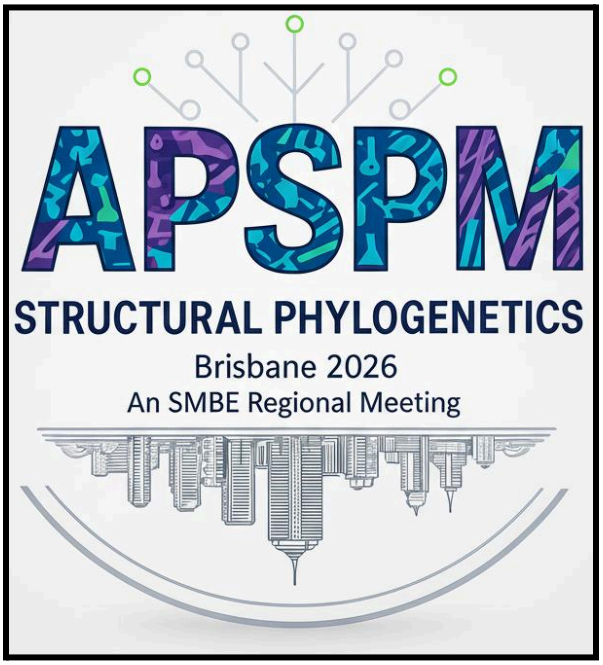
SMBE Australasian Protein Structural Phylogenetics Meeting 2026



Program Booklet

Feb 15 – 18, 2026, University of Queensland, Brisbane, Australia

Acknowledgements

We gratefully thank the Society of Molecular Biology & Evolution (SMBE) for their generous support towards this meeting, as part of the SMBE Regional Meeting program.

We also wish to acknowledge further financial and organisational support from the University of Queensland (notably the BioSig Lab and EDI Management Committee), the Australian National University (Research School of Biology), the University of Auckland (Centre for Computational Evolution), and the Australian Centre for Ecogenomics.

A special thank you to all of the speakers and participants who made this meeting possible.

Lastly, we wish to acknowledge the Traditional Owners and their custodianship of the lands on which we meet. We pay our respects to their ancestors and their descendants, who continue cultural and spiritual connections to the country. We recognise their valuable contributions to Australian and global society.

Background

Since the recent surge of AI-generated protein structure data, structural phylogenetics has rapidly evolved and is poised to reshape our understanding of molecular evolution. While the availability of structural data has exploded, its integration into evolutionary analysis has yet to reach its full potential. This meeting addresses this critical gap, providing a timely forum to explore this frontier.

New methods and tools are moving the field from a niche, computationally expensive discipline to one that is feasible and widely applicable. By bringing together structural biologists, evolutionary biologists, and computational method developers, this meeting aims to foster the cross-disciplinary collaborations needed to establish best practices, drive methodological innovation, and harness the power of protein structure to resolve deep evolutionary histories and understand the diversity of molecular function.

Times and places

All times are in Australian Eastern Standard Time (GMT +10).

Feb 15, 2026. 5:30pm. Welcome Mixer. Venue: Southbank Beer Garden Brisbane.

