly discovered regulatory allele has been maintained in polymorphic mainland populations and repeatedly spread to both beach mouse lineages, thereby facilitating their rapid and parallel evolution.

SYMP4-7

Positive selection on regulatory regions drives gene expression evolution on recently evolved sex chromosomes in threespine stickleback fish

Daniel E. Shaw, Michael A. White
University of Georgia, Athens, GA, USA

Abstract

Heteromorphic sex chromosomes (X and Y) have evolved independently many times across the tree of life. One key step in the evolution of sex chromosomes is the suppression of meiotic crossovers, which leads to the rapid degeneration of the sex-limited chromosome. Much work has been devoted to understanding how sequence degeneration leads to the loss of ancestral genes on Y chromosomes, but it remains unclear how regulatory regions (promoters and enhancers) evolve after recombination is suppressed. Recent theoretical work has led to a new theory that suggests selection on regulatory regions could lead to lower Y allelic expression and help drive Y degeneration (degeneration by regulatory evolution). The threespine stickleback (*Gasterosteus aculeatus*) has a high-quality Y chromosome assembly, allowing us to explore this hypothesis at the early stages of sex chromosome evolution. We examined sequence divergence of regulatory regions between the Y chromosome in threespine stickleback fish with the ancestral autosome outgroup, ninespine stickleback fish (*Pungitius pungitius*). We found evidence of positive selection on Y-linked promoter regions associated with changes in gene expression. We combined sequence divergence estimates with RNA-seq to find support for degeneration by regulatory evolution. We found that elevated divergence is associated with up-regulated X alleles and down-regulated Y alleles, resulting in the maintenance of ancestral levels of gene dosage. Our findings provide the first evidence that sex-linked regulatory regions undergo positive selection quickly after the suppression of recombination. This result has broad implications for understanding how Y chromosomes degenerate.

SYMP6-1

Scaling biodiversity genomics to predict species' responses to global environmental change

Ann-Marie Waldvogel
Institute of Zoology, University of Cologne, Cologne, Germany

Abstract

Contemporary ecosystems show an alarming loss of biodiversity, reduction of wildlife and fragmentation of habitats. This decline in nature is a direct result of human activity but at the same time also constitutes a direct threat to human well-being in all regions of the world. The accelerating speed of climate change poses the biggest challenge to the development of strategies for a realistic and sustainable ecosystem management. In order to overcome the status of only retrospectively documenting the biodiversity crisis as it is ongoing, there is an urgent need to understand and then ultimately forecast how biodiversity will respond to changing environmental conditions.

Studies to show the importance of integrating genomic data in various frameworks of prediction modelling are accumulating, demonstrating the power and improvement in accuracy of such predictions. Approaches of genomic forecasting are thus most relevant to becoming implemented in conservation genomic studies and eventually inform conservation management. This transition of a currently monitoring-orientated ecosystem management towards prediction-based conservation strategies, will demand a significant scaling in the availability of genomic resources – reference genomes of the critical proportion of biodiversity as well as resequencing data on the population level. Coordinated biodiversity genomics initiatives are required to generate this data fast enough to meet the urgent need of action in face of the accelerating speed of global environmental change.

SYMP6-2

The evolution of seasonal camouflage in white-tailed jackrabbits in response to past and future climates

Mafalda S. Ferreira[1,2,3], Timothy J. Thurman[3], Matthew R. Jones[3], Liliana Farelo[1], Alexander V. Kumar[4,5], Sebastian M.E. Mortimer[3], John R. Demboski[6], L. Scott Mills[4,7], Paulo C. Alves[1,2,4], José Melo-Ferreira[1,2], Jeffrey M. Good[3,4]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Porto, Portugal. [2]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal. [3]Division of Biological Sciences, University of Montana, Missoula, Montana, USA. [4]Wildlife Biology Program, College of Forestry and Conservation, University of Montana, Missoula, Montana, USA. [5]Inventory and Monitoring Division, National Park Service, Fort Collins, Colorado, USA. [6]Zoology Department, Denver Museum of Nature & Science, Denver, Colorado, USA. [7]Office of Research and Creative Scholarship, University of Montana, Missoula, Montana, USA

Abstract

Adaptation from standing genetic variation is a critical component of evolutionary responses to rapid environmental change. However, the difficulty of identifying the genetic basis of fitness-relevant traits in natural populations has limited the direct incorporation of genotype-to-phenotype information into conservation efforts. We studied the evolution of adaptive winter color variation in the white-tailed jackrabbit (*Lepus townsendii*), a North American species undergoing population declines. Using extensive museum records, we show that winter pelage color closely tracks dynamics of snow cover across the range of white-tailed jackrabbits, suggesting that geographic variation for the trait is maintained by strong selection. Using whole genomes of specimens collected during winter, we show that seasonal camouflage variation was primarily determined by additive and epistatic genetic variation at three pigmentation genes. With a phylogenetic analysis of whole genomes from 10 hare species, we further show that color variation is associated with highly divergent alleles at all three genes and link their origin to a mixture of long-term maintenance of ancestral polymorphisms and introgression from another species. Using a probabilistic model of winter coat color and snow cover projections for 2080, we predict that future selective pressures will strongly favor darker winter phenotypes across much of the white-tailed jackrabbit distribution. However, adaptation to future snow cover may be impeded by ongoing population declines that appear to differentially threaten standing adaptive genetic variation for darker winter pelage. Our study illustrates how evolutionary genomics can be used to identify functional genetic variation of critical importance for adaptation to climate change.

SYMP6-3

Parallel reduction in flowering time enabled evolutionary rescue and establishment in a colonizing *Arabidopsis* lineage

Célia Neto[1], Andrea Fulgione[1], Ahmed F. Elfarargi[1], Emmanuel Tergemina[1], Mehmet Göktay[1], Herculano Dinis[2], Nina Döring[1], Pádraic J. Flood[1], Sofia Rodriguez-Pacheco[1], Nora Walden[3], Markus Koch[3], Fabrice Roux[4], Joachim Hermisson[5], Angela M. Hancock[1]

[1]MPIPZ, Cologne, Germany. [2]Parque Natural do Fogo, Santiago, Cape Verde. [3]Ruprecht-Karls-Universität, Heidelberg, Germany. [4]Université de Toulouse, Castanet-Tolosan, France. [5]University of Vienna, Vienna, Austria

Abstract

Populations subject to abrupt environmental change must adapt quickly to avoid extinction. Small populations are especially vulnerable, confronting high extinction risk due to limited genetic variation and low efficiency of selection. Here, we use 335 newly collected and fully sequenced wild population samples from the Cape Verde Islands to investigate the mechanisms of adaptation after a sudden shift to an arid climate. We find that the time to flowering was reduced in parallel across islands, substantially increasing fitness, and that this change was mediated by *de novo* loss of function of two core flowering time genes: *FRI* on one island and *FLC* on the other. We characterize the adaptive scenario by combining inferences from population genetic data, modeling and simulations. Evolutionary reconstructions reveal a case where expansion of the new populations coincided with the emergence and proliferation of these novel variants, consistent with models of rapid adaptation and evolutionary rescue. Further, our results imply that adaptation to a more limited precipitation regime and reduced growing season – a scenario which may be common under global climate change – is predictable.

SYMP6-4

Genomic and population viability analyses suggest high recovery potential for the critically endangered vaquita

Jacqueline A Robinson[1], Christopher C Kyriazis[2], Annabel C Beichman[3], Sergio F Nigenda-Morales[4], Kirk E Lohmueller[2], Lorenzo Rojas-Bracho[5], Barbara L Taylor[6], Robert K Wayne[2], Phillip A Morin[6]

[1]University of California San Francisco, San Francisco, CA, USA. [2]University of California Los Angeles, Los Angeles, CA, USA. [3]University of Washington, Seattle, WA, USA. [4]Advanced Genomics Unit, National Laboratory of Genomics for Biodiversity (Langebio), Center for Research and Advanced Studies (Cinvestav), Irapuato, Guanajuato, Mexico. [5]Comisión Nacional de Áreas Naturales Protegidas/SEMARNAT, Ensenada, BC, Mexico. [6]Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA, USA

Abstract

Anthropogenic pressures are driving species declines across the globe. Consequently, many species are predicted to suffer the genetic impacts of reduced population size, including loss of diversity and inbreeding depression, elevating their risk of extinction. Integrating genomic information, such as the distribution of deleterious or adaptive variation, into population viability models and conservation planning remains a fundamental challenge. We used whole genome sequencing, analysis, and simulations to project population trends over the next 50 years in the most endangered marine mammal, the vaquita (Phocoena sinus). Decades of excess bycatch mortality from gillnet fishing have driven its catastrophic decline, with only ~10 individuals remaining from a population of several thousand. Due to this extreme decline and low genetic diversity in vaquitas, it has been suggested that the species is doomed to extinction by inbreeding depression, and that scarce conservation resources should no longer be devoted to its recovery. However, our analyses show that low diversity is a natural characteristic of the vaquita caused by its historically low abundance, and not a consequence of its recent decline. We corroborate this finding with stochastic, individual based simulations, which similarly suggest that the risk of inbreeding depression in vaquitas is lower due to their reduced burden of recessive deleterious variation. Our simulations suggest recovery is possible if bycatch mortality is immediately halted. However, even modest rates of continuing bycatch mortality result in an appreciable extinction risk. Our results provide hope for vaquita recovery and highlight the applicability of genomic data in conservation management planning.

SYMP6-5

Genomic insights into present local adaptation and future climate change vulnerability of a keystone forest tree species in East Asian

Yupeng Sang, Jing Wang, Zhiqin Long, Guanglei Yang
Sichuan University, Chengdu, Sichuan province, China

Abstract

ABSTRACT

Rapid global climate change is posing numerous threats to Earth's biodiversity, forcing organisms either to rapid adapt and evolve or face maladaptation and possibly extinction. Assessing the adaptive capacity of ecosystem keystone species, like forest tree species, is fundamental for preserving biodiversity and informing where conservation efforts should be most effective. In this study, we first perform a de novo chromosome-level genome assembly of *Populus koreana*, a pioneer and keystone forest tree species in local ecosystems in East Asian. We then characterize the genomic diversity of 230 individuals collected from 24 native populations of *P. koreana*. A combination of genome-wide environmental association studies and whole-genome selection scans was performed to reveal the genomic basis of local adaptation to diverse climates. We identify a set of single nucleotide polymorphisms (SNPs), indels and structural variations (SVs) strongly associated with various environmental variables and under strong natural selection. Finally, we integrate climate-associated genomic variation with environmental modelling and machine learning approaches to investigate spatiotemporal response to future climate change. The gradient forest (GF) and risk of non-adaptedness (RONA) analyses help us to identify populations being most vulnerable to future climate change. To our knowledge, our work is one of the very few studies that associate whole-genome variations (SNPs, indels and SVs) with local environmental variables and predictions of climate change vulnerability across the landscape of an ecologically important forest tree species, which may provide a key reference for other studies working on non-model species.

SYMP6-6

Predicting bleaching response from genomes in a Great Barrier Reef coral

Zachary L Fuller[1], Veronique Mocellin[2], Joseph Pickrell[3], Peter Andolfatto[1], Mikhail Matz[4], Line Bay[2], Molly Przeworski[1]

[1]Columbia University, New York, NY, USA. [2]Australian Institute of Marine Science, Townsville, Australia. [3]Gencove Inc., New York, NY, USA. [4]University of Texas, Austin, TX, USA

Abstract

Although reef-building corals are rapidly declining worldwide, there is considerable variation in bleaching response and heat tolerance within populations, which is in part heritable. To map the genetic basis of this variation and develop individual predictors of bleaching in the wild, we conducted a genome-wide association study (GWAS) of bleaching in the coral Acropora millepora from the Great Barrier Reef. Using over 200 sequenced colonies and a chromosome-scale genome assembly, we first investigated population structure and demographic history, finding high levels of gene flow across hundreds of kilometers. We show that we can reliably impute genotypes in low-coverage sequencing data to obtain millions of high confidence single nucleotide polymorphism (SNP) calls. Testing more than 6.8 million SNPs for association with bleaching, we find strong evidence for a polygenic basis for bleaching response. Moreover, we show a polygenic score constructed from these GWAS estimates to be a significant predictor of bleaching and show the proportion of phenotypic variance that can be explained by environmental, symbiotic and genetic factors. Finally, we detect a strong signal of balancing selection at a heat-shock co-chaperone. These results have implications for understanding how corals may respond to increased warming in the future and thus set the stage for the use of genomic-based prediction in coral conservation strategies.

SYMP6-7

Validating metrics of "genetic offset" and "genomic vulnerability" of populations to climate change

Katie E Lotterhos
Northeastern University, Nahant, MA, USA

Abstract

Local adaptation and adaptive potential represent high-priority parameters for incorporation into climate change projections, but are omitted from projections more often than other biological mechanisms. To address this gap, several metrics have recently been proposed that claim to estimate "genomic vulnerability" or "genetic offset" of populations to climate change. Here, I review these estimates and highlight pitfalls in the way these metrics are being applied without rigorous validation. I will then highlight shortcomings in the few empirical attempts at validation that have been made. For example, some have considered a negative association between a measure of "genomic vulnerability" and population size as a validation of the metric. However, I will show that this is not necessarily the case. Using truth-known simulations, I will show how a metric of "genomic vulnerability" (based on the machine learning algorithm gradient forests) can be associated with population size because of neutral genetic drift and not because of selection. These results highlight the importance of understanding how neutral processes affect offset/vulnerability metrics. Finally, I will suggest a framework for validating offset/vulnerability metrics based on best practices in training and testing algorithms from data science.

SYMP8-1

Model selection on empirical data using deep learning

Sebastian Burgstaller-Muehlbacher[1], Heiko A Schmidt[1], Tamara Drucks[1], Stephen Crotty[2], Arndt von Haeseler[1]

[1]Center for Integrative Bioinformatics Vienna, Vienna, Austria. [2]University of Adelaide, Adelaide, Australia

Abstract

Selecting the correct model of sequence evolution (mSE) for a multiple sequence alignment (MSA) constitutes the first step of tree reconstruction. State of the art approaches for inferring nucleotide models mostly apply maximum likelihood (ML) methods.

Here, we demonstrate that neural networks can infer the correct mSE including the shape parameter α of the Γ-distribution. A Residual Neural Network (Resnet) was trained with the six most frequently used mSE (JC, K2P, F81, HKY, TN93 and GTR), whereas a Long Short-Term Memory (LSTM) network with attention was trained to determine α.

Our results show that the CNN correctly identifies the mSE in a range of 51.94% to 100%, depending on the true mSE. Thus, it is comparable to results of IQ-Tree. Similar accuracies were obtained for α. However, the trained networks need substantially less computing time (up to a 60x speedup), depending on the size of the MSA.

We demonstrate, for the first time, that neural networks can be used to identify the correct mSE as well as rate heterogeneity of an MSA. Furthermore, we intend to generalize our approach so all mSE of relevance in phylogenetics can be inferred using neural networks, conferring a substantial reduction in computational requirements to model selection.

SYMP8-2

A LASSO-based approach to sample sites for phylogenetic tree search

Noa Ecker[1], Dana Azouri[1], Ben Bettisworth[2,3], Alexandros Stamatakis[2,3], Yishay Mansour[1], Itay Mayrose[1], Tal Pupko[1]

[1]Tel Aviv University, Tel Aviv, Israel. [2]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany. [3]Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract

In recent years, full-genome sequences have become increasingly available and as a result many modern phylogenetic analyses are based on relatively large sequences, often with over 100,000 sites. Phylogenetic reconstruction of large-scale alignments is challenging for maximum-likelihood based phylogenetic inference programs and usually requires using a powerful computer cluster. Current tools for alignment trimming prior to phylogenetic analysis do not promise a significant reduction in the alignment size and are claimed to have a negative effect on the accuracy of the obtained tree. Here, we propose an artificial intelligence approach, which provides a subset of sites and a formula by which one can compute the log-likelihood of the entire data based on this subset. Our approach is based on training a regularized regression model that optimizes the log-likelihood prediction accuracy while putting a constraint on the number of sites used for the approximation. We tested our approach during SPR searches we performed on 55 alignments sampled from the Selectome database and showed that using only around 10% of the alignment sites well approximated log-likelihood values during the search. For the vast majority of the analyzed alignments, using our site-sampling approximation did not result in an inferior tree topology, yet it substantially reduced running times.

SYMP8-3

Machine learning inference of enhancer conservation across mammalian genomes leads to new genotype-phenotype associations

Irene Kaplow[1], Daniel Schaffer[1], BaDoi Phan[1], Alyssa Lawler[1], Kathleen Foley[2], Wynn Meyer[2], Andreas R Pfenning[1]

[1]Carnegie Mellon University, Department of Computational Biology, Pittsburgh, PA, USA. [2]Lehigh University, Department of Biological Sciences, Bethlemem, PA, USA

Abstract

Rapid advances in the genome sequencing have provided a comprehensive view of cross-species conservation across small segments of nucleotides or even at individual nucleotides. These conservation measures have proven invaluable to associate phenotypic variation, both within and across species, to variation in genotype at protein-coding genes or very highly-conserved enhancers. However, these approaches have had limited success when applied to the vast majority of enhancers where the conservation level of individual nucleotides are low even when enhancer function is conserved. To overcome this limitation, we developed the TACIT-ML (Tissue Aware Conservation Inference Through Machine Learning) approach, in which convolutional neural network models learn the regulatory code connecting genome sequence to tissue-specific open chromatin, allowing us to accurately predict cases where differences in genotype are associated with differences in tissue-specific open chromatin at enhancer regions. We apply this technique to identify dozens of new associations between genetic variation in orthologous motor cortex and liver enhancers across 222 Boreoeuthrian mammals to differences in species' brain size, diet, and longevity. One of the longevity-associated liver enhancers contains a single nucleotide polymorphism in the human population that affects TDG, a gene that regulates DNA repair and methylation in an age-dependent manner. We also show that TACIT-ML substantially improves the interpretation of human genome-wide association studies for a number of complex neurological phenotypes including schizophrenia, addiction, and sleep traits.

SYMP8-4

Evolutionary Sparse Learning for Phylogenomics and Phylogenetics

Sudhir Kumar, Sudip Sharma
Temple University, Philadelphia, PA, USA

Abstract

I will introduce supervised machine learning with sparsity constraints, evolutionary sparse learning (ESL), for molecular phylogenetics and evolutionary analysis. In ESL, we view genomic loci—such as genes, proteins, genomic segments, and positions—as evolutionary model parameters. ESL uses the Least Absolute Shrinkage and Selection Operator (LASSO) to build models that retain only the most important parameters (genomic loci) associated with a phylogenetic hypothesis. ESL is distinct from traditional phylogenetics in which model parameters are branch lengths, rates of substitutions between nucleotides, and evolutionary rates among positions, all of which are estimated during phylogenetic inference. Instead, ESL directly assesses the concordance of patterns of variation at genomic loci in multiple sequence alignment to the hypothesis considered. I will also introduce positional, locus, sequence, and hypothesis sparsity scores for molecular evolutionary analysis and show how to use ESL through example applications. ESL provides a natural way to combine heterogeneous molecular data for the same set of organisms and directly incorporates biological and functional annotations during model building. Ultimately, the ESL framework complements traditional phylogenetic approaches and will encourage a new class of computational methods for molecular evolution and phylogenomics. It can democratize big data analytics via shortened analysis time and a relatively small memory footprint compared to maximum likelihood analyses.