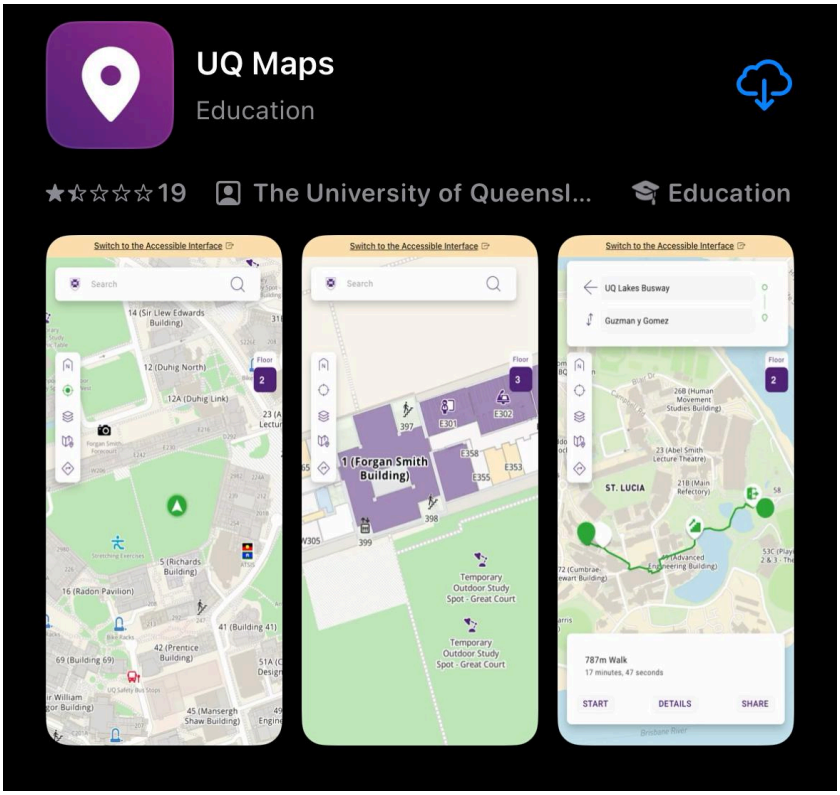
Feb 16 – 18, 2026, 8:30am – approx. 5pm. Presentations and Workshops. Location: Building 50, Room N202, University of Queensland, Brisbane.

Feb 16, 2026, 2:30pm – 3:30pm. Special talk hosted by the University of Queensland. Speaker: Martin Steinegger. Location: Forgan Smith Building 01-E109, University of Queensland, Brisbane.

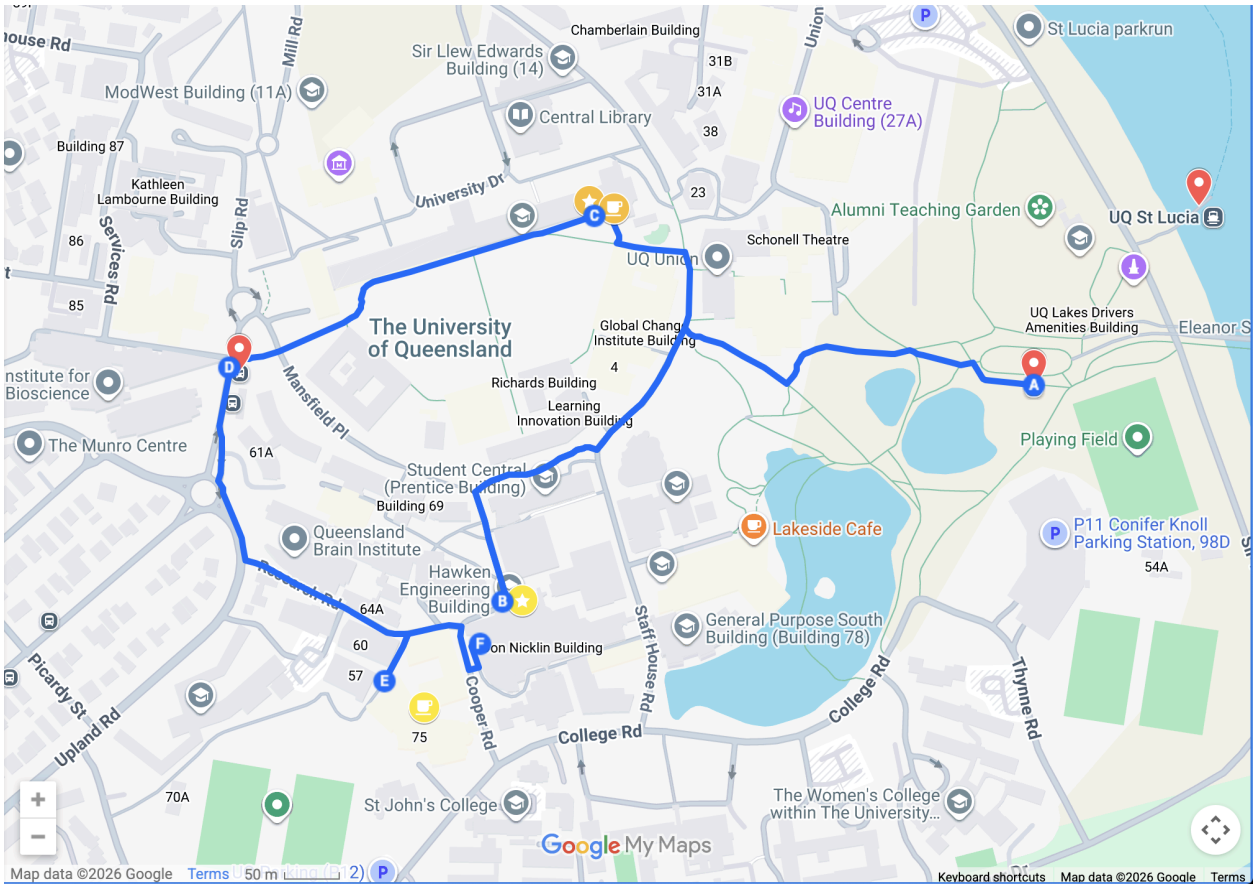
Feb 18, 2026, 6:30pm. Conference dinner. Venue: Patina at Alumni Court, University of Queensland, Brisbane.