SYMP8-5

Generative neural networks in population genetics

Burak Yelmen[1,2], Aurélien Decelle[1,3], Linda Ongaro[2], Davide Marnetto[2], Corentin Tallec[1], Francesco Montinaro[2], Cyril Furtlehner[1], Luca Pagani[2,4], Flora Jay[1]

[1]Paris-Saclay University, CNRS, INRIA, LISN, Orsay, France. [2]Institute of Genomics, University of Tartu, Tartu, Estonia. [3]Departamento de Física Téorica I, Universidad Complutense, Madrid, Spain. [4]APE Lab, Department of Biology, University of Padova, Padova, Italy

Abstract

In the era of next generation sequencing, population genetic datasets keep increasing in size and it is now common to observe millions of genomic markers sequenced for hundreds or thousands of individuals. Yet, extracting information from these genomic datasets can be complex due to their size and sometimes impossible due to privacy rules that govern several human genome databases.

Recently, deep learning approaches have been introduced at different levels of populations genetics, mostly for evolutionary parameter inference, but also for better characterizing biological processes, visualizing population genetics data or generating them. Our lab is working on demographic and selection inference based on exchangeable neural networks and ABC, and developing *dnadna* a package to easily design and share neural networks for population genetic studies.

In this talk, I will focus on genomic data generation. We leveraged two type of generative neural networks, Generative Adversarial networks (GANs) and Restricted Boltzmann Machines (RBMs) to learn the high dimensional distributions of real genomic datasets. First, these networks capture the hidden structure of population genetic datasets, which can be visualized (as a non-linear alternative to PCA). Second, they enable (unlimited) generation of artificial genomic regions. We showed that these artificial genomes have summary statistics similar to the real ones, and that, given the one-to-one correspondence between artificial and real SNPs, artificial and real data can be combined for downstream analyses, such as imputation or selection scans. Finally, we investigated statistics that can be used both for controlling privacy loss and for better calibration of generative neural networks.

Because artificial genomes retain important characteristics of the real genomes without copying them, they could serve as their surrogates and become valuable assets in future genetic studies by limiting privacy issues associated with genome donors.

SYMP8-6

Ancestry-related assortative mating and sex bias driven by social stratification in admixing American populations

Alex Mas-Sandoval[1], Andres Ruiz-Linares[2], Sara Mathieson[3], Matteo Fumagalli[1]

[1]Department of Life Sciences, Silwood Park campus, Imperial College London, Ascot, United Kingdom.

[2]UMR 7268 ADES, CNRS, Aix-Marseille Université, EFS, Faculté de médecine Timone, Marseille, France.

[3]Department of Computer Science, Haverford College, Haverford, PA, USA

Abstract

Culture and socioeconomic differences stratify human societies and shape the genetic structure beyond geography. However, despite evidence of population stratification in most human populations, demographic models often assume panmixia within geographical demes. In a scenario of sociocultural stratification, the permeability across strata regulates mating and configures the population structure. In admixing populations, such as most from the Americas, sociocultural stratification is correlated with genetic ancestry and is frequently asymmetric between women and men.

To comprehend the complexity of admixture events, we derive a mechanistic model of ancestry-related assortative mating and ancestry-related sex bias. Under this model, we simulate a range of admixture scenarios iterating assortative mating, sex bias and other demographic parameters. Then, we implement an artificial neural network using the ancestry tract length information from the simulations, exploring both fully connected and convolutional architectures, and multiple output regression layers. Once the neural network is trained, we infer mating and demographic parameters of the admixing American populations from genomic data. We show how ancestry-related assortative mating and sex bias have constrained the admixture process between Native American, European and sub-Saharan African genetic components since the European colonization and the subsequent Atlantic slave trade.

Overall, we demonstrate the potential of machine learning to disentangle unsolved challenges in population genetics, such as the complexity of admixture events. Furthermore, we highlight the importance of the cultural and socioeconomic context for understanding the evolutionary processes of the genetic history of human populations.

SYMP8-7

Accurate detection of interspecific positive selection using convolutional neural networks

Conor R Walker[1,2], Nicola De Maio[1], Nick Goldman[1]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridgeshire, United Kingdom. [2]Department of Genetics, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

Abstract

Detecting adaptive changes in multiple sequence alignments of protein-coding DNA sequences is typically performed using likelihood-based tests to identify positive selection, as embodied in software such as PAML and HyPhy. These tests typically involve statistical evaluation of the nonsynonymous to synonymous substitution rate ratio (“dN/dS” or “ω”), at individual sites, or on individual branches, or both. Likelihood-based tests perform well in idealised scenarios involving perfect alignments, but increasingly poorly as greater levels of divergence lead to multiple indels and alignment errors. We show that convolutional neural networks (CNNs) can learn to accurately detect selection using alignments containing many of these errors, demonstrating that test accuracy does not have to be constrained entirely by alignment quality.

We treat this as a binary classification problem. CNNs are tasked with identifying if either positive selection (conventionally ω>1), or no selection, has occurred within an alignment at any time since the sequences' divergence from a common ancestor. We simulate sequence evolution under a range of realistic conditions, using alignments of these sequences for training, validating, and testing our networks. We show that CNNs trained on Clustal Omega alignments can classify selection with high accuracy in the presence of alignment error, performing favourably when compared to best-in-class aligner+PAML combinations. We use a technique for calculating global site class saliency maps, which explain the regions of an input alignment that were informative for classification; these show that CNNs learn site-wise information that may permit expanding this approach into evaluating selection at individual sites.

SYMP8-8

A deep learning framework for dimensionality reduction of genotype data

Kristiina Ausmees, Carl Nettelblad
Uppsala University, Uppsala, Sweden

Abstract

Dimensionality reduction is a widely employed technique in the field of genomics. The lower-dimensional representation allows for visualization and can be used for various characterizations of genetic variation, including detection of population structure and relatedness.

However, many of the standard statistical tools used for dimensionality reduction come with limitations and may be unsuitable for particularly challenging cases. Examples of such scenarios are cohorts with high levels of genetic variation, the presence of rare variation or LD in the data, as well as high levels of missing data, as is often the case when analyzing ancient DNA.

In this project, we explore the use of deep learning methods to obtain a more robust method that can overcome some of the above mentioned challenges. Convolutional neural networks are a type of deep learning model that can capture local spatial dependencies in data and have been successfully used in applications involving noisy and sparse data. We present a convolutional autoencoder architecture for performing nonlinear dimensionality reduction of genotype data in a manner that exploits the sequential nature of genetic information.

The proposed model is tailored to capture local as well as global patterns in sequence data, and we demonstrate that it can learn a useful representation of human genetic variation. We compare performance of our model to the alternative techniques of PCA and t-SNE, and show that the autoencoder can achieve dimensionality reduction with richer visual information for highly diverse cohorts, as well as yield a more accurate population classification model.

SYMP11-1

Leveraging ancient and modern DNA to investigate adaptive introgression

Emilia Huerta-Sanchez[1,2], Kelsey Witt[1], Alyssa Funk[1], Fernando Villanea[3], David Peede[1], Elle Loughran[2], Maria Avila Arcos[4]

[1]Brown University, Providence, Rhode Island, USA. [2]Trinity College Dublin, Dublin, Ireland. [3]University of Colorado Boulder, Boulder, Colorado, USA. [4]International Laboratory for Human Genome Research, UNAM, Queretaro, Mexico

Abstract

Several studies have shown that archaic introgression facilitated adaptations in Eurasian populations, but much less is known about the role of archaic variation in facilitating adaptation to environments in the Americas. Humans expanded into North and South America very recently and rapidly, and therefore we would expect that positive selection may have acted on standing archaic variants to facilitate adaptations to these new environments. To characterize the landscape of adaptive introgression in these populations, we scan the genomes of admixed populations from the Americas (from the 1000 Genomes Project) to identify signatures of adaptive introgression. As these populations are admixed, archaic haplotypes from Denisovans and Neanderthals can lie in regions of European or indigenous American ancestry. We use local ancestry calls and available DNA sequence data from ancient individuals from the Americas to identify a set of candidate genes with archaic variation that may have accelerated adaptations to the new environments, compared to the process of selection on de novo mutations alone.

SYMP11-2

Investigation of *CYP450* genes in archaic individuals and the implications for evolutionary medicine

Tadeusz H Wroblewski[1], Kelsey E Witt[2], Seung-been Lee[3], Ripan S. Malhi[4], Emilia Huerta-Sanchez[2], Fernando Villanea[5], Katrina G Claw[1]

[1]Division of Biomedical Informatics and Personalized Medicine, Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [2]Ecology and Evolutionary Biology and Center for Computational and Molecular Biology, Brown University, Providence, RI, USA. [3]Precision Medicine Institute, Macrogen Inc., Seoul, Korea, Republic of. [4]Department of Anthropology, University of Illinois Urbana-Champaign, Champaign, IL, USA. [5]Department of Anthropology, University of Colorado Boulder, Boulder, CO, USA

Abstract

Modern humans carry archaic genome elements that play important roles in human adaptation to their environments and continue to influence various phenotypes. The cytochrome P450 (*CYP450*) genes encode oxidase enzymes that function in metabolism of endogenous small molecules and in detoxification of exogenous compounds, often toxic chemicals produced by plants, fungi, and bacteria, in the local environment. These genes show evidence of positive selection and high allele frequency variation in modern humans, suggesting some role in adaptation, but archaic variation and introgression remains unstudied. We investigate eleven key *CYP450* genes, involved in metabolizing 75% of small molecule drugs, in three Neanderthal and one Denisovan individuals. Using our computational pipeline, we identify genetic variants, predict phased diplotypes, and infer metabolizer phenotypes of eleven *CYP450* genes in the archaic individuals. To examine introgression, we calculate the allele frequencies of identified archaic variants in modern human populations and cluster haplotypes. We observed over 1500 variants in the archaic individuals, including potentially deleterious variants in many of the genes investigated, suggesting that some enzymes may not have been essential in archaic hominins. We found the structural variant *CYP2A6*12* – responsible for a slow metabolizer phenotype of nicotine – present in all Neanderthal individuals and identified an introgressed *CYP2J2* Neanderthal variant at high frequencies in modern populations. We show that the contribution of archaic ancestry to pharmacologically relevant genes, such as the *CYP450* genes, is important to health and provides insights into the role of archaic alleles in human adaptation to the environment.

SYMP11-3

Natural selection and adaptive introgression in Pacific islanders

Jeremy Choin[1,2], Javier Mendoza-Revilla[1], Lara R. Arauna[1], Sebastian Cuadros-Espinoza[1,3], Olivier Cassar[4], Maximilian Larena[5], Albert Min-Shan Ko[6], Christine Harmant[1], Romain Laurent[7], Paul Verdu[7], Guillaume Laval[1], Anne Boland[8], Robert Olaso[8], Jean-François Deleuze[8], Frédérique Valentin[9], Ying-Chin Ko[10], Mattias Jakobsson[5,11], Antoine Gessain[4], Laurent Excoffier[12,13], Mark Stoneking[14], Etienne Patin[1], Lluis Quintana-Murci[1,15]

[1]Human Evolutionary Genetics Unit, Institut Pasteur, UMR 2000, CNRS, Paris 75015, France. [2]Université Paris Diderot, Sorbonne Paris Cité, Paris 75013, France. [3]Sorbonne Université, Collège doctoral, Paris 75005, France. [4]Oncogenic Virus Epidemiology and Pathophysiology, Institut Pasteur, UMR 3569, CNRS, Paris 75015, France. [5]Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden. [6]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China. [7]Muséum National d'Histoire Naturelle, UMR7206, CNRS, Université de Paris, Paris, France. [8]Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, Evry, France. [9]Maison de l'Archéologie et de l'Ethnologie, UMR 7041, CNRS, Nanterre, France. [10]Environment-Omics-Disease Research Center, China Medical University and Hospital, Taichung, Taiwan. [11]Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [12]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [13]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [14]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [15]Collège de France, Paris, France

Abstract

Pacific islanders primarily descend from two ancestral groups, who can be related to the out-of-Africa dispersal ~50,000 ya and the Austronesian expansion ~5,000 ya, the most recent human expansion into empty territories. Importantly, their genomes also carry the highest degree of combined Neanderthal and Denisovan ancestry of all human groups. However, the detailed genomic history of Oceanians remains largely uncharacterised and how Neandertals and Denisovans contributed adaptive genetic material to Oceanians needs to be explored in further detail. Here, we generated high-coverage genomes for 317 individuals from 20 populations from Near and Remote Oceania. First, we show that the Neanderthal legacy is highly similar among Pacific populations (~2,5%), while that of Denisovans is highly variable (~0-3,2%) and generally correlate with Papuan-related ancestry. Furthermore, while Neanderthals facilitated human adaptation related to multiple phenotypes (e.g., pigmentation, neuronal development, etc), Denisovan introgression was primarily beneficial for the regulation of innate and adaptive immune functions. We also report evidence of selective sweeps and polygenic adaptation associated with pathogen exposure and lipid metabolism in the Pacific. Finally, although subtle differences in the genetic load of Pacific islanders exist, we find that deleterious variants, including loss-of-function variants, tend to segregate at higher frequency in some Oceanian groups, likely due to stronger genetic drift. Collectively, our analyses provide novel insights into the demographic and adaptive history of Pacific populations, including their various interactions with archaic hominins, and increased our understanding of the mechanisms of biological adaptation to island environments.

SYMP11-4

Reconstructing ancient traits and recovering variants under positive selection at the Mesolithic-Neolithic transition using 1,490 ancient imputed genomes

Alba Refoyo Martínez[1], GeoGenetics MesoNeo Consortium[1], Evan Irving-Pease[1], Martin Sikora[1], Morten E Allentof[1,2], Olivier Delaneau[3,4], Kristian Kristiansen[5], Rasmus Nielsen[1,6,7], Eske Willerslev[1,8,9,10], Fernando Racimo[1]

[1]Lundbeck Foundation GeoGenetics Centre, Globe Institute, University of Copenhagen, Copenhagen, Denmark. [2]Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, Western Australia, Australia. [3]Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland. [4]Swiss Institute of Bioinformatics (SIB), University of Geneva, Geneva, Switzerland. [5]Department of Historical Studies, University ofGothenburg, Gothenburg, Sweden. [6]Department of Integrative Biology, UC Berkeley, Berkeley, CA, USA. [7]Department of Statistics, UC B, Berkeley, CA, USA. [8]Department of Zoology, University of Cambridge, Cambridge, United Kingdom. [9]. The Danish Institute for Advanced Study, University of Southern Denmark, Odense, Denmark. [10]The Wellcome Trust Sanger Institute, Cambridge, United Kingdom

Abstract

Between 12,000 and 5,000 years ago, Western Eurasia underwent a dramatic social transformation: the development of agricultural practices in the Middle East was followed by a mass movement of people from Anatolia into Europe. This process – known as the Mesolithic-Neolithic transition – had major consequences for human societal organization, cultural practices, health and genetics. To understand its genomic impact across time and space, we generated a dataset of 318 ancient genomes from this period at an average of 0.75X coverage. The dataset was combined with other published sequences, and a total of 1,490 ancient genomes were imputed and phased using present-day reference panels. We then reconstructed polygenic scores for phenotypes of ancient hunter-gatherers, farmers and steppe nomad populations, using effect size estimates from genome-wide association studies conducted on the UK Biobank. We found that the most significantly over-dispersed scores correspond to variants associated with traits related to pigmentation, anthropometric traits and disorders associated to diet and sugar levels, suggesting strong population trait differences preceding the transition, followed by trait homogenization via subsequent admixture. We also looked for strong episodes of positive selection on particular genetic variants, and recovered several novel candidate genes involved in cardiovascular disorders, and glucose and lipid metabolism. Overall, this dataset provides a highly detailed picture of changes in genetic variation over several millennia, and of ancestral relationships between ancient humans and humans living in Western Eurasia today.

SYMP11-5

Predicting the effects of selection in the lab: a sexual selection experiment in *D. pseudoobscura*

Carolina Barata, Carolin Kosiol
University of St Andrews, St Andrews, Scotland, United Kingdom

Abstract

Experimental evolution studies have taught us to observe replication as a means of increasing the power to predict the outcomes of selection. Taking advantage of this experimental set up and combining it with next generation sequencing and statistical methods for estimating selection, allows us to analyse genome-wide polymorphism datasets as never before. Understanding short-term adaptation signatures in the genome is, thus, within reach.

We have investigated allele frequencies in a sexual selection experiment in flies that lasted for 200 generations (Snook *et al*, 2005). *Drosophila pseudoobscura* females are observed to be promiscuous in the wild. The experiment altered the flies' social environment such that each female was housed with one – monogamy (M) – or six – elevated polyandry (P) – males. One prediction is that sexual selection is alleviated in monogamy lines where there is no competition for mates. One also expects differences between autosomes and the X chromosome since more causative loci should be located on the X. Previous comparisons between the treatments suggest that this might be the case (Wiberg *et al*, 2020).

We have estimated the effective population size – N_e – across the genome for this experiment. Results indicate that average N_e is much lower than previously estimated (~90 and ~121 for M and P treatments, respectively). N_e is higher in P replicates, suggesting that selection might not have relaxed as a result of monogamy. Nevertheless, these results highlight the importance of how genetic drift affects allele frequencies. Finally, we apply a genome scan for selective sweeps, and estimate selection parameters.

SYMP11-6

Estimating selection coefficients and testing their changes from ancient DNA data

Xiaoyang Dai[1], Mark Beaumont[2], Feng Yu[2], Zhangyi He[3]
[1]Queen Mary University of London, London, United Kingdom. [2]University of Bristol, Bristol, United Kingdom. [3]Imperial College London, London, United Kingdom

Abstract

Recent advances in ancient DNA preparation and sequencing techniques have made available an ever-increasing amount of time serial samples of segregating alleles in ancestral populations. The temporal dimension provided by the incoming ancient DNA data can boost power to infer the strength and timing of natural selection, thereby generating insight into events that have shaped the phenotypes and genotypes of current species and populations. However, despite all its lure, utilising ancient DNA to study natural selection also involves considerable obstacles such as how to estimate the strength of natural selection applied to specific phenotypic traits while considering confounding factors such as demographic histories and genetic interactions. To address these issues, in this work we introduce a novel Markov chain Monte Carlo framework for Bayesian inference of natural selection from ancient DNA data while taking into account demographic histories, genetic linkages and epistatic interactions. Moreover, our framework provides a Bayesian procedure for testing hypotheses on the drivers of past selection events such as domestication. The performance and utility of our approach are shown with an application to the ancient DNA data associated with the loci encoding coat base colour and white pattern in horses.

SYMP11-7

Understanding mechanisms of rapid evolution using museum genomics

Angela McGaughran
University of Waikato, Hamilton, Waikato, New Zealand

Abstract

Some of my recent research uses DNA from Australian pest moths that were collected over the last 100 years, including populations that were never exposed to insecticides and those that were exposed to different classes of insecticide following the rapid occurrence of field resistance. In this project, I compared DNA sequences from these temporal samples to examine the genomic makeup of a rapid adaptive response to strong selective pressure (i.e., insecticide application). In my talk, I will present preliminary results that look at: (i) population structure across time and space; (ii) outlier analysis, including shifts in allele frequencies over time. Together, the examination of genomic shifts between pre-, mid-, and post-insecticide sampling points furthers our understanding of the potential mechanisms underlying rapid evolution in a major pest moth.

SYMP13-1

Guaranteeing unbiasedness in tests for polygenic adaptation

Jennifer Blanc, Jeremy J Berg
University of Chicago, Chicago, Illinois, USA

Abstract

In recent years, many groups have worked to detect polygenic adaptation events in humans using polygenic scores, which are computed as a sum of individual genotypes, weighted by the effect that each site has on a trait, as measured by a genome wide association study.

One key assumption of these analyses is that the effect sizes used to compute the polygenic scores are unbiased. However, if the phenotypes of individuals in the GWAS panel are stratified across an ancestry gradient, estimated effect sizes can become systematically biased, leading to biased polygenic score predictions, and spurious signals of natural selection. While standard principal component based approaches to controlling for stratification can remedy this issue in principle, there is currently no way to tell whether this procedure has worked.