Getting here:

- The best way to navigate on campus is to either download the UQ Maps application available on the app store, or use https://maps.uq.edu.au/ and use the left hand panel for search.



- Labels A and D represent the two bus stops. Stop A is a ~15 minute walk from the main venue (Label B) and ~7-10 minute walk from the Biochemistry Lecture venue (Label C).
- Labels B and C represent the two rooms reserved for proceedings. (Building 50, Lecture room N202 for the meeting; Building 1, Lecture room E109 for the Alumni Lecture). Walk time between these two venues is ~10 minutes.
- The two coffee cups indicate cafes closest to the venues.



The two bus stops (Labels A and D) are called Chancellors Place (left) and UQ Lakes (right). Buses arriving at Chancellors Place (left) include 411, 412 from the City and 414, 427,428,432 from the Indooroopilly suburb. Buses arriving at UQ Lakes (right) include M2. From the city a ferry can also be taken.

For more details use Google Maps or Campus maps. The UQ Lakes station should appear as “UQ Lakes station, St Lucia QLD 4067” and the Chancellors Place will appear as “Chancellors Pl, St Lucia QLD 4067”.

If you are coming from the City side, use UQ Lakes. If you are in the western suburbs (relative to the City) use Chancellors Place. All public transport accepts Visa cards (direct tap or via NFC in phone; fare is 50 cents), if you are unsure if your card will work or not, buy single use tickets from ticketing machines available on all hubs. If your visa card works on the machine, it will work inside the bus. The 50 cent deduction will appear in your card statement by midnight.

Bus cards can also be purchased from the same machines that sell single use tickets @ AUD$5 and then topped up. However if you are here for 4 days and will be using the bus to get to and from the venue, it's best to use a visa card directly (or via phone) or go with single use tickets (valid for 90 minutes from time of purchase).

Uber

Uber entry is restricted to the UQ campus from the left hand side of the map as the access bridge on UQ Lakes, connecting the Campus to Dutton Park and onwards to the city is for foot traffic and public buses. So if you use Uber and enter UQ Lakes, you will drive past Chancellors place, so best to use Chancellors place as a drop off point.

Emergency (The UQ Maps application has more details).

While on campus in case of emergencies contact UQ security directly on **(07) 3365 3333**. Emergencies include:

- fire, evacuations
- medical emergencies, first aid and medical transport
- environmental, spills and explosions
- bombs, arson and life threats
- criminal activity, armed offenders
- hold up and duress alarms
- vehicle accidents
- assault, violent behaviour
- people trapped in lifts
- emergency callout points
- power failures.