Using theory from population and statistical genetics, we show how to guarantee unbiasedness in polygenic adaptation tests and other polygenic score analyses. Specifically, by identifying the specific test to be performed before conducting the GWAS, one can compute a covariate to include in the GWAS which will render the polygenic adaptation test unbiased. Our procedure does not require explicit inference of a demographic model, knowledge of the specific environmental confounders present in the GWAS, and follows immediately from patterns of shared genetic covariation between the GWAS and polygenic score panel. These results place the study of polygenic adaptation via polygenic scores back on firm theoretical footing, and have broader relevance for phenotypic prediction in both modern and ancient individuals.

SYMP13-2

Filling the gap: a comparative analysis of imputation methods using trait datasets

Jacqueline A. May, Zeny Feng, Sarah J. Adamowicz
University of Guelph, Guelph, Ontario, Canada

Abstract

Missing observations in trait datasets pose an obstacle for analyses in the fields of molecular evolution, ecology, and biodiversity science. Imputation offers an alternative to removing cases with missing values from datasets, and techniques that incorporate phylogenetic information into their estimations demonstrate improved accuracy over standard techniques. However, previous studies of phylogenetic imputation tools are largely limited to simulations of continuous data and specific taxa. In addition, phylogenetic information is usually included in the form of a multigene tree; it remains to be explored whether the type of genetic data included in the tree affects imputation accuracy. We conducted a study to compare the performance of several imputation methods on datasets comprised of real trait data. Numerical and categorical data were assembled from 17 trait databases for taxa across the tree of life (mammals, fishes, reptiles, birds, insects). Gene trees were constructed from mitochondrial and nuclear markers and used as the source of phylogenetic information. Selected methods included k-nearest neighbour, random forests, multiple imputation by chained equations, and Rphylopars. Known data were removed at varying levels from complete datasets (e.g. 10%-40%). To assess imputation accuracy, each method (with and without phylogenetic information) was used to fill in the best estimate of the missing value. Given a taxa and trait combination, recommendations are provided for an imputation strategy and set of informative genes to mitigate the issue of missing data. These insights may be useful for studies investigating the evolutionary history of traits, particularly in rare or poorly studied taxa.

SYMP13-3

Structures by genetic connectedness between modern biobanks and imputed ancient genomes

Toomas Kivisild
KU Leuven, Leuven, Belgium. Institute of Genomics, University of Tartu, Tartu, Estonia

Abstract

How population structures form, are maintained, and decay within and among populations are questions with relevance for the study of the evolutionary history of populations, diversity of simple and complex traits. The knowledge about these processes has a range of downstream applications in biological, medical sciences, and humanities. Combination of fast and scalable methods to detect identical-by-descent segments and to extract community structures from large genomic data sets allow us to study the genetic structure of modern biobanks in unprecedented detail. In parallel, application of ancient DNA methods on increasing numbers of individuals from the past enable us to directly test in time evidence for migration, admixture, and adaptation at increasingly finer resolution. However, while community extraction methods require high quality genotype information most ancient genomes remain characterized at low coverage. In this presentation the potential use and limitations of imputed ancient genomes for the interpretation of genetic relatedness and ancestry-based patterns of connectedness in modern biobank data will be discussed.

SYMP13-4

A novel statistical approach to detect selection in admixed individuals reveals new adaptation signals in Latin Americans associated with complex phenotypes

Javier Mendoza-Revilla[1,2], Juan Camilo Chacon-Duque[1], Macarena Fuentes-Guajardo[3], Victor Acuña-Alonzo[1], Louise Ormond[1], Ke Wang[4], Rolando Gonzalez-Jose[5], Maria Catira Bortolini[6], Samuel Canizales-Quinteros[7], Carla Gallo[8], Giovanni Polleti[8], Gabriel Bedoya[9], Francisco Rothhammer[10], Kaustubh Adhikari[1], Andres Ruiz-Linares[11,12], Garrett Hellenthal[1]

[1]University College London, London, United Kingdom. [2]Institut Pasteur, Paris, France. [3]Universidad de Tarapaca, Arica, Chile. [4]Max Planck Institute for the Science of Human History, Jena, Germany. [5]Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina. [6]Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [7]Instituto Nacional de Medicina Genómica, Mexico City, Mexico. [8]Universidad Peruana Cayetano Heredia, Lima, Peru. [9]Universidad de Antioquia, Medellin, Colombia. [10]Instituto de Alta Investigación, Arica, Chile. [11]Aix-Marseille Université, Marseille, France. [12]Fudan University, Shanghai, China

Abstract

Recent admixture is ubiquitous in human populations, yet many methods to detect selection do not model it. Approaches that infer the ancestral source of each DNA segment in admixed individuals (i.e. local-ancestry-inference, LAI) can be powerful in this setting, but only if admixing sources are sufficiently genetically different. Here we introduce an alternative approach that, without requiring LAI, identifies SNPs under selection in admixed populations, infers selection coefficients, and accurately classifies whether selection occurred post-admixture or prior to admixture in one of the ancestral source populations.

In a new analysis of ~4,000 admixed Latin Americans, our approach infers evidence of selection at >30 previously unreported loci. By leveraging results of expression quantitative loci (eQTL) and genome-wide association (GWAS) studies, we provide evidence of adaptation associated with immune, pigmentation and metabolic phenotypes. This includes signals of selection in Native Americans prior to admixture at SNPs associated with (i) white cell count and gene expression in T-cells; (ii) expression in adipose tissue of a gene associated with glucose levels during pregnancy; and (iii) preeclampsia, a pregnancy condition prevalent among Andean women. Among our signals of post-admixture selection, we report a SNP strongly associated with the expression of a nearby gene involved in T cell proliferation. Our findings emphasize the benefits of combining selection scans with published information on complex traits, and suggest that diverse ecological and historical changes, such as the introduction of novel pathogens to the Americas, have shaped the genomes of Indigenous and admixed Americans by affecting phenotypes of biomedical relevance.

SYMP13-5

Quantifying signatures of diverse evolutionary forces genome-wide across >600 human traits

Abin Abraham[1], Abigail LaBella[1], John A. Capra[2], Antonis Rokas[1]
[1]Vanderbilt University, Nashville, TN, USA. [2]UCSF, San Francisco, California, USA

Abstract

Natural selection has shaped the genetic architecture of many human traits. While the action of specific evolutionary forces on some traits has been identified, the role of different modes of selection across the human phenome remains poorly understood. Using summary statistics from genome-wide association studies (GWASs), we developed a computationally efficient permutation-based framework to calculate enrichment for 11 evolutionary metrics based on sequence constraint, populations differentiation, and allele age that accounts for linkage disequilibrium, allele frequency, and other confounders. We applied the framework to >600 GWASs to generate a comprehensive map linking selection to the human phenome. On a smaller subset of high-confidence GWASs, we detected pervasive (42/44 traits), statistically significant enrichments for sequence conservation at trait-associated loci across species and within human populations. The trait ‘number of children’ had the highest enrichment for sequence conservation (mean LINSIGHT=0.52). Additionally, we found enrichments between 1000 Genomes super-populations for population differentiation (mean Fst: 0.08 to 0.20; XP-EHH: -0.53 to 0.74) among hair, skin, and pigmentation traits. Evolutionary metric enrichment also varied based on GWAS effect size. Loci associated with the smallest absolute effect size had statistically significant enrichment for population differentiation (mean Fst=0.15) and depletion of balancing selection (mean BetaScore=4.7). Our enrichments not only reinforce the role of negative selection in the human genome, but also quantify diverse modes of selection acting on specific subsets of the phenome. The catalog of natural selection aggregated across many traits will enable exploration of the relationship between the genetic architecture and selection.

SYMP13-6

Reconstructing recent human evolution with ancient DNA

Iain Mathieson
University of Pennsylvania, Philadelphia, PA, USA

Abstract

Throughout our history, humans have experienced–and adapted to–dramatic changes in environment and selective pressures. While the historical evidence of adaptation can be found in our genomes today, recent advances in ancient DNA technology allows us to directly observe genetic changes associated with evolution. Here, we describe three cases where ancient DNA has provided novel insights into the evolution of complex traits.

First, variation at the FADS locus involved in lipid metabolism has been under selection for tens of thousands of years across multiple human populations. In Europe, the most recent episode of selection, which started in the Bronze Age, is still ongoing today. Because of the complexity of the historical selection at this locus, ancient DNA is invaluable for reconstructing its evolutionary history and association with changes in diet. Second, we investigate the evolution of variation associated with skin pigmentation. Although pigmentation is highly polygenic, we show that selection in Europe was in fact oligogenic–driven by a relatively small number of coding variants–not by polygenic selection on gene expression. Finally, we show that polygenic scores for height can be used to predict the height of ancient individuals and populations. We identify changes in height and other anthropometric traits that are likely driven by environmental and genetic changes. Overall, our results show how ancient DNA can be combined with data from present-day populations, as well as archaeological and environmental data, to provide new insights into the evolution of complex traits.

SYMP16-1

Pervasive translation of lncRNAs and the birth of new proteins

M. Mar Alba
Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra (UPF), Barcelona; Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

Abstract

Recent studies based on the analysis of ribosome profiling (Ribo-Seq) data have shown that there are thousands of transcripts annotated as long non-coding RNAs (lncRNAs) that contain translated small open reading frames (smORFs). A subset of the smORFs are well conserved across relatively distant species and thus represent misannotated protein-coding genes. Many of them, however, show poor phylogenetic conservation. We have used polymorphism data to investigate the signatures of purifying selection in mouse smORFs detected by Ribo-Seq. We have found that, whereas some especies-specific smORFs show evidence of selection, a large fraction of them appear to encode neutrally evolving peptides. We argue that the pervasive translation of the transcriptome provides the raw material for the birth of new funcional proteins.

SYMP16-2

Evolutionary flexibility in a bacterial tRNA set

Gökçe B Ayan[1], Wing Yui Ngan[1], Hye Jin Park[1,2], Jenna Gallie[1]
[1]Max Planck Institute for Evolutionary Biology, Plön, Germany. [2]Asia Pacific Center for Theoretical Physics, Pohang, Korea, Democratic People's Republic of

Abstract

During translation, consecutive mRNA codons are matched with their respective amino acids by transfer RNA (tRNA) molecules. This matching process occurs by successive, random sampling of tRNA molecules from the mature tRNA pool. Hence, the average speed at which any given mRNA is translated depends on the composition of the tRNA pool. In organisms where rapid translation is important, selection is expected to favour tRNA pools that minimize search time. In support of this expectation, we recently observed the adaptive evolution of the tRNA pool of the rapidly-growing bacterium, *Pseudomonas fluorescens* SBW25, in response to changing translational needs.

Deletion of *serCGA* from SBW25 resulted in the elimination of tRNA-Ser(CGA) from the mature tRNA pool. The accompanying growth defect was rapidly compensated during a serial transfer evolution experiment, *via* two routes: (i) duplication of *serTGA* (the gene encoding tRNA-Ser(UGA)), and (ii) a point mutation in the *serTGA* promoter. These genetic changes were reflected in the mature tRNA pool; *serCGA* deletion resulted in tRNA-Ser(CGA) elimination, while both compensatory mutations were accompanied by an increase in tRNA-Ser(UGA). These changes are hypothesized to affect translational speed and bacterial growth as follows: tRNA-Ser(CGA) elimination is expected to slow translation of its cognate codon UCG, a relatively high-use codon that can also be translated by tRNA-Ser(UGA). The increased translational demand for tRNA-Ser(UGA) is then presumably relieved by *serTGA* mutations that serve to elevate tRNA-Ser(UGA) levels.

Overall, our results demonstrate that a bacterial tRNA pool can rapidly adapt to changing translational needs.

SYMP16-3

Evolution of the Ssd1 family of translational regulators as inactive pseudonucleases, arising from a fungal branch of the Dis3/RNase II family of nucleases, with a novel RNA-binding surface.

Rosemary A. Bayne[1], Elizabeth R.B. Ballou[2,3], Uma Jayachandran[4], Aleksandra Kasprowicz[4], Stefan Bresson[4], David Tollervey[4], Atlanta G. Cook[4], Edward W. J. Wallace[1]

[1]Institute for Cell Biology and SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom. [2]MRC Centre for Medical Mycology, University of Exeter, Exeter, United Kingdom. [3]Institute for Microbiology and Infection, University of Birmingham, Birmingham, United Kingdom. [4]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

Abstract

The RNase II family of 3′-5′ exoribonucleases are present in all domains of life, and eukaryotic family members Dis3 and Dis3L2 play essential roles in RNA degradation. Ascomycete yeasts contain both Dis3 and inactive RNase II-like “pseudonucleases”. The latter function as RNA-binding proteins that regulate translation and affect cell growth, cytokinesis, and fungal pathogenicity. However, the evolutionary origins of these pseudonucleases are unknown: what sequence of events led to their novel function, and when did these events occur?

Here, we show how RNase II pseudonuclease homologs, including *Saccharomyces cerevisiae* Ssd1, are descended from active Dis3L2 enzymes. During fungal evolution, active site mutations in Dis3L2 homologs have arisen at least four times, in some cases following gene duplication. Our new crystal structure of Ssd1 shows that the ancestral RNA-binding "funnel" leading to the active site is blocked by loop insertions, implying emergence of a novel RNA-binding site. In contrast, N-terminal cold-shock domains and regulatory features are conserved across diverse dikarya and mucoromycota. We map the RNA-binding sites of Ssd1 by UV crosslinking and high-throughput sequencing, and show that mutations to a conserved surface of the cold-shock domains reduce RNA binding.

In the basidiomycete pathogenic yeast *Cryptococcus neoformans*, the single Ssd1/Dis3L2 homolog is required for cytokinesis from polyploid “titan” growth stages. This phenotype is consistent with those of inactive fungal pseudonucleases, yet the protein retains an active site sequence signature. We propose that a nuclease-independent function for Dis3L2 arose in an ancestral hyphae-forming fungus, involving RNA-binding on the surface of the cold shock domains.

SYMP16-4

Determinants of genome-wide distribution and evolution of uORFs in eukaryotes

Hong Zhang[1], Yirong Wang[1,2], Xinkai Wu[1], Xiaolu Tang[1], Changcheng Wu[1], Jian Lu[1]
[1]State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China. [2]College of Biology, Hunan University, Changsha, China

Abstract

Upstream open reading frames (uORFs) play widespread regulatory functions in modulating mRNA translation in eukaryotes, but the principles underlying the genomic distribution and evolution of uORFs remain poorly understood. Here, we analyze ~17 million putative canonical uORFs in 478 eukaryotic species that span most of the extant taxa of eukaryotes. We demonstrate how positive and purifying selection, coupled with differences in effective population size (N_e), has shaped the contents of uORFs in eukaryotes. Besides, gene expression level is important in influencing uORF occurrences across genes in a species. Our analyses suggest that most uORFs might play regulatory roles rather than encode functional peptides. We also show that the Kozak sequence context of uORFs has evolved across eukaryotic clades, and that noncanonical uORFs tend to have weaker suppressive effects than canonical uORFs in translation regulation. This study provides insights into the driving forces underlying uORF evolution in eukaryotes.

SYMP16-5

Combinatorial regulation of gene expression by uORFs and microRNAs in *Drosophila*

Hong Zhang[1], Yirong Wang[1,2], Xiaolu Tang[1], Shengqian Dou[1], Yuanqiang Sun[1], Qi Zhang[1], Jian Lu[1]
[1]Peking University, Beijing, China. [2]Hunan University, Changsha, China

Abstract

To ensure the accuracy of developmental programs and maintain cellular homeostasis, the process of gene expression is tightly controlled at all levels by multiple *cis*- and *trans*-regulators. However, little is known about the organization of different regulatory layers and its significance. Here, we showed that uORFs and miRNAs tend to regulate overlapping sets of genes in *Drosophila melanogaster*, and genes under such coregulation are enriched for transcription factors and signaling genes. We then provided evidence that the combination of miRNA targeting with uORFs repress translation in an additive and fail-safe manner. Moreover, we found that mRNAs coregulated by uORFs and miRNAs tend to have lower PABP occupancy and higher decay rate, which cannot be explained by codon optimality of coregulated mRNAs. We also demonstrated that the fail-safe repression mediated by uORFs and miRNA targeting is evolutionarily conserved in animals. Altogether, this present study provides novel insights into our understanding of post-transcriptional gene expression regulation.

SYMP16-6

Widespread ribosome collisions promote cotranslational protein folding

Taolan Zhao[1], Yan-Ming Chen[1], Xiang Yu[2,3], Jia Wang[1], Siyu Chen[1], Ning Gao[2], Wenfeng Qian[2]
[1]Institute of Genetics and Developmental Biology, Beijing, China. [2]Peking University, Beijing, China. [3]Tsinghua University, Beijing, China

Abstract

The folding of proteins is challenging in the highly crowded and sticky environment of a cell. Regulation of translation elongation, which should have been optimized during evolution, may play a crucial role in ensuring the correct folding of proteins. Much of our knowledge regarding translation elongation comes from the sequencing of mRNA fragments protected by single ribosomes by ribo-seq. However, larger protected mRNA fragments have been observed, suggesting the existence of an alternative and previously hidden layer of regulation.

Here, we applied the high-throughput sequencing technologies to detect ribosome collisions, a subsequential event of translational pauses during which the 5′-elongating ribosome collides with the 3′-paused one. We detected widespread ribosome collisions that are related to slow ribosome release when stop codons are at the A-site, slow peptide bond formation from proline, glycine, asparagine, and cysteine when they are at the P-site, and slow leaving of polylysine from the exit tunnel of ribosomes. The structure of these collided ribosomes obtained by cryo-electron microscopy suggests a different conformation from the substrate of the ribosome-associated protein quality control pathway. Collisions occurred more frequently in the gap regions between α-helices, where a translational pause can prevent the folding interference from the downstream peptides. Paused or collided ribosomes are associated with specific chaperones, which can aid in the cotranslational folding of the nascent peptides.

Therefore, cells use regulated ribosome collisions to ensure protein homeostasis.

SYMP18-1

Haplotype-resolved *de novo* assembly of a Tujia genome suggests the necessity for high-quality population-specific genome references

Haiyi Lou[1], Yang Gao[1,2], Bo Xie[1], Yimin Wang[1], Haikuan Zhang[3], Miao Shi[3], Sen Ma[1], Xiaoxi Zhang[1,2], Chang Liu[1], Shuhua Xu[1,2,4,5,6]

[1]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [2]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [3]Berry Genomics, Beijing, China. [4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. [5]Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China. [6]Collaborative Innovation Center of Genetics and Development, Shanghai, China

Abstract

While the human reference assembly is continually being improved, it remains debatable whether a population-specific reference is necessary for each ethnic group. We applied multiple sequencing technologies to *de novo* assemble an individual genome (TJ1) from the Tujia population, an ethnic minority group most closely related to the Han Chinese. TJ1 provides a haplotype-resolved assembly of chromosome-scale high quality with N50 scaffold size >78 Mb. Notably, compared with GRCh38 and other *de novo* assemblies, TJ1 remarkably improved short-read mapping by ~2%, comparable to inter-individual genome-length difference, and enhanced calling precision by ~6-16% for structural variants. Furthermore, TJ1 facilitates detecting rare or low-frequency variants and identifying the fine-scale difference between closely-related populations, outstanding examples including population-stratified variants between Tujia and Han Chinese on genes like *LCT* and *UBXN8*. Our results support the necessity of a population-specific assembly and exemplify its particular value in the genetic analysis, especially for studying close-related populations.

Keywords

de novo assembly; Tujia; Population-specific reference genome; Haploid genome; Structural variation.

All authors contributed equally to this work.

SYMP18-2

Genetic admixture in the culturally unique Peranakan Chinese population in Southeast Asia

Degang Wu[1], Roger Foo[2], Chaolong Wang[1]

[1]Huazhong University of Science and Technology, Wuhan, Hubei, China. [2]National University of Singapore, Singapore, Singapore

Abstract

The Peranakan Chinese are culturally unique descendants of immigrants from China who settled in the Malay Archipelago ~300-500 years ago. Today, among large communities in Southeast Asia, the Peranakans have preserved Chinese traditions with strong influence from the local indigenous Malays. Yet, whether or to what extent genetic admixture co-occurred with the cultural mixture has been a topic of ongoing debate. We performed whole-genome sequencing (WGS) on 177 Singapore (SG) Peranakans and analyzed the data jointly with WGS data of Asian and European populations. We estimated that Peranakan Chinese inherited ~5.62% (95% confidence interval [CI]: 4.75-6.46%) Malay ancestry, much higher than that in SG Chinese (1.08%, 0.69-1.53%), southern Chinese (0.86%, 0.57-1.31%), and northern Chinese (0.25%, 0.18-0.33%). A sex-biased admixture history, in which the Malay ancestry was contributed primarily by females, was supported by X chromosomal variants, and mitochondrial (MT) and Y haplogroups. Finally, we identified an ancient admixture event shared by Peranakan Chinese and SG Chinese ~1,612 (95% CI: 1,345-1,923) years ago, coinciding with the settlement history of Han Chinese in southern China, apart from the recent admixture event with Malays unique to Peranakan Chinese ~190 (159-213) years ago. These findings greatly advance our understanding of the dispersal history of Chinese and their interaction with indigenous populations in Southeast Asia.

SYMP18-3

A dynamic 6,000-year genetic history of Eurasia’s eastern steppe

Choongwon Jeong[1], Ke Wang[2], Shevan Wilkin[2], William Timothy Treal Taylor[2], Bryan K. Miller[2], Jan H. Bemmann[3], Raphaela Stahl[2], Chelsea Chiovelli[2], Florian Knolle[2], Sodnom Ulziibayar[4], Dorjpurev Khatanbaatar[5], Diimaajav Erdenebaatar[6], Ulambayar Erdenebat[7], Ayudai Ochir[8], Ganbold Ankhsanaa[9], Chuluunkhuu Vanchigdash[5], Battuga Ochir[10], Chuluunbat Munkhbayar[11], Dashzeveg Tumen[7], Alexey Kovalev[12], Nikolay Kradin[13], Bilikto A. Bazarov[14], Denis A. Miyagashev[14], Prokopiy B. Konovalov[14], Elena Zhambaltarova[15], Alicia Ventresca Miller[2], Wolfgang Haak[2], Stephan Schiffels[2], Johannes Krause[2], Nicole Boivin[2], Myagmar Erdene[7], Jessica Hendy[2], Christina Warinner[2]
[1]Seoul National University, Seoul, Korea, Republic of. [2]Max Planck Institute for the Science of Human History, Jena, Germany. [3]Rheinische Friedrich-Wilhelms-Universitat, Bonn, Germany. [4]Institute of Archaeology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia. [5]Mongolian University of Science and Technology, Ulaanbaatar, Mongolia. [6]Ulaanbaatar State University, Ulaanbaatar, Mongolia. [7]National University of Mongolia, Ulaanbaatar, Mongolia. [8]International Institute for the Study of Nomadic Civilizations, Ulaanbaatar, Mongolia. [9]National Centre for Cultural Heritage of Mongolia, Ulaanbaatar, Mongolia. [10]Institute of History and Ethnology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia. [11]University of Khovd, Khovd, Mongolia. [12]Institute of Archaeology, Russian Academy of Sciences, Moscow, Russian Federation. [13]Far East Branch of the Russian Academy of Sciences, Vladivostok, Russian Federation. [14]Siberian Branch of the Russian Academy of Sciences, Ulan-Ude, Russian Federation. [15]East Siberian State Institute of Culture, Ulan-Ude, Russian Federation

Abstract

The Eastern Eurasian Steppe was home to historic empires of nomadic pastoralists, including the Xiongnu and the Mongols. However, little is known about the region’s population history. Here, we reveal its dynamic genetic history by analyzing new genome-wide data for 214 ancient individuals spanning 6,000 years. We identify a pastoralist expansion into Mongolia ca. 3000 BCE, and by the Late Bronze Age, Mongolian populations were biogeographically structured into three distinct groups, all practicing dairy pastoralism regardless of ancestry. The Xiongnu emerged from the mixing of these populations and those from surrounding regions. By comparison, the Mongols exhibit much higher eastern Eurasian ancestry, resembling present-day Mongolic-speaking populations. Our results illuminate the complex interplay between genetic, sociopolitical, and cultural changes on the Eastern Steppe.

SYMP18-4

Genetic connections and shared evolution of dark-skinned indigenous peoples in Asia

Lian Deng[1], Yuwen Pan[1], Yinan Wang[1], Hao Chen[1], Kai Yuan[1], Sihan Chen[2], Dongsheng Lu[1], Siti Shuhada Mokhtar[3], Thuhairah Abdul Rahman[4], Boon-Peng Hoh[5], Shuhua Xu[1,6]
[1]Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Shanghai, China. [2]Human Phenome Institute, Fudan University, Shanghai, China. [3]Institute of Medical Molecular Biotechnology, Faculty of Medicine, Universiti Teknologi MARA, Selangor, Malaysia. [4]Clinical Pathology Diagnostic Centre Research Laboratory, Faculty of Medicine, Universiti Teknologi MARA, Selangor, Malaysia. [5]Faculty of Medicine and Health Sciences, UCSI University, Kuala Lumpur, Malaysia. [6]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China

Abstract

Dark-skinned indigenous (DSI) people attract much attention for their unique and outstanding appearance, nevertheless, their genetic history and adaptive evolution remain mysteries. Here we conducted a population genomic study to dissect the genetic distinction and connection of broad geographical DSIs. Despite DSI groups show diverse genetic makeup and large inter-area genetic differentiation, we identified a basal Asian ancestry (bASN) specifically shared by the Asian DSIs. Interestingly, bASN was relatively enriched in ancient Asian human genomes dated as early as ~50,000 years before present, and diminished in more recent history. Notably, bASN was not likely derived from archaic hominins but rather modeled as a survived lineage of the initial peopling of Asia. Shared adaptations associated with the bASN were identified among DSI groups (e.g., *LIMS1* for hair morphology) and enriched in neurological functions at an identical locus (e.g., *NKAIN3*) or different loci in an identical gene (e.g., *TENM4*). It remains debatable whether the dark skin phenotype is an ancestral feature or a result of genetic convergence. We show that the phenotypic convergence of the dark skin in DSIs could have resulted from parallel evolution (e.g., *DDB1*), convergence driven by genetic admixture (e.g., *MTHFD1* and *RAD18*) or novel mutations (e.g., *STK11*), as well as notably purifying selection (e.g., *MC1R*). Our results provided new insights into the initial peopling of Asia and advanced understanding of the phenotypic convergence of DSI peoples.

SYMP18-6

Population admixture in the Neolithization of East Asia inferred from ancient genomes

Chuan-Chao Wang[1], Panxin Du[2], Guanglin He[1], Kongyang Zhu[1], Shaoqing Wen[3]

[1]Department of Anthropology and Ethnology, Institute of Anthropology, State Key Laboratory of Cellular Stress Biology, State Key Laboratory of Marine Environmental Science, National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, Fujian, China. [2]MOE Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai, Shanghai, China. [3]Institute of Archaeological Science, Fudan University, Shanghai, Shanghai, China

Abstract

The establishment of the complex societies in Neolithic China appears to have been associated with rapid population growth and cultural innovation. However, it remains unclear whether substantial human migrations mediated the culture changes due to a lack of ancient DNA. We here sampled and sequenced 20 individuals dating to 6000-4000 BP from Gansu and Shandong provinces in the Upper and Lower Yellow River Basin, respectively, and genotyped more than 1000 present-day individuals from Tibeto-Burman and Tai-Kadai speaking groups. Through the population genomic analysis, we observed a genetic structure change in the farming populations of the late Neolithic period compared with the earlier hunter-gatherers. The cultural innovation in the late Neolithic period had been associated with massive population migration and genetic admixture from the Neolithic farmers from the middle reaches of the Yellow River.

SYMP18-7

Human back migration from Sundaland to South Asia was driven by sea-level rises during the Last Glacial Maximum

HIE LIM KIM, Namrata Kalsi, Tanghua Li, Timothy Adam Shaw, Stephan C Schuster, Benjamin Horton
Nanyang Technological University, Singapore, Singapore

Abstract

Rapid sea-level rise between the Last Glacial Maximum (LGM) and the mid-Holocene flooded Sundaland, which changed dramatically Southeast Asian coastal landscapes. To understand the impact of the geographical changes on human demography, here we addressed the question by an interdisciplinary approach. We reconstruct sea level and paleogeography in Southeast Asia since the LGM with fine resolution of time and inferred human population history using 742 high-coverage whole-genome sequencing datasets from 59 ethnic groups in Southeast and South Asia. We inferred that rapid sea-level rise, in particular, meltwater pulses 1A (MWP1A, 4cm/year ~14,500–14,000 years ago) and 1B (MWP1B, 2cm/year ~11,500–11,000 years ago) caused flooding which broke land bridges and split landmass, and at the same time, multiple population splits occurred in Southeast Asia. Increasing population density by population expansion in the reduced land area was inferred, and it might trigger the migration of Southeast Asians from Sundaland toward South Asia as we found the signal of admixture in the same period. Our novel approach revealed one of the earliest instances of human migration driven by sea-level rise.

SYMP18-8

Genetic origins and sex-biased admixture of the Huis

Xixian Ma[1], Wenjun Yang[2], Yang Gao[1,3], Yuwen Pan[1], Yan Lu[1,4], Hao Chen[1], Dongsheng Lu[1], Shuhua Xu[1,3,4,5,6,7]

[1]Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [2]Key Laboratory of Fertility Preservation and Maintenance, the General Hospital, Ningxia Medical University, Yinchuan, China. [3]ShanghaiTech University, Shanghai, China. [4]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China. [5]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. [6]Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China. [7]Human Phenome Institute, Fudan University, Shanghai, China

Abstract

The Hui people are unique among Chinese ethnic minorities in that they speak the same language as Han Chinese (HAN) but practice Islam. However, as the second-largest minority group in China numbering well over 10 million, the Huis are under-represented in genomic studies. Here, we present the first whole-genome sequencing effort of 234 Hui individuals (NXH) aged over 60 who have been living in Ningxia, where the Huis are mostly concentrated. NXH are genetically more similar to East Asian than to any other global populations. In particular, the genetic differentiation between NXH and HAN (F_{ST} = 0.0015) is only slightly larger than that between northern and southern HAN (F_{ST} = 0.0010), largely attributed to the western ancestry in NXH (~10%). Highly-differentiated functional variants between NXH and HAN were identified in genes associated with skin pigmentation (e.g., *SLC24A5*), facial morphology (e.g., *EDAR*), and lipid metabolism (e.g., *ABCG8*). The Huis are also distinct from other Muslim groups such as the Uyghurs (F_{ST} = 0.0187), especially, NXH derived much less western ancestry compared with the Uyghurs (~50%). Modeling admixture history indicated that NXH experienced an episode of two-wave admixture. An ancient admixture occurred ~1,025 years ago, reflecting the west-east contacts during the Tang and Song Dynasty. A recent admixture occurred ~500 years ago. Notably, we identified considerable sex-biased admixture, i.e., excess of western males and eastern females contributing to the NXH gene pool. The origins and the genomic diversity of the Hui people imply the complex history of contacts between western and eastern Eurasians.

SYMP18-9

Genetic continuity of Indo-Iranian speakers since Iron Age in Southern Central Asia

Perle Guarino-Vignon[1], Nina Marchi[1,2], Evelyne Heyer[1], Céline Bon[1]

[1]UMR7206 Éco-Anthropologie, MNHN, Paris, France. [2]CMPG group, Institute of Ecology and Evolution (IEE - UNIBE), Bern, Switzerland

Abstract

Since prehistoric times, South Central Asia has been a region at the crossroads of the movement of people, cultures, and goods. Today, Central Asia is populated by populations divided into two cultural and linguistic groups: the Indo-Iranian group and the Turko-Mongolian group. Genetics unveiled that migrations from East Asia contributed to the spread of Turko-Mongolian populations in Central Asia and to the partial replacement of Indo-Iranian population. However, the origin of the latter is still little known. To shed light on this, we compare the genetic data on two current-day populations– Yaghnobis and Tajiks – with the increasing number of genome-wide data from published ancient individuals. Using PCA, Admixture, f3 and D-statistics we show that the present Indo-Iranian populations from Central Asia show a strong genetic continuity with the Iron Age samples from Turkmenistan and Tajikistan. With qpAdm, we model Yaghnobis as a mixture of 93% Iron Age individuals from Turkmenistan and 7% from Baikal. For the Tajiks, we observe a more important Baikal ancestry and an additional admixture event with a South Asian population. Our results therefore suggest that beside complex history and settlement, Central Asia shows a remarkable genetic continuity since the Iron Age, with only limited gene flow.

SYMP19-1

Global patterns of genome diversity in great apes and across the primate radiation

Tomas Marques-Bonet
Institute of Evolutionary Biology (UPF-CSIC), Barcelona, Spain

Abstract

Despite great advances in sequencing technologies, we are just starting to understand the genetic diversity and recent adaptations of our evolutionary close's relatives, the primates. In this talk, I will cover two projects that aims to expand our view on the genetic variation in apes and other primates and will discuss some of their consequences in conservation and in evolution.

First, I will discuss how non-invasive samples are an excellent source of DNA to study genome diversity. As an example, we have captured the complete chromosome 21 from > 800 fecal samples collected widely distributed across Africa discovering ~50% of new variation. We provide evidence for new barriers impeding genetic exchange that overlap with known geographical barriers among subspecies and the study of genetic connectivity at different time scales suggest that these barriers might have been recently permeable.

Finally, and despite 60% of the world's primate species are currently threatened with extinction only a few of them have been studied in depth, and still lack genomic resources. To this end, we have compiled high coverage whole genome sequences of >800 individuals including > 230 distinct species of primates representing all 16 extant families and thus dramatically increase the number of species with available genomic resources.

Taken together, our results provide a snapshot of the current of the current genetic status of different primate species, which will prove a valuable future asset for conservation management, particularly in regions of increased pressure in extinction risk.

SYMP19-2

Human-chimpanzee hybrid stem cells as a platform for exploring primate evolution

Rachel Agoglia, Danqiong Sun, Hunter Fraser
Stanford University, Stanford, CA, USA

Abstract

Among primates, humans display a unique trajectory of development responsible for the many traits specific to our species. However, the inaccessibility of primate tissues—particularly from our closest living relative, the chimpanzee—has limited our ability to study the evolution of human-specific traits.

We have developed a new platform to address these limitations. We first generated a panel of tetraploid hybrid stem cells by fusing human and chimpanzee induced pluripotent stem cells. We next applied these hybrid cells to study species divergence in two of our most distinctive traits: the brain and face. We differentiated our hybrid stem cells into brain organoids, performed RNA-seq to explore genome-wide patterns, and conducted follow-up experiments to investigate a candidate gene involved in human-specific neuropsychiatric traits. In sum, we have found this system has promising potential to reveal the molecular basis of human evolution.

SYMP19-3

New gene emergence in human genome: origination mechanism, molecular function and dedicated database

Shengjun Tan[1], Hao Yuan[1,2], Tianhan Su[2], Dan Zhang[1,2], Yi Shao[1,2], Chunyan Chen[1,2], Yong E. Zhang[1,2]

[1]Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [2]University of Chinese Academy of Sciences, Beijing, China

Abstract

Various mechanisms (*e.g.*, duplication) underlie a flux of lineage- or species-specific new genes, which shape phenotypic evolution. Mounting efforts including our own studied the origination process and biological function of new genes in numerous species especially in humans. First, we found that tandem duplication together with transposon (retrotransposons and DNA transposons) mediated duplication mechanisms generate new genes in human genome. In spite of the different nature of these mechanisms, they often generate incomplete duplications and chimeric gene structures. Second, we found that new genes tend to be involved in fast-evolving processes such as spermatogenesis or brain development. Different from this general picture, *HBBP1*, which appears to be an unexpressed pseudogene in non-human primates, confers human-specific essentiality in a seemingly conserved process, *i.e.*, erythropoiesis. Finally, despite the significance of primate- or human-specific new genes for lineage-specific traits, these genes tend to be uncharacterized. We thus have been maintaining and updating a new-gene-focused database for human genome (GenTree, http://gentree.ioz.ac.cn), which facilitate users to evaluate when and how a gene arises and what type of function it may have.

SYMP19-4

Functional genetic variation within and between primate species: Examples with biomedical and evolutionary significance from strepsirrhines and catarrhines

Jeffrey Rogers[1], Muthuswamy Raveendran[1], Ronald A. Harris[1], Elaine E. Guevara[2], Roger Wiseman[3], Rui Chen[1], Sara M. Thomasy[4], Ala Moshiri[4], Timothy Stout[1], Timothy H. Webster[5], Richard R. Lawler[6], Brenda J. Bradley[7], Jeannin Ranaivonasy[8], Joelisoa Ratsirarson[8], Anne D. Yoder[2], David O'Connor[3]
[1]Baylor College of Medicine, Houston, Texas, USA. [2]Duke University, Durham, North Carolina, USA. [3]Univ. of Wisconsin, Madison, Wisconsin, USA. [4]Univ. of California, Davis, California, USA. [5]University of Utah, Salt Lake City, Utah, USA. [6]James Madison University, Harrisonburg, Virginia, USA. [7]George Washington Univ., Washington, District of Columbia, USA. [8]Univ. of Antananarivo, Antananarivo, Madagascar

Abstract

Recent advances in genomic technologies have made large-scale analysis of DNA sequence differences within and between primate species practical. As a result, the amount and scope of genomic information concerning nonhuman primates is expanding rapidly. We have investigated whole genome sequences from a wide range of nonhuman primates, including Old World monkeys, New World monkeys and strepsirrhines. In this presentation, we will describe whole genome analyses of 853 rhesus macaques (*Macaca mulatta*) using Illumina short read sequencing methods. Our results document high levels of both total and protein-coding sequence polymorphism among the macaques. Average heterozygosity genome-wide is >0.0023 with 790,377 protein-coding single nucleotide variants. These results, combined with whole exome sequence data from an additional 1314 rhesus macaques provide extensive documentation of the nature, distribution and potential functional impact of this diversity. We will illustrate the impact of this variation by describing spontaneous macaque models of human genetic disease (e.g. achromatopsia and elevated risk for behavioral disorders) identified through sequencing. To illustrate the coming revolution in analyses of between species differences and ecological adaptations, we will describe derived genetic changes found in four sifaka species (*Propithecus coquerelli, P. verreauxi, P. tattersalli and P. diadema*) that have diets heavy in leaves. These species share coding sequence changes in genes relevant to folivory (*ACE* and *RNASE1),* along with rapid evolution in other genes related to digestion, intestinal absorption and lipid metabolism. Molecular changes shared with folivorous colobine monkeys constitute examples of parallel genomic adaptation.