Presentation Schedule

Monday 16 February, 2026

8:30	Morning coffee	
9:00 – 9:10	*APSPM Opening Comments*	
	Clocks, Characters, and Complexes (Chair: Bostjan Kobe)	
9:10 – 9:40	Plenary Speaker: Nicholas Matzke	Protein structure characters in the light of phylogenetic systematics
9:40 – 10:00	Lindell Bromham	Estimating evolutionary time from protein structure: what do we need to make it work?
10:00 – 10:15	Yogapriya Subramaniyan	Structural Phylogeny of Flagellar FliF with Sporulation and T3SS homologs
10:15 – 10:30	Pietro Ridone	From Ions To Force: Structure and evolvability of the building blocks of ion-powered rotary motors
10:30 – 10:45	Michael Landsberg	Structures of the YenTc holotoxin prepore and pore reveal the evolutionary basis for chitinase incorporation into ABC toxins
10:45	Morning tea (45 minutes)	
	Emerging Phylogenetic Methods (Chair: Nicholas Matzke)	
11:30 – 11:50	David Moi	FoldTree2: Integrative Sequence–Structure Phylogenetics with GNN-Derived Structural Alphabets
11:50 – 12:05	Kevin Downard	Bridging Protein Structure Analysis and Phylogenetics with Mass Spectrometry
12:05 – 12:20	Huaiyan Ren	A new information criterion favors mixture models over partition models and influences partition model selection.
12:20 – 12:40	Scott Schmidler	Incorporating Flexibility and Site-Dependence into Bayesian Structural Phylogenetics
12:40 – 1:00	Remco Bouckaert	PhyloPotts models for phylogenetics with dependent sites
1:00 – 1:15	Clementine Yan	Bayesian Phylogenetic Inference of Protein Structure Evolution Using Angular Diffusion Model
1:15 – 1:30	Jasmine Saghafifar	Exploring Time-Nonreversibility and

		Site-Mixtures in Virus Phylogenetics with ABySS
1:30	Lunch (60 minutes)	
	Special talk hosted by University of Queensland. Forgan Smith Building 01-E109	
2:30 – 3:30	Martin Steinegger	Exploring the Protein Universe with Multimers, Motifs & Interfaces
3:30	Afternoon tea (45 minutes)	
	Virus Evolution (Chair: Caroline Puente-Lelievre)	
4:15 – 4:35	Spyros Lytras	Leveraging structural predictions to understand virus glycoprotein evolution on the macro and micro-scale
4:35 – 4:50	Cinthy Lorena Jimenez Silva	Mycovirome Online Platform (MVoP): A collaborative online tool for mycovirus discovery
4:50 – 5:05	Mihnea Bostina	From Molecular Machines to Evolutionary Blueprints
5:05 – 5:20	Jonathon Mifsud	The Promise and Challenges of Viral Protein Structural Phylogenetics Across Deep Time

Tuesday 17 February, 2026

8:30	Morning coffee	
	Searching the Twilight Zone (Chair: Peter Wills)	
9:00 – 9:20	Claudia Alvarez Carreno	Charting the protein universe: Deep evolutionary signals in sequence-structure space
9:20 – 9:35	Fabian Ruperti	Cross-phyla protein annotation by structural prediction and alignment
9:35 – 9:50	Juan Manuel Trinidad-Barnech	Expanding kinetoplastid genome annotation through protein structure comparison
9:50 – 10:05	Jamiema Sara Philip	Easy come, easier go: mapping the loss of flagellar motility across the tree of life
10:05 – 10:20	Bostjan Kobe	A grammar-based discovery engine for structural relationships in the global TIRome
10:20 – 10:35	Ashar Malik	Structome-DeepRoots: On inferring homology using high-dimensional embeddings
10:35	Morning tea (40 minutes)	
11:15 – 12:30	**Workshop 1: Structural Homology Detection** Instructors**:** Cameron Gilchrist and Sukhwan Park Because of tools like AlphaFold, there are now hundreds of millions of high quality protein structural predictions available online. This workshop will show how our suite of tools, including Foldseek, Foldseek-cluster, Foldseek-multimer, Folddisco and FoldMason, can be used to rapidly search and compare homologous structures from these massive structure databases.	
12:30	Lunch (75 minutes)	
1:45 – 3:00	**Workshop 2: Phylogenetic Inference from Structure** Instructors: Caroline Puente-Lelievre and Jordan Douglas Protein structures are well-conserved over long evolutionary timescales, making them particularly useful for phylogenetics when sequence divergence is high. Using IQ-tree, this workshop will demonstrate how phylogenetic trees can be estimated under the 3Di structural alphabet employed by Foldseek.	
3:00	Afternoon tea (30 minutes)	
3:30 – 5:00	**Workshop 3: Protein Language Models and Phylogenetics** Instructor: Ashar Malik Protein Language Models (pLMs) allow latent space representation of proteins, both sequence and structure. This workshop will explore how this latent space can be used to extract evolutionary signals and put it towards	