SYMP19-5

Cone ratio variation in free-ranging Rhesus macaques (*Macaca mulatta*)

Amanda D Melin[1], Rachel A Munds[1], Eve Cooper[2], Mareike C Janiak[1,3], Linh Lam[1], Alex DeCasien[2], Sam Bauman[4], Michael Montague[5], Melween Martinez[4], Shoji Kawamura[6], James P Higham[2]
[1]University of Calgary, Calgary, Alberta, Canada. [2]New York University, New York, New York, USA. [3]University of Salford, Manchester, United Kingdom. [4]University of Puerto Rico, San Juan, Puerto Rico. [5]University of Pennsylvania, Philadelphia, Pennsylvania, USA. [6]University of Tokyo, Kashiwanoha, Japan

Abstract

Catarrhine primates (African and Asian monkeys and apes) differ from other mammals in that males and females both routinely possess trichromacy, i.e. color vision based on three retinal cone types maximally sensitive to long (L, reddish), medium (M, greenish), and short (S, blueish) wavelengths of light. The spectral tuning of the L, M and S opsins is conserved across the infraorder. However, the ratio of L to M cone expression within and between species is surprisingly variable and poorly understood. We took advantage of a rare sampling opportunity to study the retinal transcripts of 189 rhesus macaques (*Macaca mulatta*) from the large pedigreed population living on Cayo Santiago, Puerto Rico. We quantified expression of L, M and S opsin genes using droplet digital PCR of cDNA transcribed from retinal (foveal) RNA. We conducted heritability analyses using animal models to estimate the additive genetic variance of the cone ratio. We find an essentially equal ratio of L and M retinal cones, on average, with L:M ratio ranging from *ca* 1:2 to 2:1. The ratio of S cones relative to the other (L+M) cone population is *ca.* 1:7, with relatively less interindividual variation. We detected a weak impact of sex on L:M ratio and evidence that S, but not L:M, cone ratios are heritable. Our results provide new data on the biology of color vision variation and evolution in a model species of high biomedical relevance, and contextualize the extraordinary variation observed in human L and M cones.

SYMP19-6

Inferring speciation times and ancestral population history throughout the primate phylogeny using incomplete lineage sorting

Marjolaine Rousselle[1], Iker Rivas-González[1], Fang Li[2,3], Long Zhou[3], Josefin Stiller[2], Dongdong Wu[4,5,6], Kasper Munch[1], Mikkel H Schierup[1], Guojie Zhang[2,4,6,7]

[1]Bioinformatics Research Center, Aarhus University, Aarhus, Denmark. [2]Villum Center for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [3]BGI-Research, BGI-Shenzhen, Shenzhen, China. [4]State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China. [5]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan, China. [6]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan, China. [7]China National Genebank, BGI-Shenzhen, Shenzhen, China

Abstract

Inferred primate genomic divergence times are at odds with the fossil record. However, we know that ancestral polymorphism causes divergence times to be overestimations of the time when new species separate. To disentangle divergence from speciation times in the primate phylogeny, we used coalHMM, which combines a hidden Markov model to reconstruct the pattern of genealogies along an alignment, and the coalescent theory to infer speciation times and ancestral population sizes. Applying it on a genome-wide alignment encompassing 50 primates species from apes to lemurs, we inferred the genealogies of each fragment of the genome, and estimated speciation times and ancestral population sizes in 28 branches of the primate phylogeny. We report ancestral population sizes that are in the high range of values previously reported in the literature, with extensive variation from apes (~100,000) to lorises (~500,000). This variation explains the discrepancies between split time and divergence times, which are particularly pronounced in strepsirrhini, where we infer a split time between lemurs and lorises at ~10.5 Mya versus ~50-60 Mya in divergence time. Notably, we date the ancestor of hominoidea at ~13 Mya, and cercopithecidae at ~10 Mya, which is at least 5 Mya smaller than divergence estimates. In addition, we used the asymmetry of the divergence of fragments attributed to incongruent genealogies to detect potential ancestral introgression events. We assessed that most of the incongruencies are due to incomplete lineage sorting and not due to ancestral introgression.

SYMP19-7

Frequent lineage-specific substitution rate changes support an episodic model for protein evolution

Neel Prabh, Diethard Tautz
Max Planck Institute for Evolutionary Biology, Ploen, Germany

Abstract

Since the inception of the molecular clock model for sequence evolution, the investigation of protein divergence has revolved around the question of a more or less constant rate of overall sequence information change. Although anomalies in clock-like divergence are described for some proteins, nowadays, the assumption of a constant decay rate for a given protein family is taken as the null model. Still, so far, a systematic test of this null model has not been done at a genome-wide scale despite the databases' enormous growth. We focus here on divergence rate comparisons between closely related lineages, since this allows clear orthology assignments by synteny and unequivocal alignments, which are crucial for the determination of substitution rate changes. Thus, we generated a high-confidence dataset of syntenic orthologs from four ape species, including humans. Further analysis revealed that despite the appearance of an overall clock-like substitution pattern, a substantial number of proteins show lineage-specific acceleration and deceleration in divergence rates, or combinations of both in different lineages. Interestingly, when aggregated, even the families showing large lineage-specific rate perturbations can show overall rate equality. Our analysis uncovers a much more dynamic history of substitution rate changes in protein families. Which invalidates a pan-genome null model of constant decay, on the one hand, but remains compatible with the existing notion that aggregated data can be reliably used to estimate species splitting time. Ultimately, our data shows that a null model of constant change is not suitable to predict the evolutionary trajectories of individual proteins.

SYMP23-1

Patterns of nucleotide diversity under background selection and evolving recombination rates

Tom R Booker
University of British Columbia, Vancouver, BC, Canada. University of Calgary, Calgary, AB, Canada

Abstract

Analysis of population genomic datasets in many species has revealed that natural selection shapes patterns of genetic variability in response to recombination rate variation across the genome. In particular, background selection, the effect of purifying selection on linked variation, is thought to be ubiquitous across the genome and can give rise to the positive correlation between diversity and recombination rate reported in many species. The interpretation of patterns of genetic variability typically makes the implicit assumption that recombination rates are invariant over time, however recombination rates may evolve very rapidly in some lineages, for example due to the evolution of recombination hotspots. In such cases, empirical estimates of recombination rate variation in contemporary populations may not fully reflect ancestral recombination rate landscapes, which are relevant to patterns of genetic variability. In this study, I examine the house mouse *Mus musculus*, a species in which recombination rates have evolved rapidly. I use a combination of theory, simulations and data analysis to determine how rapid evolution of the recombination rate landscape may have influenced patterns of variability across the mouse genome. In particular, I focus on how rapid recombination rate evolution may have muted the correlation between nucleotide diversity and local recombination rate across the mouse genome.

SYMP23-2

Decreased adaptation at human disease genes as a possible result of low recombination between deleterious and advantageous variants

Chenlu Di, Diego Salazar Tortosa, M. Elise Lauterbur, David Enard
the University of Arizona, Tucson, Arizona, USA

Abstract

Despite our expanding knowledge of gene-disease associations and the medical importance of disease genes, their evolution has not been thoroughly studied across diverse human populations. In particular, understanding the relationship between disease and adaptation at the gene level is severely hampered by the fact that we don’t even know whether disease genes have experienced more, less, or as much adaptation as non-disease genes during recent human evolution. Here, we compared the rate of strong recent adaptation in the form of selective sweeps between disease and non-disease genes across 26 human populations from the 1,000 Genomes Project. After controlling for confounding factors that may impact the rate of selective sweeps, we found that disease genes have experienced far less selective sweeps compared to non-disease genes. Investigating further the possible causes of the sweep deficit at disease genes, we found that the sweep deficit is very strong at disease genes with both low recombination rates and with high numbers of associated disease variants, but is mostly inexistant at disease genes with higher recombination rates or lower numbers of associated disease variants. These observations strongly suggest that adaptation has been slowed down by the presence of interfering deleterious variants at disease genes, especially in low recombination regions. These results clarify the evolutionary relationship between disease genes and recent genomic adaptation, and suggest that disease genes are cursed not only by a higher load of segregating deleterious mutations, but also by an inability to adapt as much as the rest of the genome.

SYMP23-3

The impact of chromosomal fusions on recombination rate variation in wild populations

Covadonga Vara[1], Andreu Paytuví-Gallart[1,2], Yasmina Cuartero[3], Lucía Álvarez-González[1], Laia Marín-Gual[1], Cristina Marín-García[1], Francisca Garcia[1], Laia Capilla[1], Rosa Ana Sanchéz-Guillén[1,4], Zaida Sarrate[1], Riccardo Aiese-Cigliano[2], Walter Sanseverino[2], Jeremy B Searle[5], Jacint Ventura[1], Marc A Marti-Renom[3], François Le Dily[3], Aurora Ruiz-Herrera[1]

[1]Universitat Autònoma de Barcelona, Barcelona, Spain. [2]Sequentia Biotech, Barcelona, Spain. [3]CNAG-CRG, Barcelona, Spain. [4]Instituto de Ecología AC, Veracruz, Mexico. [5]Cornell University, Ithaca, USA

Abstract

The spatial folding of chromosomes and their organisation in the nucleus has profound regulatory impacts, yet how this affects recombination rates in germ cells remains unclear. Here, we use a multidisciplinary approach taking advantage of chromosome conformation capture following by deep sequencing (Hi-C) in combination with SNP genotyping and cytological analysis of crossover events to asses the fine scale genomic impact of chromosomal fusions on recombination. To this aim we analyzed a unique wild population of house mice characterised by a recent evolutionary origin and the presence of chromosomal fusions in polymorphic state. Our results show that chromosomal fusions alter the nuclear architecture of spermatocytes in meiotic prophase, causing a reduction in total number and chromosomal re-distribution of crossovers, ultimately affecting mice fertility. The presence of chromosomal fusions also resulted in differences in molecular diversity between populations, providing novel evidence that disturbances in recombination rates due to chromosome fusions result in detectable genomic footprints. Overall, our results provide new insights into how genome reshuffling perturbs recombination rates in the context of 3D genome folding and nuclear architecture of germ cells. We anticipate that our results will provide impetus for exploration of the functional and structural basis of recombination in a broad context, reinforcing the link between the 3D genome architecture and evolution.

SYMP23-4

Anti-recombination and hybrid sterility in budding yeasts

Jasmine Ono[1], G. Ozan Bozdag[2], David W Rogers[3], Ellen McConnell[3], Jai A Denton[4], Emre Karakoc[5], Neil Hunter[6], Jun-Yi Leu[7], Duncan Greig[1]

[1]University College London, London, United Kingdom. [2]Georgia Institute of Technology, Atlanta, GA, USA. [3]Max Planck Institute for Evolutionary Biology, Plön, Germany. [4]Okinawa Institute of Science and Technology, Onna-son, Japan. [5]Wellcome Sanger Institute, Hinxton, United Kingdom. [6]University of California, Davis, Davis, CA, USA. [7]Academia Sinica, Taipei, Taiwan

Abstract

Reproductive isolation, and consequently speciation, builds up from a combination of isolating barriers. We investigated the potential genetic causes of intrinsic reproductive isolation between two closely related species of budding yeast, Saccharomyces cerevisiae and S. paradoxus. F1 crosses between these species are viable but infertile. Two potential causes for this sterility are chromosome mis-segregation during meiosis and genic incompatibilities. All chromosomes in Saccharomyces yeast are essential. Thus, a gamete missing a single chromosome is inviable. We show that chromosome mis-segregation due to anti-recombination is common in this hybrid, and that it is linked to genome-wide sequence divergence. Build-up of genic incompatibilities, where genes that work well in their parental genetic background do not work well together in a hybrid genetic background, is thought to be a common mechanism for speciation. We break the first species barrier by restoring recombination in the hybrid, allowing us to uncover evidence of genic incompatibilities in this system.

SYMP23-5

Interference hypothesis for recombination suppression in chromosomal inversion heterozygotes: A formal genetics analysis in *Drosophila melanogaster*

Spencer A Koury
Stowers Institute for Medical Research, Kansas City, MO, USA. Stony Brook University, Stony Brook, NY, USA

Abstract

Chromosomal inversions are commonly associated with both adaptive genetic variation and the accumulation of deleterious mutations. Paradoxically, strong recombination suppression in inversion heterozygotes can act to both preserve beneficial epistatic interactions while simultaneously reducing efficiency of purifying selection. Surprisingly, recombination suppression extends outside the inverted regions where there are no barriers to normal pairing, synapsis, double strand break formation, or recovery of crossover products. This unexplained suppression outside of inversions accounts for a large fraction of the total phenotype, decaying over 10 Mb in *Drosophila melanogaster*, and may account for varying “strength” of recombination suppression observed for different inversions. The interference hypothesis of recombination suppression proposes heterozygous inversion breakpoints possess chiasma-like properties, predicting recombination suppression extends from breakpoints in a process analogous to crossover interference. To test this hypothesis, I adapted and parameterized a probabilistic model of crossover interference with gamma-distributed interevent distances. To test predictions, I scored >40,000 meioses for crossovers in intervals external and adjacent to four different cosmopolitan inversions of *Drosophila melanogaster*. The interference hypothesis accurately predicted recombination suppression in intervals adjacent to heterozygous inversion breakpoints but failed to explain elevated recombination in genetic intervals spanning the centromere. This result is consistent with well-documented, but counterintuitive, negative crossover interference associated with pericentric heterochromatin. Therefore, to understand the magnitude and extent of recombination suppression and the effects of inversions on genomic polymorphism will require models of crossover patterning that incorporate both the local effects of crossover interference as well as the regional effects of centromeres and pericentric heterochromatin.

SYMP23-6

Modeling Recombination Rate as a Quantitative Trait Reveals New Insight into Selection in Humans

Austin L Drury, Jean-Francois Gout, Amy L Dapper
Mississippi State University, Mississippi State, MS, USA

Abstract

Meiotic recombination is both a fundamental biological process required for proper chromosomal segregation during meiosis and a fundamental genomic parameter that shapes major features of the genomic landscape. However, despite the central importance of this phenotype, we lack a clear understanding of the selective pressures that shape its variation in natural populations, including humans. While there is strong evidence of fitness costs of low rates of recombination, the possible fitness costs of high rates of recombination are less defined. To determine whether a single lower fitness bound can explain the variation in recombination rate observed in human populations, we simulated the evolution of recombination rate as a quantitative trait. Under each scenario, we statistically compared the resulting trait distribution to the observed distribution of recombination rates from a published study of the Icelandic population. To capture the genetic architecture of recombination rate in humans, we modeled it as a moderately complex trait (20 contributing loci) with modest heritability (0.3 and 0.18 for males and females, respectively). For our fitness function, we implemented a hyperbolic tangent curve with several flexible parameters to capture a wide range of existing hypotheses. We found that a lower bound alone is insufficient to explain current variation in recombination rate in both males and females, supporting the existence of fitness costs of high rates of recombination in humans. With simulations using both upper and lower bounds, we describe a parameter space for an upper bound on recombination rate that produces results consistent with empirical observations.

SYMP24-1

Testing for allelic imbalance between environments

Lauren M McIntyre[1], Alison R Miller[1], Allison M Morse[1], Jacqueline E Borgert[1], Zihao Liu[1], Kelsey Sinclair[1], Katrina Sherbina[2], Sergey Nuzhdin[2], Gavin Gamble[1], Fei Zou[3], Jeremy R B Newman[1], Luis G León-Novelo[4], Fabio Marroni[5]
[1]University of Florida, Gainesville, Florida, USA. [2]University of Southern California, Los Angeles, California, USA. [3]University of North Carolina, Chapel Hill, North Carolina, USA. [4]University of Texas at Houston, Houston, Texas, USA. [5]University of Udine, Udine, Italy

Abstract

Allelic imbalance (AI) occurs when alleles in a diploid individual are differentially expressed and indicates *cis* acting regulatory variation. What is the distribution of allelic effects in a natural population? Are all alleles the same? Are all alleles distinct? While testing for AI is common testing AI differences between conditions has been more heuristic. Tests of allelic effect are generally performed by crossing individuals and comparing expression between alleles directly in the F1. However, a crossing scheme that compares alleles pairwise is a prohibitive cost for more than a handful of alleles as the number of crosses is at least (**n2-n)/2** where **n** is the number of alleles. A testcross design followed by a hypothesis test of AI between testcrosses can be used to infer differences between nontester alleles, allowing **n** alleles to be compared with **n** crosses. Using a mouse data set where both testcrosses and direct F1 comparisons have been performed, we show that the predicted differences between nontester alleles are validated at levels of over 90% when a parent-of-origin effect is present and of 60%−80% overall. Power considerations for a testcross, are similar to those in a reciprocal cross. Once a minimum level of coverage is present power depends on the number of replicates per cross more than on increased coverage.

SYMP24-2

Pervasive GxE interactions shape the exploration of the phenotypic landspace: lessons from Saclay divergent selection experiments on maize flowering time.

Arnaud Desbiez-Piat[1], Adrienne Ressayre[2], Élodie Marchadier[1], Maud I Tenaillon[3], Christine Dillmann[1]
[1]Université Paris-Saclay, Gif-Sur-Yvette, France. [2]INRAE, Gif-Sur-Yvette, France. [3]CNRS, Gif-Sur-Yvette, France

Abstract

Multidimensional phenotypic shifts arise from drift, selection, and correlative pleiotropic responses. However, understanding the adaptive dynamics that generates such shifts in complex traits is challenging, not only because heritable variation originates from many loci of small effects, but also because these traits are under strong environmental influence. Indeed, GxE interactions can change the fitness effect size of mutations from parents to offspring that experience different environment. To explore the determinants of phenotypic responses to selection and its interplay with environment, we focused on two artificial experiments where divergent selection for flowering time was applied for 20 years generations in field conditions. We conducted common garden experiments across two years to collect data from selected genotypes across all generations, for three traits (flowering time, plant height and leaf length). Using a derivative of the animal model, we distinguished two successive adaptive phases: first the selection of beneficial standing mutations, and second the fixation of beneficial de novo mutations. Moreover, we found significant GxE interactions in all populations and traits. The broadness of the exploration of the phenotypic space changed from one environment to another so that we observed a relative decoupling of traits known to be correlated at the species level. Using association mapping we showed that selected de novo mutations were highly pleiotropic and associated with the three traits, contrary to initially standing variants displaying rare associations. Our results raise the question of the detection of adaptive variants in association mapping settings where kinship is fully controlled.

SYMP24-3

A natural transposable element affects gene regulation and fitness-related traits depending on the developmental stage and the environmental conditions in *D. melanogaster*

Miriam Merenciano, Josefa González
Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain

Abstract

Transposable elements (TEs) are repetitive DNA sequences with the ability to move along the genome. TEs have been considered a genome-wide source of regulatory elements capable of regulating nearby gene expression. In *Drosophila melanogaster*, the *FBti0019985* natural TE insertion has been previously reported to add a transcription start site to the *Lime* transcription factor. In this work, we performed *in vivo* enhancer assays and gene expression analysis with CRISPR/Cas9 mutants and natural populations to explore the effects of *FBti0019985* on *Lime* expression under different stress conditions and different developmental stages. We found that this insertion acts as an enhancer in the adult stage under immune-stress conditions. Indeed, the deletion of predicted immune-related binding sites in the TE significantly reduces its enhancer activity in infected conditions, confirming that it harbours functional cis-regulatory elements. We also found that the TE up-regulates *Lime* in embryos, however, in this case we could not pinpoint the molecular mechanism. On the other hand, in adult stages or under cold- and ethanol-stress conditions, *FBti0019985* is associated with *Lime* down-regulation most likely because it affects the spacing between regulatory motifs in the gene promoter region. Finally, we found that TE-induced *Lime* up-regulation was associated with tolerance to bacterial infection and with increased egg-to-adult viability probably due to increased glucose release. Our results suggest that different developmental stages and environmental conditions should be tested in order to fully characterise the molecular and functional effects of a genetic variant.

SYMP24-4

Fluctuating environments result in variable trajectories of copy number variant evolution

Farah Abdul-Rahman, Angela Hickey, David Gresham
New York University, NYC, New York, USA

Abstract

Copy number variants (CNVs) are a class of mutation with large fitness effects in which a genomic locus varies in repeat number. Genes that undergo replication-dependent alterations in copy number offer a mechanism in which a population can rapidly adapt to environmental change. Despite the prevalence of fluctuating selection in natural environments, it is not well understood how it influences the dynamics and outcomes of CNV evolution in microbial populations.

Previously, we have found that in budding yeast, genes encoding a glutamine transporter (GAP1) and proline transporter (PUT4) undergo gene amplification in chemostats limited for their respective nitrogen source. Here, we developed a dual-fluorescence CNV reporter system to track gene amplification and deletions at GAP1 and PUT4 simultaneously. We evolved budding yeast populations in chemostats over hundreds of generations alternating between glutamine and proline-limitation. We found that static conditions selected for gene amplification of the respective nitrogen transporter with highly repeatable dynamics. Increases in transporter allele frequency were driven by CNV-specialists, individuals with amplifications of either GAP1 or PUT4. We also observed increases in GAP1 deletion alleles under proline-limitation, suggesting the presence of a fitness-trade off. Fluctuating environments resulted in greater variation between replicate populations in which observed dynamics were either driven by CNV-specialists which oscillated in frequency or by CNV-generalists, individuals with gene amplifications of both GAP1 and PUT4, which steadily rose to high frequencies.

Our study highlights the impact of variable selection on the evolution of CNVs which have condition-dependent fitness trade-offs resulting in complex evolutionary dynamics.

SYMP24-5

Using polygenic scores to detect gene-environment interactions associated with body mass index

Pelin Poyraz, Andrew Marderstein, Andrew G Clark
Department of Computational Biology, Cornell University, Ithaca, New York, USA

Abstract

Many common human diseases and traits are polygenic, with thousands of small-effect causal variants across the genome. While genetic effects are strongly modulated by the environment (GxE) in model organisms, single-locus GxE interactions have been a challenge to detect in humans. On the other hand, polygenic gene-environment (GxE) interactions may play an underappreciated role in the genetic architecture of many common human phenotypes. In our study, we evaluate how lifestyle factors modulate the genome-wide contribution to body mass index (BMI). First, we summarized the BMI effects of genome-wide variants within 337,208 unrelated British European individuals from the UK Biobank using a polygenic score (PGS). By leveraging summary statistics from a non-overlapping, previous GWAS meta-analysis of 339,224 mostly European descent individuals, we constructed a BMI PGS that included 617,515 SNPs from across the genome and explained 6.9% variation in BMI. Finally, we tested for significant interactions between the PGS and environmental factors related to physical activity, sedentary behavior, and alcohol intake frequency, identifying highly significant interactions for all three ($P < 10\text{-}15$). Notably, we find that the effect size of the BMI PGS for individuals with high levels of physical activity is nearly 34% less than individuals with low physical activity. Overall, we demonstrate that the aggregation of genetic effects within polygenic scores is a powerful method for detecting GxE, and that GxE can have an important role in the architecture and prediction of human complex traits such as BMI.

SYMP24-6

Rare Variants and Extremes in Gene Expression in a Population of Indica Rice

Zoe N Lye[1], Adrian E Platts[2], Michael D Purugganan[3]

[1]New York University, New York Univeristy, New york, USA. [2]Michigan State University, Ann Arbor, Michigan, USA. [3]New York University, New York, New York, USA

Abstract

The role of rare genetic variation in phenotypes has often been difficult to investigate on a population scale due to the large sample sizes required for conventional methods such as genome wide association studies. Rare genetic variants are thought to be largely deleterious, as purifying selection keeps them at low frequencies in a population. In human systems where rare genetic variation is best studied, it is associated with disease and phenotypic extremes. This work investigates the role of rare genetic variants in gene expression using a diverse population of indica rice varieties. We leveraged a population-level transcriptome data set representing two environmental conditions, to demonstrate that rare variants are associated with extremes in gene expression under both environmental conditions. We further dissected this pattern, comparing the relationship between rare variants and expression extremes for different classes of gene expression, and also comparing different classes of genetic variants including SNPs, INDELs, and structural variants (SV)s. We find that the association of rare variants with extremes of gene expression is more pronounced in a dry environment compared to the control (wet) environment. This suggests that rare genetic variation may act as a source of cryptic genetic variation. Additionally, we calculate selection coefficients for each gene and show that genes with negative directional coefficients have a stronger association compared to those under positive directional selection. These analyses are the first of their kind in rice, and together they provide important clues into how rare genetic variation contributes to patterns of gene expression.

SYMP24-7

Latent phenotypic complexity of adaptation in a single environment

Grant Kinsler[1], Kerry Geiler-Samerotte[2], Dmitri A Petrov[1]
[1]Stanford University, Stanford, CA, USA. [2]Arizona State University, Tempe, AZ, USA

Abstract

Organisms have an amazing ability to adapt to diverse challenges, yet recent observations suggest most mutations affect many traits and most of those effects are deleterious. How is adaptation possible if mutants that influence one trait in a beneficial way influence many other traits in a deleterious way? Evolution experiments with DNA barcodes to track millions of independent evolving lineages have quantified and identified the spectrum of unique single mutations that can each help yeast adapt to glucose limitation. Using recent developments in DNA barcoding, we precisely measure the fitness of a collection of hundreds of these adaptive mutants in many subtly different environments. These data allow us to understand how these mutants interact with the environment, identify the number of fitness-relevant phenotypes affected by these mutants, and uncover the genotype-phenotype-fitness map. We find that while individual genetic mutations do affect many phenotypes (pleiotropy), only a small number of these traits matter for adaptation to the original evolution condition. This finding suggests that adaptation can proceed despite widespread pleiotropy because not all phenotypes affected by adaptive mutations have substantial fitness effects in the current environment. However, we are able to show that these phenotypes often do have strong fitness effects in other environments. In this way we show that adaptation in a single condition can generate a set of mutations that appear very similar phenotypically locally, yet quite phenotypically diverse globally - the pattern we term latent phenotypic complexity.

SYMP26-1

Exploiting multi-way identity-by-descent to detect pedigree relationships, cluster relatives, and infer long-range phase

Ying Qiao[1], Jens G Sannerud[1], Sayantani Basu-Roy[1], Caroline Hayward[2], Amy L Williams[1]
[1]Department of Computational Biology, Cornell University, Ithaca, New York, USA. [2]MRC Human Genetics Unit, MRC Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Identity-by-descent (IBD) segments have become a nearly ubiquitous analysis tool in human genetics, with ever increasing utility as sample sizes increase. In their basic form, IBD segments uncover relatedness between two individuals, yet more elaborate inferences arise when multiple individuals are jointly related. CREST is one such inference method that uses data from a second degree relative pair and their mutual relatives to infer the relationship type of the second degree pair---half-sibling (HS), avuncular, or grandparent-grandchild (GP). The signal CREST exploits is that, with some assumptions, fewer meioses separate genetically older individuals from the mutual relatives, so, e.g., an uncle is expected to have twice as much IBD sharing to a mutual relative as his niece. CREST also detects whether HS or GP pairs are paternally or maternally related, leveraging the pronounced differences between male and female genetic maps. A second multi-way IBD method seeks to identify two clusters of genetic relatives corresponding to the relatives of a focal individual F's two parents. One feature this method uses is that individuals that share IBD to F at a locus should themselves also share IBD if they are related to F through the same parent; conversely, such relatives will not have IBD at the locus if they relate to F through distinct parents. As an important application, the IBD segments from these clusters allow the inference of long-range phasing, with the promise of ultimately assigning chromosome-wide phase with high fidelity in large outbred samples.

SYMP26-2

Impact of reduced local population size on runs of homozygosity in the Florida Scrub-Jay

Tram N Nguyen[1,2], Elissa J Cosgrove[1], Felix Beaudry[3], Nancy Chen[3], John W Fitzpatrick[2,1], Andrew G Clark[1]

[1]Cornell University, Ithaca, NY, USA. [2]Cornell Lab of Ornithology, Ithaca, NY, USA. [3]University of Rochester, Rochester, NY, USA

Abstract

Assuming neutrality, patterns of runs of homozygosity (ROH) convey a rich record of the recent past demography of a population. The existence of recessive deleterious variation can result in selection against long and numerous ROH, a phenomenon generally called inbreeding depression. With an accurate high-resolution linkage map, it is possible to assess the fit of the distribution of ROH to neutral demographic models. But even in the absence of quality background information, the contrast between ROH of related populations can be highly informative. Here, we leverage a unique opportunity to compare the relative abundance and lengths of ROH across four populations of Florida Scrub-Jays that have faced contrasting demographic trajectories. We use high-coverage whole-genome sequences of 225 individuals to compare the abundance and lengths of ROHs in a natural case and control scenario using two large, stable populations versus two small, declining populations. Departure from neutrality is seen in greater than expected heterogeneity along the genome in regions with both excess and a deficit of ROH, consistent with inhomogeneity in the fitness consequences of homozygosity. Populations with recently reduced population size, driven by habitat loss, show inflated size of ROH regions, as well as greater coverage of the genome by ROH. Our results underscore the importance of considering heterogeneity across populations when informing conservation decisions on the species level in the wild.

SYMP26-3

Detecting ROH and IBD in low coverage ancient DNA - New insights into human demography from haplotype sharing

Harald Ringbauer[1], David Reich[2], John Novembre[3], Matthias Steinrücken[3]
[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [2]Harvard Medical School, Boston, USA. [3]University of Chicago, Chicago, USA

Abstract

Here we present newly developed methods to infer long shared haplotypes in human ancient DNA data. To identify shared haplotypes in ancient individuals with low coverage, we leverage haplotype structure from a contemporary reference panel using a modified haplotype copying model. Simulations and down-sampling data of high quality show that our methods can reliably infer ROH and IBD in the low coverage regime (0.5-1x average depth) and tolerate high error rates (up to 5%) typical of human ancient DNA data.

We screened thousands of ancient humans with these new methods for both ROH and IBD. First, by calling ROH, we inferred surprisingly low rates (<5%) of first cousin or closer unions across most ancient populations. Moreover, our ROH results evidence a very substantial impact of the adoption of agricultural lifestyles on background relatedness. Second, the genomic signposts of recent relatedness (IBD) allowed us to robustly identify relatives up to 6th degree and to estimate the fraction of more distant genetic cousins beyond that. Our results establish fine-scale links between ancient cultures and generate new insights into past mobility and population sizes.

SYMP26-4

Nationwide biobank in Mexico unravels demographic history and complex trait architecture from 6,000 genomes

Mashaal Sohail[1,2,3], Amanda Y Chong[4], Maria J Palma[1], Consuelo D Quinto-Cortes[1], Aaron Ragsdale[1], Santiago G Medina[1], Andres Jimenez-Kaufmann[1], Carmina Barberena[1], Paulina Nunez[1], Guadalupe Delgado-Sánchez[5], Luis Pablo Cruz-Hervert[5,6], Leticia Ferreyra-Reyes[5], Hortensia Moreno-Macías[7], Carlos A Aguilar-Salinas[8], Adrian Cortes[9], Selene L Fernandez-Valverde[1], Victor Acuña-Alonzo[10], Genevieve Wojcik[11], Christopher R Gignoux[12], Carlos D Bustamante[13], Adrian Hill[4,14], Maria T Tusie-Luna[7], Alexander J Mentzer[4,9], John Novembre[2,15], Lourdes García-García[5], Andrés Moreno-Estrada[1]
[1]Laboratorio Nacional de Genómica para la Biodiversidad (UGA-LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico. [2]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [3]Centro de Ciencias Genómicas (CCG), Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, Mexico. [4]The Wellcome Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom. [5]Instituto Nacional de Salud Pública (INSP), Cuernavaca, Morelos, Mexico. [6]Facultad de Odontología, Universidad Nacional Autónoma de México (UNAM), México City, México City, Mexico. [7]Unidad de Biología Molecular y Medicina Genómica, UNAM-INCMNSZ and Universidad Autónoma Metropolitana, México City, México City, Mexico. [8]Division de Nutrición, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, México City, Mexico City, Mexico. [9]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, Oxfordshire, United Kingdom. [10]Escuela Nacional de Antropología e Historia (ENAH), México City, México City, Mexico. [11]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [12]Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA. [13]Department of Biomedical Data Science, Stanford University, Stanford, California, USA. [14]The Jenner Institute, University of Oxford, Oxford, Oxfordshire, United Kingdom. [15]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA

Abstract

We present the Mexican Biobank Project, genotyping 6,059 individuals from 32 states at ~2 million SNPs, with linked complex trait and environmental information. We call genetic ancestries and identity-by-descent tracts genome-wide to estimate ancestry-specific effective population size (Ne) for the last 300 generations, revealing elaborate fine-scale structure reflecting the different cultural histories of Mesoamerican regions. While the Gulf, Occident, Mayan region, and Oaxaca reach their peak Ne in the classical period, followed by a slow decline in the post-classical, the center of Mexico shows an increase in Ne in the post-classical period and into half of the colonial period. Later in the colonial and post-colonial periods, while the center decreases in Ne, the Mayan region and Oaxaca increase. European and African ancestries also present varying dynamics and strengths of founder effects. We further call runs of homozygosity (ROH) and find that they are positively correlated with native ancestry. We find that ROH distributions also show fine-scale structure, with a larger genomic imprint of short ROH (< 8 Mb) moving southward and eastward, reflecting smaller long-term Ne due to higher native ancestry, or due to variation in Ne within native ancestry. We model trait values using a set of predictors including genetic ancestry, ROH, socioeconomic and biogeographical covariates, and GRM and state as random effects, finding strong effects of genetic ancestry and/or ROH for height, BMI, triglyceride and glucose levels. Our work threads the causes, consequences, and utility of IBD and ROH using a newly generated genomic resource on under-represented individuals.

SYMP26-5

Haplotype-based recent ancestry sharing methods reveal fine-scale genetic structure in West and Central Africa

Nancy Bird[1], Paschal Awah[2], Neil Bradman[3], Bruce Connell[4], Yves Monino[5], Nana Nketsia[6], Louise Ormond[1], Eno-Abasi Urua[7], David Zeitlyn[8], Mark Thomas[1], Garrett Hellenthal[1]
[1]University College London, London, United Kingdom. [2]University of Yaoundé, Yaoundé, Cameroon. [3]Henry Stewart Group, London, United Kingdom. [4]York University, Toronto, Canada. [5]Centre National de la Recherche Scientifique, Paris, France. [6]Essikado Traditional Council, Essikado, Ghana. [7]Univeristy of Uyo, Uyo, Nigeria. [8]University of Oxford, Oxford, United Kingdom

Abstract

While numerous studies have reported fine-scale genetic sub-structure within countries of (e.g.) Europe, similar dissections of sub-structure in African populations lag behind. Here we present analyses of unpublished genetic data comprising >500,000 polymorphic loci typed in ~1250 individuals from ~100 ethnolinguistic groups from Cameroon, Ghana, Nigeria, and the Republic of the Congo. We illustrate how haplotype-based methods, which infer the proportion of haplotypes for which individuals share a most recent common ancestor, detect a previously unappreciated degree of African population sub-structure that can be missed when analyzing SNPs independently. Examples include a strong correspondence between genetics and geography within Ghana, and significant sub-structure among Cameroonians living within 20km of one another. We further demonstrate that this structure can be inferred at different timescales by analyzing different lengths of identical-by-descent (IBD) segments shared among people, and we compare these patterns to linguistic and cultural records. We also use this haplotype-sharing approach to infer the presence of, and date, events where genetically distinct populations intermixed. By cross-referencing our inference with linguistic and archaeological records, we provide evidence for previously unreported climate-induced migrations occurring more than 3000 years ago, and interactions induced by the Kanem-Bornu (700-1400AD) and Ghana (300-1100AD) empires. Our findings emphasize the importance of denser sampling of west African regions to recover a more refined picture of genetic structure and history, which may have implications for population stratification correction in genotype-phenotype association studies that include people with African ancestry.

SYMP26-6

The impact of identity-by-descent on fitness and disease in natural and domesticated Canid populations

Jazlyn A Mooney[1], Abigail Yohannes[2], Kirk E Lohmueller[2]
[1]Stanford, Stanford, CA, USA. [2]UCLA, Los Angeles, CA, USA

Abstract

Domestic dogs have experienced population bottlenecks, recent inbreeding, and strong artificial selection. These processes have simplified the genetic architecture of complex traits, allowed deleterious variation to persist, and increased both identity-by-descent (IBD) segments and runs of homozygosity (ROH). As such, dogs provide an excellent model for examining how these evolutionary processes influence disease. We assembled a dataset containing 4,414 breed dogs, 327 village dogs, and 380 wolves genotyped at 117,288 markers and phenotype data for clinical and morphological phenotypes. Breed dogs have an enrichment of IBD and ROH, relative to both village dogs and wolves and we use these patterns to show that breed dogs have experienced differing severities of bottlenecks in their recent past. We then found that ROH burden is associated with phenotypes in breed dogs, such as lymphoma. We next test the prediction that breeds with greater ROH have more disease alleles reported in Online Mendelian Inheritance in Animals (OMIA). Surprisingly, the number of causal variants identified correlates with the popularity of that breed rather than the ROH or IBD burden, suggesting an ascertainment bias in OMIA. Lastly, we use the distribution of ROH across the genome to identify genes with depletions of ROH as potential hotspots for inbreeding depression and find multiple exons where ROH are never observed. Our results suggest that inbreeding has played a large role in shaping genetic and phenotypic variation in dogs, and that there remains an excess of understudied breeds that can reveal new disease-causing variation.

SYMP26-7

Intermittent recombination generates novelty in a small, sexually reproducing parasite population

Angela M Early[1,2], Manuela Carrasquilla[2], Aimee R Taylor[2], Angélica Knudson[3], Diego F Echeverry[4,5], Timothy J.C. Anderson[6], Sónia Gonçalves[7], Dominic P Kwiatkowski[7], Caroline O Buckee[2], Julian C Rayner[8], Fabián E Sáenz[9], Daniel E Neafsey[2], Vladimir Corredor[3]
[1]Broad Institute, Cambridge, MA, USA. [2]Harvard T.H.Chan School of Public Health, Boston, MA, USA. [3]Universidad Nacional de Colombia, Bogota, Colombia. [4]Universidad del Valle, Cali, Colombia. [5]Centro Internacional de Entrenamiento e Investigaciones Médicas (CIDEIM), Cali, Colombia. [6]Texas Biomedical Research Institute, San Antonio, TX, USA. [7]Wellcome Sanger Institute, Hinxton, United Kingdom. [8]University of Cambridge, Cambridge, United Kingdom. [9]Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Abstract

Selection is predicted to be most efficacious when effective population size is large, yet in the sexual, eukaryotic malaria parasite *Plasmodium falciparum*, novel drug resistance genotypes have repeatedly emerged first in small populations. This raises questions about how selection dynamics operate in regions of low malaria transmission where effective population sizes are small and population-level recombination rate is low. Here, we report results from a longitudinal (1999-2016) identity-by-descent (IBD) analysis of 526 parasites from the Pacific Coast Region of Colombia and Ecuador. Modern-day genomes (2013-2016) display high background relatedness (median pairwise IBD > 0.26) and low haplotypic diversity with only 17 unique genomic backgrounds represented by 186 parasites. While this confounds or underpowers many analyses, IBD analysis proved powerful. First, intra-chromosomal patterns of IBD differentiate signatures of hard and soft sweeps, mirroring patterns we found with haplotype-based tests (e.g. H_{12}) in more diverse South American populations. Second, IBD segments mark recent recombination breakpoints, the first such analysis possible in a natural population of *P. falciparum.* Third, temporal and spatial analysis of IBD segments show instances of loss and introgression around known drug resistance alleles. Finally, while most *Plasmodium* populations are too large and too outbred to sample recombinant progeny, IBD segments identify five clear crosses, including a three-generation pedigree. These crosses created novel combinations of drug resistance mutations and occurred during periods of heightened case counts. Periodic outbreaks may therefore fuel parasite adaptation in small populations through bursts of outbreeding followed by selection on successful recombinant progeny.