	phylogenetic inference. The workshop will introduce pLMs, look at state of the art models like ESM-Cambrian and SA-Prot, show how to extract latent space representations and their respective computational transformation for phylogenetic inference.
6:30	Conference dinner – Patina at Alumni Court (~15 min walk from venue)

Wednesday 18 February, 2026

8:30	Morning coffee	
	LUCA and Beyond (Chair: David Moi)	
9:00 – 9:30	Plenary Speaker: Joanna Masel	What can we learn from ancient (protein) history?
9:30 – 9:45	Natalia Mrnjavac	From rocks to life: the primordial assembly of microbial metabolism
9:45 – 10:05	Charles Carter, Jr.	Beyond Phylogenetic Inference from Structure
10:05 – 10:25	Peter Wills	Genetic coding represented as the first local example of intelligence in the cosmos
10:25 – 10:40	Aleksandar Radakovic	Structural phylogenetics of aminoacyl-tRNA synthetases and the implications for the evolution of the genetic code
10:40	Morning tea (30 minutes)	
	Ancestral Reconstruction (Chair: Matthew Baker)	
11:10 – 11:30	Elizabeth M. J. Gillam	Lost in translation: the sequence verification challenge in studies of protein evolution by ancestral sequence reconstruction
11:30 – 11:45	Jacob Scadden	A Tale of Tails: Investigating the Evolution and Adaptation of the Bacterial Flagellar Filament
11:45 – 12:00	Divyangi Pandit	Hook, Line, and Swimmer: A look into the bacterial flagellar hook and filament capping proteins
12:00 – 12:15	Ulban Adhikary	Evolutionary trajectory and inhibition of ancestral metallo-β-lactamases
12:15 – 12:30	Jordan Douglas	Stochastic mapping and ancestral gapping: reconstructing the evolutionary pathways in protein sequences, structures and indels
12:30	Lunch (70 minutes)	
	Genome Analysis and Functional Prediction (Chair: Cinthy Jimenez-Silva)	
1:40 – 2:00	Evelyne Deplazes	Mosaic Evolution of Zinc Selectivity: Decoupling Chemical Optimization from Structural Constraints in Bacterial Solute-Binding Proteins
2:00 – 2:15	Marco Antonio Chávez Tinoco	More Than One Way to Lose a Gene: Distinct Evolutionary Fates of NOX5 in Rodents

2:15 – 2:30	Hassan Sibroe Abdulla Daanaa	Genome-wide analysis of protein intrinsic disorder in Drosophila melanogaster
2:30 – 2:45	Samuel Ingloff-Richards	Evolution of Antibiotic Resistance and Structure of the RNA Polymerase β Subunit
2:45 – 3:00	Phil Hugenholtz	The Genome Taxonomy Database (GTDB): a potential use case for structural phylogenetics
3:00	Afternoon tea (40 minutes)	
	The Future of Structural Phylogenetics (Chair: Spyros Lytras)	
3:40 – 4:00	Caroline Puente-Lelievre	Protein structure and the beginning of a third era in phylogenetics
4:00 – 5:00	*Discussion session and future directions*	
5:00 – 5:05	*Closing remarks*	

Abstracts

Speakers are sorted alphabetically by first name.