SYMP27-1

Under pressure: Microbial communities in hospitalized patients

Ami S. Bhatt
Departments of Medicine (Hematology & BMT) and Genetics, Stanford University, Stanford, California, USA

Abstract

From human health to the oceanic food chain, microbes are at the base of every major biological system. Far from being passive, these organisms strongly interact with the environment. Yet, for all of this interaction, the dynamics between humans and the microbiome has only been explored for the last twenty years, and even then, most studies have collapsed spectacular strain heterogeneity into monolithic “species”. Positing that “strains matter”, our lab has developed and applied molecular and computational methods that link microbes to specific biological phenomena. By studying patients undergoing hematopoietic cell transplantation, who are typically hospitalized for weeks and are exposed to wide-spectrum antibiotics and chemotherapies, we hope to understand ecological and evolutionary changes in the gut microbiome in clinical time scales. Furthermore, we hope to understand how microbial communication systems shape community structure in these settings. In this session, I will discuss recent, unpublished data on these two topics and describe how our observations are shaping our growing thinking on microbial genomic evolution in short time scales and microbe-microbe and microbe-host communication.

SYMP27-2

Adaptation of a human bacterial pathogen in response to fungal antagonism

Matthew Barber, Caitlin Kowalski
University of Oregon, Eugene, Oregon, USA

Abstract

Bacteria encounter a range selective pressures during host colonization including nutrient limitation, competing microbes, predatory phages, and host immune defenses. Strains of commensal bacteria native to humans produce antibiotics and other toxic effectors that repel invasive pathogens, a process termed colonization resistance. Interactions between pathogenic bacteria and commensal fungi, however, are far less understood. *Staphylococcus aureus* is a particularly versatile bacterial pathogen which passively colonizes 30% of the human population but can also cause severe skin infections, sepsis, and other life-threatening conditions. *S. aureus* skin infections occur most frequently through a cut or abrasion, and the factors that limit *S. aureus* fitness on healthy skin remain unclear. We recently discovered that human commensal fungi possess antimicrobial activity that potently kills a wide range of *S. aureus* clinical isolates. Using an experimental evolution approach, we find that *S. aureus* can readily evolve tolerance to this fungal antagonist while simultaneously altering resistance to conventional antibiotics. Our observations from laboratory evolution are consistent with previous studies suggesting that particular strains of commensal fungi effectively limit *S. aureus* colonization on human skin. Together this work identifies new facets of inter-kingdom microbial conflict as well as revealing mechanisms of rapid adaptation in a major human pathogen.

SYMP27-3

Population genomics of *Saccharomyces cerevisiae* in a continental island reveal abundant wild lineages undergoing independent evolutionary path to its mainland counterparts

Tracy Jiaye Lee, Yu-Ching Liu, Wei-An Liu, Yu-Fei Lin, Hsin-Han Lee, Chia-Lun Hsieh, Kuo-Fang Chung, Huei-Mien Ke, Isheng Jason Tsai
Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

Abstract

Saccharomyces cerevisiae, a budding yeast, is one of the most important model organisms in genetics, cellular and molecular biology. While research interests have previously focused on the evolution of multiple domestication, recent phylogenomic studies have started to address questions regarding the origin of “wild” populations and their natural life histories. Wild *S. cerevisiae* populations can be found on a variety of habitats and exhibit high genetic diversity and clear population structure, providing an excellent model for studying its natural diversity and ecology. To fully capture the abundance, distribution and diversity of natural *S. cerevisiae*, we collected a total of 2,428 samples which resulted in 105 isolates that were subsequently sequenced. We constructed a global phylogeny comprised of 340 isolates including previously published genomes from both domestic and natural populations. In addition to the discovery of multiple natural populations in Taiwan, we also found cases where Taiwanese and Chinese isolates formed monophyletic groups, and among them the TW1+CHN-IX represents the most divergent lineage to date. Moreover, even though natural populations tend to show strong population structure and lack of genetic admixture, we identified mosaic isolates, which provided an opportunity to investigate genetic admixture on various substrates for a mostly asexual species. Surprisingly, diversity estimates in Taiwanese populations were found to be comparable to populations on a continent scale. These results will assist us in quantifying its natural diversities, and understanding the dispersal and demographic events of the wild populations.

SYMP27-4

Multiple Levels of Genetic Drift Constrain the Evolutionary Rate of Prokaryotic Plasmids

Ana Garoña, Nils F. Hülter, Devani Romero Picazo, Tal Dagan
Institute of General Microbiology, Kiel University, Kiel, Germany

Abstract

Extrachromosomal genetic elements are ubiquitous in nature and have been recognized as drivers of prokaryotic evolution. They can be present in multiple copies inside the host cell leading to an increased mutational supply that favours the emergence of intracellular genetic diversity. Plasmids are an undeniable source of evolutionary novelty for microbes, commonly thought to be rapidly evolving genetic entities. Nonetheless, plasmid intracellular genetic diversity is subjected to genetic drift upon cell division; the effect of intracellular processes on plasmid evolution remains understudied. Here we show that genetic drift at the host cell level is an important determinant of plasmid allele dynamics. We developed a genetic system to trace the dynamics of single plasmid alleles at the resolution of individual cells in evolving populations. By focusing on vertical inheritance, we studied the genetic drift that accompanies cell division, termed segregational drift. We found that segregational drift leads to rapid loss of novel plasmid alleles in comparison to chromosomal alleles under non-selective conditions. Furthermore, we show that the mode of plasmid inheritance is an important determinant of plasmid evolutionary rate. Our study provides empirical evidence for multilevel drift in plasmid evolution, which likely constrains evolutionary rates of plasmid genomes.

SYMP27-5

Bacterial adaptation is constrained in complex communities

Thomas Scheuerl[1,2,3], Meiron Hopkins[1], Reuben Nowell[4], Damian Rivett[5], Tim Barraclough[4], Tom Bell[1]
[1]Imperial College London, Ascot, United Kingdom. [2]University of Cambridge, Cambridge, United Kingdom. [3]University of Innsbruck, Mondsee, Austria. [4]University of Oxford, Oxford, United Kingdom. [5]Manchester Metropolitan University, Manchester, United Kingdom

Abstract

A major unresolved question is how bacteria living in complex communities respond to environmental changes. In communities, biotic interactions may either facilitate or constrain evolution depending on whether the interactions expand or contract the range of ecological opportunities. A fundamental challenge is to understand how the surrounding biotic community modifies evolutionary trajectories as species adapt to novel environmental conditions. Here we show that community context can dramatically alter evolutionary dynamics using a novel approach that 'cages' individual focal strains within complex communities. We find that evolution of focal bacterial strains depends on properties both of the focal strain and of the surrounding community. In particular, there is a stronger evolutionary response in low-diversity communities, and when the focal species have a larger genome and are initially poorly adapted. We see how community context affects resource usage and detect genetic changes involved in carbon metabolism and inter-specific interaction. The findings demonstrate that adaptation to new environmental conditions should be investigated in the context of interspecific interactions.

SYMP27-6

Natural selection on the level of molecular crowding in bacterial cells

Tin Yau Pang, Martin Lercher
Heinrich Heine University Duesseldorf, Duesseldorf, Germany

Abstract

The dry mass dissolved in bacterial cells, comprising proteins, metabolites, and other molecules, occupies a substantial fraction of the cytosolic volume. While a higher density of catalysts and substrates might boost biochemical turnover rates, it may also slow down diffusion, change the reactions' Gibbs free energies, and reduce the catalytic efficiency of proteins. Hence, there exists an optimal density that maximizes biochemical efficiency and cellular growth; this optimal density depends on the size and shape distributions of the molecules. In nutritionally poor environments, the cytosol of E. coli is largely filled with metabolic enzymes and metabolites; in nutrient-rich environments, the cytosol is instead dominated by the much larger ribosomes and tRNAs. Here, to understand how molecular crowding affects cellular efficiency at the network level, we simulated a simple model cell, whose growth rate depends on the concentrations of the individual molecule species and the total cytosol density. This model accounts systematically for the slowdown of diffusion and the perturbation of Gibbs free energies through crowding. We find that the biochemical efficiency of reactions with larger catalysts and substrates is more sensitive to variations of the cytosol density. The theoretical dependence of the optimal cytosol density on the size distribution of the macromolecules is consistent with experimental observations of a ~10% difference in E. coli density between nutrient-rich and poor environments. We conclude that the observed dependence of cytosol density variation in E. coli is consistent with an optimality principle of cellular efficiency.

SYMP27-7

Interspecies recombination in agricultural *Campylobacter*: Is it the song or the singer?

Evangelos Mourkas[1], Koji Yahara[2], Sion C Bayliss[1], Jessica K Calland[1], Ben Pascoe[1], Samuel K Sheppard[1]

[1]University of Bath, Bath, United Kingdom. [2]National Institute of Infectious Diseases, Tokyo, Japan

Abstract

Evolutionary analyses of bacteria often consider lineages and species that inhabit different niches. For example, within the genus *Campylobacter* there are species that inhabit the gut of different animals and species that inhabit different niches within a single animal. The maintenance of these species as discrete entities depends on barriers to genetic exchange between them. These can be physical - with species inhabiting different niches, or adaptive - implying selection against hybrid lineages, but quantifying the relative importance of these barriers can be challenging. Considering genes, rather than lineages, as units of selection provides a theoretical solution to this. While understanding clonal population structure and phylogenetics remains important, new theoretical approaches consider the genes that underlie the collective functions of a microbiome (songs) rather than the lineages in which they are found (singers). Here, we analysed >600 genomes of multiple *Campylobacter* species isolated from birds, mammals and reptiles. By characterizing interspecies core and accessory genome recombination in isolates from the same and different hosts we quantify the extent to which genes, rather than lineages, inhabit the niche. Specifically, for some species pairs there was ~0.6 times more recombination between cohabiting isolates than host segregated ones. By broadly defining the limits of interspecies recombination and the function of mobile genes, we provide real-world data by which to interrogate influential theories about the levels of the biological hierarchy (genes, lineages, species) at which selection operates to maintain what we know as ‘species’.

SYMP30-1

The mutual illumination of molecular evolution and cancer biology

Jeffrey P Townsend
Yale University, New Haven, CT, USA

Abstract

Cancer progression is a molecular evolutionary process. The tools of molecular evolution—e.g. model-based phylogenetic inference, the detection of signals of selection—can be applied to enlighten cancer evolution. Phylogenetic tools are essential to understanding the evolutionary history of cancer and illuminating past trajectories, and the rules regarding the temporal progression of cancer that we learn from them illuminate both general molecular evolution and the special circumstances of cancer. The convergent selective effects—observed in tumor after tumor in large analysis of large datasets that enable us to “replay the tape” of evolution—provide an ability to predict the evolutionary trajectory of cancer based on heterogeneous underlying mutation rates, selective impacts of mutations, and epistatic interactions. Efforts to prevent or delay cancer will be enormously aided by molecular evolutionary approaches engaging these concepts. I’ll discuss a suite of concepts and tools from molecular evolutionary theory that can inform cancer biology in new and meaningful ways; highlight current challenges to applying these concepts; and propose ways in which incorporating these concepts could identify new therapeutic modes and vulnerabilities in cancer.

SYMP30-2

The Role of Multi-level Genetic Diversity in Cancers

Li Liu[1], Navid Ahmadinejad[2]
[1]Arizona State University, Phoenix, AZ, USA. [2]Illumina Inc., San Diego, CA, USA

Abstract

Background: In ecology, genetic diversity has been associated with population fitness under fluctuating environmental conditions. A tumor is composed of multiple cell populations (i.e., subclones) that cooperate and compete during tumorigenesis and cancer treatment, forming a dynamic ecological system. Tumor mutational burden, as a gross measure of genetic diversity in cancers is predictive of responses to immunotherapies. However, no studies investigate clinical significance of genetic diversity at subclonal levels.

Methods: We analyzed 5,754 tumor exomes from the TCGA project. For each tumor, we estimated genetic diversity in clonal, subclonal, ancestral, and derived cell populations. For each diversity measure, we tested its association with age of diagnosis, tumor stage and patient overall survival adjusted for tumor types and sex. P-values <0.05 indicated significant associations.

Results: In pan-cancer analysis, subclonal counts and entropy were both positively associated with age of diagnosis and negatively associated with patient overall survival. In individual cancer types (colorectal cancers, liver cancers, ovarian cancers and head and neck cancers), the number of driver mutations in derived subclones was associated with patient overall survival. However, the number of driver mutations in ancestral clones showed no significant associations. Furthermore, although tumor mutational burden measures showed significant associations with age at diagnosis in multiple cancer types, such associations disappeared when adjusted for tumor stages.

Discussion: Consistent with expectations from ecological biology, genetic diversity plays important roles in cancer development. However, diversity at the subclonal level instead of at the clonal level contributes to tumor fitness, especially as responses to treatments.

SYMP30-3

Finding the evolutionary roots of cancer cells migrations

Antonia Chroni[1], Lauren Hamilton[1], Tracy Vu[1], Jeffrey Townsend[2], Sudhir Kumar[1]
[1]Temple University, Philadelphia, PA, USA. [2]Yale University, New Haven, CT, USA

Abstract

Metastasis is the result of the ongoing evolutionary progression of cancer cells with metastatic potential. Deciphering the origin and trajectory of cancer cells is key for fundamentally understanding disease progression. Here, we present an analysis of migration trajectories of cancer cells based on clone phylogenies in patients with different types of metastatic cancer. We reconstructed migration histories of cancer cells seeding metastases and identified their potential sources during cancer progression for multiple cohorts. We found that majority of metastases are seeded by solitary clones that come from primary tumors. In many patients, metastases were seeded by primary tumors, but in a similar number of patients, we found metastases to be seeded by clones from other metastases. We also observed metastatic cascades involving multiple tumors and inter-tumor clone exchanges. Our findings are consistent with emerging experimental and clinical data that paint a more complex picture of metastatic migration networks. We suggest that the knowledge of inferring the migration history of cancer cells will be beneficial for identifying mutations, genes, and mutational signatures that modulate the dynamics of metastatic processes.

SYMP30-4

Quantifying and Describing Contributors to Cancer progression and Therapeutic Resistance via Molecular Phylogenetic Analysis

J Nicholas Fisk[1], Katerina Politi[1], Scott Gettinger[2], Alexander Dornburg[3], Stephen G Gaffney[1], Christopher Cross[1], Amandeep R Mahal[2], James Yu[2], Jeffrey P Townsend[1]

[1]Yale University, New Haven, CT, USA. [2]Yale School of Medicine, New Haven, CT, USA. [3]University of North Carolina at Charlotte, Charlotte, NC, USA

Abstract

The acquisition by cancer of resistance to targeted molecular therapies remains a immense clinical challenge to implementing precision medicine in oncology. Mechanisms of cancer theraputic resitance are poorly understood—partly because they must be reconstructed from biopsies whose timing and content are dictated by patient care. Molecular evolutionary techniques, including phylogenetic analysis and calculation of selection intensities, are well suited to offer guiding insights in describing and overcoming therapeutic resistance.

We developed a molecular phylogenetic approach to perform these evolutionary analyses on cancer tumor sequences, enabling precise determination of the genetic variants that are under selection and influence recurrence. We 1) infer ancestral states in tumor lineages to better identify clinical timelines of disease progression in individual patients, 2) trace the evolution of mutational signatures across a tumor phylogeny with superimposed clinical information to examine the shifting exogenous and endogenous contributors to cancer and 3) quantify the selective advantage conferred by somatic variants in a response to treatment, revealing how cancer evades elimination and recurs. To demonstrate this approach, we present results in EGFR-driven lung adenocarcinoma and in clear-cell renal cell carcinoma. In LUAD we reveal the strikingly high effect size for the EGFR T790M resistance mutation and note its consequences regarding theraputic strategies. In ccRCC, it remains unclear small renal masses all constitute viable precursors to large masses, or if there are distinct molecular etiologies associated with distinct evolutionary trajectories meriting distinct therapeutic approaches. To test this hypothesis of linear development, we apply a synergistic machine learning-evolutionary biology approach.

SYMP30-5

Mutational processes in somatic cancer cell populations

Sayaka Miura, Sudhir Kumar
Temple University, Philadelphia, PA, USA

Abstract

Mutational processes in somatic cancer cell populations are constantly changing, leaving their signatures in the accumulated genomic variation in tumors. The inference of mutational signatures from the observed genetic variation enables spatiotemporal tracking of tumor mutational processes that evolve due to cellular environmental changes, mutations, and treatment regimes. Ultimately, mutational patterns illuminate the mechanistic understanding of their evolution in cancer progression. We show that the integration of cancer cell phylogeny with mutational signature deconvolution enables higher-resolution detection of gain and loss of mutational processes within the phylogeny. This approach to analyzing somatic genomic variations in 61 lung cancer patients revealed a high turn-over of mutational processes over time and closely related clonal lineages. Some mutational signatures (e.g., smoking-related) showed a higher propensity to be lost, whereas others (e.g., AID/APOBEC) were gained during lung tumors evolution. In addition to the usefulness of phylogeny-aware approaches to reveal the turn-over of mutational processes, their usefulness in general will be briefly mentioned in other applications, such as reconstructing clone genotypes from bulk sequencing data, imputing missing data and correcting base calls in single-cell sequences, inferring clone phylogenies, and reconstructing cancer migration paths.

SYMP30-6

Adapted Binary State-Dependent Speciation and Extinction Phylodynamic Model Infers Boundary-Driven Growth in Tumors

Maya Lewinsohn[1], Trevor Bedford[2], Alison Feder[3]
[1]University of Washington, Seattle, WA, USA. [2]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [3]University of California Berkeley, Berkeley, CA, USA

Abstract

Spatial properties of tumor growth have profound implications for cancer progression, therapeutic resistance and metastasis, yet how space governs tumor cell division remains an open question. Xenograft and organoid studies suggest that tumors expand preferentially on the periphery (i.e.,“boundary-driven growth”), while sequencing efforts have suggested faster progression in the tumor interior. Boundary-driven growth likely affects the shape of tumor phylogenies and is therefore theoretically observable from multi-region sequencing data. However, phylodynamic methods have been largely under-utilized to infer growth dynamics in clinical tumors. In this study, we show that boundary-driven growth can be well-approximated by a two-state model permitting different growth rates in the tumor edge and center. We then adapt phylodynamic tools for inferring binary state-dependent speciation and extinction (BiSSE) to quantify these heterogeneous rates. We validate this approach on simulated tumors sampled across multiple spatial regions, and demonstrate its ability to quantify spatially-varying diversification rates under a range of growth conditions and sampling strategies. We then apply BiSSE to multi-region sequencing data from MMR-deficient gastro-esophageal cancers and find evidence that these tumors diversify more rapidly near the tumor edge than in the center. As multi-region and single cell sequencing increases in resolution and availability, this approach could interrogate proposed spatial growth dynamics in diverse clinically resected specimens and be extended to test other two-state growth models, e.g. metastasis or driver gene effects. More generally, this approach demonstrates the potential power of these phylodynamic models to quantify tumor evolutionary dynamics.

SYMP30-7

Cancer genome evolution after whole genome duplication

Elle Loughran, Aoife McLysaght
Trinity College Dublin, Dublin, Dublin, Ireland

Abstract

Whole genome duplication is the second most common genetic aberration in cancer after TP53 mutation, occurring in ~30% of primary tumours and most cases of metastatic cancer. Following WGD, the tetraploid cancer cell rapidly loses chromosomes, resulting in a highly aneuploid, genetically diverse and aggressive tumour.

It is not clear what factors govern the outcome of this rediploidisation process in cancer, but analyses of remnants left in the human genome from two ancestral vertebrate WGDs have revealed non-random duplicate gene retention and loss and suggested selection to retain dosage-constrained genes (the dosage balance hypothesis). Here we investigate to what extent the dosage balance hypothesis explains duplicate gene retention and loss after WGD in cancer.

Using copy number profiles of post-WGD tumour samples from The Cancer Genome Atlas, we found that WGD genes retained in duplicate in humans (ohnologs) are not more likely to be retained after cancer WGD, and that known dosage-sensitive genes in the human genome show little evidence of dosage constraint in cancer samples. Dosage balance does not appear to be a major factor in tumour rediploidisation; it remains to be seen whether other factors have a significant influence or whether the process is essentially random.

SYMP32-1

Scalable divide-and-conquer strategies for phylogeny estimation

Tandy Warnow, Paul Zaharias, Minhyuk Park
University of Illinois, Urbana, IL, USA

Abstract

Phylogeny estimation, especially of large datasets, presents enormous computational and statistical challenges. In this talk, I will describe a new type of divide-and-conquer approach where the species set is divided into disjoint subsets, trees are computed on the disjoint sets, and then merged together using auxiliary information (e.g., a distance matrix). This kind of approach has strong mathematical guarantees, including providing running-time advantages and maintaining statistical consistency. In addition, these approaches can also provide improved accuracy over leading methods for phylogenomic species tree estimation, or for gene tree estimation on large sequence datasets exhibiting substantial non-stationarity.

SYMP32-2

Recovering tree models via spectral graph theory

Yariv Aizenbud[1], Ariel Jaffe[1], Meng Wang[1], Amber Hu[1], Noah Amsel[1], Boaz Nadler[2], Joseph T Chang[1], Yuval Kluger[1]

[1]Yale University, Hew Haven, Connecticut, USA. [2]Weizmann institute of Science, Rehovot, Israel

Abstract

In many modern phylogenetic applications, the sequence length and number of taxa pose enormous computational and statistical challenges for existing tree recovery methods. One approach to address these challenges is “divide-and-conquer”, where a tree structure is inferred separately for multiple subsets of terminal nodes.

We present a novel divide-and-conquer method for tree recovery based on tools from spectral graph theory. We show that the tree topology is strongly related to the spectral properties of a fully connected graph, defined over the terminal nodes of the tree. This relation forms the theoretical basis of our method. Comparing our approach to several competing methods, we show that in many settings, our spectral method has stronger theoretical guarantees and works better in practice.

SYMP32-3

Generating species-trees from a set of incomplete overlapping unrooted subtrees and its application in supertree and supermatrix phylogenetic inference

Olga Chernomor[1], Arndt von Haeseler[2]

[1]CIBIV, University of Vienna, Vienna, Austria. [2]CIBIV, University of Vienna & Medical University of Vienna, Vienna, Austria

Abstract

In phylogenomics one infers a species-tree for a group of species using genetic information from multiple genes. Supertree methods typically infer gene trees separately and then summarise them on a single topology. While supermatrix approaches concatenate gene alignments and infer trees from the resulting alignment, typically assuming a partition model.

Both methods are affected by missing data, i.e. when some species do not have sequences for some genes. Identifying, whether incomplete gene trees are compatible, i.e. can be represented on a single tree topology, is computationally hard. Furthermore, depending on the approach, amalgamating incomplete gene trees can lead to different species trees in supertree methods. On the other hand, sparse concatenated alignments in supermatrix approaches might result into huge collections of trees with identical scores, termed phylogenetic terraces.

Here, we present a deterministic algorithm that given a set of incomplete unrooted subtrees identifies, if they are compatible, and, if yes, generates a collection of all corresponding compatible species trees. For supertree, where indeed gene trees are rarely compatible, the algorithm can be modified to identify groups of compatible gene trees. The presented algorithm also tackles the problem of enumerating trees on the same phylogenetic terrace. Hence, it provides means to ascertain existence (supertree) and uniqueness (supertree and supermatrix) of species trees in the presence of missing data.

The algorithm is implemented in IQ-TREE2. We provide an exhaustive evaluation of its feasibility and show, that it can deal with large phylogenomic datasets in reasonable amount of time.

SYMP32-4

Fast likelihood calculation for pandemic-scale data

Nicola De Maio
EMBL-EBI, Hinxton, Cambridge, United Kingdom

Abstract

Recent years have seen a steady increase in the application of genome sequencing technologies to the epidemiological field. This trend is being further emphasised with respect to the COVID-19 pandemic, which has seen an unprecedented response from the sequencing community. Vast genome datasets can provide invaluable information regarding microbial spread and evolution, but leveraging large sequence datasets is computationally demanding, and currently we are unable to investigate large portions of the available SARS-CoV-2 genomic data using classical phylogenetic approaches due to excessive computational demand.

I will present novel strategies to reduce the computational burden of likelihood-based phylogenetic inference. First, I introduce a novel technique for the calculation of phylogenetic likelihoods. The method is an alternative to the classical Felsenstein pruning algorithm tailored for datasets with many closely related sequences, as typical in genomic epidemiology. For example, this includes current SARS-CoV-2 alignments of 100s of 1000s of genomes, with which the new algorithm can be orders of magnitude faster.

I will also present a more concise yet readable way to represent and process sequence data in genomic epidemiology; I will describe a simple way to reduce dataset size in maximum likelihood tree inference by removing sub-informative sequences; and I will discuss efficient ways to estimate substitution rates with such large data.

These new methods have the potential of considerably extending the computational reach of classical maximum likelihood and Bayesian phylogenetic approaches implemented in popular phylogenetic packages.

SYMP32-5

Summary statistics of ranked tree shape distributions

Julia A. Palacios, Samyak Rajanala
Stanford University, Stanford, California, USA

Abstract

Rooted and ranked binary trees are mathematical objects of great importance used to model hierarchical data and evolutionary processes with applications ranging across many fields including evolutionary biology and infectious disease transmission. While Bayesian methods allow exploration of the posterior distribution of trees, assessing uncertainty and summarizing tree distributions remains challenging for these types of structures. Similarly, in many instances, one seeks to summarize samples of trees obtained with different methods, or from different samples and environments, and wishes to assess stability and generalizability of these summaries. In this talk I will present new metrics on the space of ranked tree shapes and ranked genealogies and provide an efficient combinatorial optimization algorithm for estimating Fréchet means and variances. I will show the applicability of our summary statistics for studying popular tree distributions and for studying the evolution of SARS-CoV-2.

SYMP32-6

Reconstructing CRISPR/Cas9-based single-cell phylogenies with Cassiopeia

Matthew G Jones[1,2,3], Alex Khodaverdian[1], Richard Zhang[1], Sebastian Prillo[1], Jonathan Weissman[3], Nir Yosef[1]

[1]UC Berkeley, Berkeley, CA, USA. [2]UCSF, San Francisco, CA, USA. [3]MIT, Boston, MA, USA

Abstract

The pairing of CRISPR/Cas9-based gene editing with massively parallel single-cell readouts now enables large-scale lineage tracing. However, the rapid growth in complexity of data from these assays has outpaced our abilities to accurately infer phylogenetic relationships. To address this, we have recently developed Cassiopeia: an end-to-end pipeline for reconstructing phylogenies from these single-cell lineage tracing assays. We first introduce the Cassiopeia algorithms, which specifically addresses the scale and nuance of single-cell lineage tracing data with perfect phylogeny-based heuristics and Steiner-Tree optimization. We utilize two resources to showcase Cassiopeia's performance across several parameter regimes: first, a flexible simulation framework for evaluating algorithms and exploring lineage tracer design principles. Second, we benchmark algorithmic performance using the most complex experimental dataset generated to date -- consisting of 34,557 human cells continuously traced over approximately 15 generations, 71% of which are uniquely marked. With both resources, we demonstrate that Cassiopeia outperforms a panel of traditional algorithms. Moreover, through this analysis, we exhibit how a user can use the Cassiopeia software suite for simulating trees, implementing new approaches, and benchmarking these new algorithms. Finally, we will end by discussing how these phylogenies can be integrated with other single-cell data modalities to gain new insights into dynamic processes like cancer metastasis. The full Cassiopeia suite is publicly available at https://github.com/YosefLab/Cassiopeia.

SYMP32-7

Polymorphism-Aware Phylogenetic Models and their Applicationsb

Carolin Kosiol
University of St Andrews, St Andrews, Scotland, United Kingdom

Abstract

The increased availability of sequenced genomes both from closely related species and from individuals of the same species, offers a great opportunity to study the speciation and evolutionary history of populations, provided we can properly model the process of sequence evolution using inter and intraspecific data together.

In my group, we have developed a new method called POlymorphisms-aware phylogenetic MOdel (PoMo). 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 in a continuous Markov process. It is a selection-mutation model which separates the mutation process from the fixation process. Thereby, a Moran process is used to model genetic drift. PoMo naturally accounts for incomplete lineage sorting because ancestral populations can be in a polymorphic state.

Our method can accurately and time-efficiently estimate the parameters describing evolutionary patterns for phylogenetic trees of any shape (species trees, population trees, or any combination of those). We have implemented the approach into Maximum Likelihood software package and recently developed a Bayesian approach for molecular dating. I will present what can be learned by applying these methods to genome-wide data sites of great ape populations about ancestral population history of this species. Finally, I will also discuss how the new methods could be applied to populations of fruit flies that have recently been subject to an experimental evolution study for sexual mating system.

SYMP32-8

Fast and accurate bootstrap confidence limits on genome-scale maximum-likelihood phylogenies using little bootstraps

Sudip Sharma[1,2], Sudhir Kumar[1,2]

[1]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA.

[2]Department of Biology, Temple University, Philadelphia, PA, USA

Abstract

As the accessibility of genome sequence databases and the assembly of multi-species and multi-gene alignments containing hundreds of thousands of bases have become widespread, Felsenstein's bootstrap approach is being applied to increasingly larger datasets. These large datasets have the power to reconstruct hard-to-resolve evolutionary relationships with high confidence. But, their bootstrapping imposes increasingly onerous computational demands because the computational complexity of phylogenomic analyses using the maximum likelihood (ML) method increases exponentially with the number of sequences and linearly with sequence length. While many heuristics reduce the burden caused by a large number of sequences, there is a need for more efficient and statistically effective approaches to reduce the onerous computational burden imposed by an increase in sequence length due to the transition to next-generation sequencing methods. Consequently, the classical Felsenstein bootstrap resampling procedure can take a long time for phylogenomic datasets. We will introduce the bag of little bootstraps approach to place confidence limits on ML phylogenies inferred using long multiple sequence alignment. We show that the bootstrapping of a collection of little subsamples, coupled with median bagging of subsample confidence limits, produces accurate bootstrap confidence for phylogenetic relationships in a small fraction of computational time and memory. We suggest that the little bootstraps will enhance rigor, efficiency, and parallelization in big data phylogenomics, even on personal computers.

SYMP33-1

Some curiosities of evolution at linked sites

Brian Charlesworth
Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Abstract

The effects of balancing selection, selective sweeps and background selection on patterns of evolution and variation across the genome are by now familiar. Other aspects of the effects of evolutionary processes on linked sites are less well known, e.g., that the recent fixation of a neutral mutation is accompanied by a reduction in variability. Some examples of such effects will reviewed, with a special emphasis on cases where mutations spreading towards fixation can cause increases rather than reductions in variability. Evidence is presented that such effects may be operating in genomic regions with low levels of recombination and in very small populations, where the products of effective population size and selection coefficients can often be close to one in magnitude.

SYMP33-2

Genomic consequences of range expansion and gene surfing

Rémi Matthey-Doret[1,2], Flávia Schlichta[1,2], Antoine Moinet[1,2,3], Laurent Excoffier[1,2]
[1]CMPG, Institute of Ecology and Evolution, University of Berne, Berne, Berne, Switzerland. [2]Swiss Institute of Ecology and Evolution, Lausanne, Vaud, Switzerland. [3]Interfaculty Bioinformatics Unit, University of Bern, Berne, Berne, Switzerland

Abstract

Range expansions occurred repeatedly in the history of most species and previous work has shown that they leave distinct genetic footprints. Notably, neutral and even deleterious variants happening on expansion fronts can spread and even fix in newly colonized territories, a phenomenon known as “gene surfing”. The rapid frequency increase of such variants resembles the effect of positive selection (i.e., selective sweeps), but the effects of gene surfing on linked neutral diversity are still not very well characterized. To better understand the genomic distribution and impact of gene surfing events, we simulated range expansions with two forward-in-time genetic simulators (SLiM and SimBit), and studied the spatial and temporal change in genome-wide diversity, focusing on the appearance and dynamics of diversity troughs, based on nucleotide diversity or coancestry. We explore the effects of various demographic parameters (migration rates, founder effects and growth rate), background selection and selective sweeps on trough density, prevalence and depth. We show that migration considerably delays loss of genetic diversity, decreasing the proportion of the genome within troughs, while simultaneously increasing their number. We observe that individual trough size is inversely proportional to the composite parameter N×r, where N is the effective population size and r is the recombination rate. During range expansions, troughs can progressively dominate the whole genomic landscape, and be tens of megabases wide. As such large troughs are often interpreted as the hallmark of selective sweeps, we advise caution when trying to identify positive selection based on patterns of genetic diversity.

SYMP33-3

Modeling neutral genetic variation linked to loci under selection in a population with variable size

Eric Friedlander, Matthias Steinruecken
University of Chicago, Chicago, IL, USA

Abstract

Natural selection on beneficial or deleterious alleles results in an increase or decrease, respectively, of their frequency within the population. Due to chromosomal linkage, these dynamics affect genetic variation in nearby neutral regions. Accurately modeling these dynamics is crucial to understanding how selection and demographic processes impact patterns of genomic variation. This is particularly important as changes in population size can yield patterns that mimic the effects of selection. We use the Wright-Fisher diffusion, a mathematical framework describing the evolution of haplotype frequencies, to study the impact of selection on linked neutral variation. In general, explicit solutions are not known for the dynamics of this diffusion when selection and recombination act simultaneously.

Thus, we present a method for numerically evaluating the Wright-Fisher diffusion dynamics at two linked loci, under selection. Arbitrary population size histories can be explicitly accounted for in this approach. A key step in the method is to express the moments of the associated transition density as solutions to ordinary differential equations. Numerically solving these ODEs relies on an accurate and numerically efficient technique to estimate higher order moments from those of lower order. This approach can also be applied to the general problem of estimating allele frequency spectra for large samples from smaller samples. We demonstrate how this numerical framework can be used to elucidate patterns of heterozygosity, linkage disequilibrium, and distortions in the site-frequency-spectra for regions linked to loci under selection in various demographic settings.

SYMP33-4

Why does genetic diversity vary so little across the human genome?

Adam Eyre-Walker, Vivak Soni
University of Sussex, Brighton, East Sussex, United Kingdom

Abstract

Genetic diversity is known to vary across the genomes of many species. For neutral diversity this variation is due variation in the rate of mutation, the genealogy length and the effective population size, with variation in the effective population size being due to linked selection. But to what extent do these two factors contribute to the variation in diversity. To investigate this we divided the human genome up into non-overlapping 10KB windows, and quantified the variation in SNP density, and the variation in the mutation rate using de novo mutation data. Surprisingly we find that there is more variation in the mutation rate than variation in diversity. We explore a number of explanations for this, and conclude that it is most likely due a negative relationship between the effects of linked selection and the mutation rate; this is not expected since the power of linked selection can depend on the mutation rate. However, our models suggest that linked selection is extremely prevalent in the human genome, reducing diversity by more than 40% on average.

SYMP33-5

Mutation rate variation shapes the distribution of diversity along Drosophila chromosomes

Gustavo V. Barroso[1], Julien Y. Dutheil[2]

[1]University of California, Los Angeles, Los Angeles, California, USA. [2]Max Planck Institute for Evolutionary Biology, Ploen, Schleswig-Holstein, Germany

Abstract

What shapes the distribution of nucleotide diversity along the genome? Attempts to answer this question have sparked debate about the roles of neutral stochastic processes and natural selection in molecular evolution. However, integrative models that simultaneously consider multiple factors are lacking; without them, confounding factors lurk in the estimates. Here we present iSMC – a new statistical method that jointly infers the genomic landscapes of genealogies, recombination rates and mutation rates. In doing so, iSMC captures the effects of demography, linked selection and local mutation rates on patterns of genomic variation. We then use linear modelling to estimate their individual contributions to levels of nucleotide diversity. Our analyses reveal the signature of selection in Drosophila melanogaster, but the mutation landscape explains the majority of the genome-wide distribution of diversity in this species. Furthermore, our simulation study suggests that in many scenarios the mutation landscape will be the most relevant factor shaping diversity, overcoming the effects of drift and selection. Since the local density of SNPs is a crucial statistic in the inference of selection and introgression, we propose that incorporating mutation rate variation into the null model of molecular evolution should lead to more realistic inference in population genomics.

SYMP33-6

Bottlenecks and range expansion can induce neutral profiles of nucleotide diversity that mimic hard selective sweep signals

Antoine Moinet[1], Stephan Peischl[2], Laurent Excoffier[1]
[1]IEE, Bern, Switzerland. [2]IBU, Bern, Switzerland

Abstract

Since the introduction of the neutral theory of evolution by Kimura in 1968, the relative importance of genetic drift and natural selection in shaping the genetic diversity of living organisms has been hotly debated. It is generally accepted that a strong reduction in diversity around a specific locus indicates the recent rapid fixation of a positively selected allele at this locus, a phenomenon called a selective sweep. Here, we demonstrate that such rapid fixations, although very unlikely to happen in a neutrally evolving population of constant size, are possible under specific demographic scenarios where populations experience changes in population size, e.g. bottlenecks or range expansions. These demographic processes lead to signals of nucleotide diversity very similar to signals of selective sweeps. We quantitatively investigate the shape of troughs in neutral genetic diversity and compare these to signals of selective sweep. We empirically test our framework in a particular case of a putative selective sweep signal around the gene Quetzalcoatl in D. melanogaster, and show that both the trough signal in the vicinity of the gene and the genome-wide profile of diversity observed in the data are compatible with a short bottleneck scenario without selection. Our findings show that valleys of genetic diversity may falsely be attributed to positive selection, when it is equally possible that demographic scenarios can lead to identical signatures in the genome.

SYMP33-7

The genetic diversity reduction due to background selection increases the disease prevalence under the liability threshold model

Xinyi Li[1], Jeremy Berg[2], John Novembre[2,3]

[1]Committee on Genetics, Genomics & Systems Biology, University of Chicago, Chicago, IL, USA. [2]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

Abstract

While the importance of background selection in shaping patterns of neutral genetic diversity is well-studied, its influence on the genetic architecture and the prevalence of complex disease is not well understood.

To address this, we extended the liability threshold model to include background selection. In the model, mutational pressure increases the liability while selection acts to reduce it, and the equilibrium disease prevalence arises due to a balance between the two forces. We found that when background selection is included, it alters this equilibrium by reducing the genetic variance for liability, which drives an increase in disease prevalence. To validate our theoretical results, we performed SLIM simulations and found scenarios where, for example, when the diversity reduction due to background selection is roughly 20% (as estimated in humans), the disease prevalence increases 10%.

Background selection also distorts the allele frequency spectrum at linked loci, as rare variants are less impacted. To examine whether the distortion of the site frequency spectrum of liability sites changes the disease risks, we used simulation to vary the number of rare variants while controlling the genetic diversity reduction. Under the scenarios we examined, the degree of distortion of the site frequency spectrum due to background selection has no meaningful effect on disease risk.

From these investigations, we conclude that background selection can impact disease prevalence and does so primarily through overall levels of diversity and not its effect on the relative abundance of rare versus common variants.