Structural phylogenetics of aminoacyl-tRNA synthetases and the implications for the evolution of the genetic code

Aleksandar Radakovic, University of Chicago, USA

Aminoacyl-tRNA synthetases (aaRS) are essential enzymes that implement the genetic code. Resolving their evolutionary relationships is critical for understanding the evolution of the code itself; however, traditional amino acid sequence-based phylogenetics has been hindered by the low sequence conservation of these ancient proteins, resulting in conflicting topologies with several key branching points lacking robust statistical support. Here, we overcome these limitations by presenting a comprehensive structural phylogeny for both aaRS classes. By leveraging the high conservation of their 3D folds – using AlphaFold for accurate structure prediction and Foldseek for structural alignment – we successfully aligned the catalytic domains of all enzymes in each class without the need for expert manual curation or complex stepwise alignment procedures. Notably, a parallel "one-shot" alignment attempt using sequence-based methods yielded an uninterpretable alignment, highlighting the advantage of the structural approach. Using the structural alignments and a structure-specific substitution matrix within a maximum likelihood (ML) framework, we inferred well-resolved and statistically supported phylogenies with moderate branch lengths for the two aaRS classes. In contrast to the state-of-the-art Bayesian phylogenies of the two classes, our structural ML phylogenies reveal several distinct groupings:

- HisRS groups with the Ser/Thr/ProRS clade.
- AlaRS groups with the Asp/Asn/LysRS clade.
- PylRS and PheRS form a clade with strong support.
- GluRS is placed within the Tyr/TrpRS and LysRS-I clade with moderate support.

Unlike previous studies, our phylogenies remain unrooted to avoid the biases of computational root inference. Instead, we present a curated list of candidate roots that remain viable under current evidence. Each candidate root carries unique and significant implications for our understanding of how the genetic code evolved.

Structome-DeepRoots: On inferring homology using high-dimensional embeddings

Ashar Malik, The University of Queensland, Australia

It's established that structure is more conserved than sequence, a fact rooted in the constraints of 3D geometry. This is why structural comparisons often succeed in recovering deeper signals where sequence alignments fail. This begs a powerful question: if moving from a 1D sequence to 3D geometry reveals so much more evolutionary history, what if we increased the dimensionality even further? Could analyzing proteins in hundreds of dimensions uncover deeper-resolved signals invisible even in 3D? This is where protein language models that encode the complex interplay of sequence, structure, and function into rich, high-dimensional embeddings become helpful. I will introduce a novel method, Structome-DeepRoots, that uses these powerful representations to move beyond conventional 3D analysis, tapping into latent space to reconstruct evolutionary history with greater resolution than ever before. In this talk, I will show how it recovers known splits on curated data and outperforms our baseline method, Structome-TM.

A grammar-based discovery engine for structural relationships in the global TIRome.

Bostjan Kobe, University of Queensland, Australia

TIR (Toll/interleukin-1 receptor) domains are widely distributed in animals, plants and bacteria, and function through self-association and homotypic interactions with other TIR domains [1]. Across phyla, these domains feature in proteins with immune functions – e.g. Toll-like receptors, plant NLRs and bacterial antiphage defence proteins. Many TIR domains across kingdoms also feature enzymatic activities - cleavage of nucleotides such as NAD^+ [2,3]. The TIR-domain structure is related to the flavodoxin fold and structurally similar proteins (which are not usually associated with TIR domains) bind nucleotides (e.g., glycosyltransferases, nucleoside hydrolases, and flavodoxins) or carbohydrates (e.g., bacterial isomerases) in the analogous region of the protein.

Phylogenetic analyses could be valuable in understanding the evolution of TIR-domain activities and informing the functions of uncharacterized TIR domains and structurally related proteins. However, limited sequence similarities of TIR domains make phylogenetic analyses using sequence-based evolutionary models unreliable. We therefore introduce a structure-based "grammar" approach that calibrates a reference TIR signature as an ordered set of secondary-structure elements, and scans candidate structures by aligning them to the reference and projecting the reference elements onto each target. From this projection, we compute a scaffold score that captures fold-level agreement and a complementary loop-variation score that highlights local divergence in connecting regions often associated with function.

We will report representative examples where the method identifies informative structural relationships, providing testable hypotheses for activities associated with TIR-domain functions.

[1] Nimma et al (2021) Front Immunol 12, 784484
[2] Horsefield et al (2019) Science 365, 793
[3] Manik et al (2022) Science, eadc8969

Protein structure and the beginning of a third era in phylogenetics

Caroline Puente-Lelievre, University of Auckland, New Zealand

The rising integration of protein structure data into molecular phylogenetics analyses marks the beginning of a third era in phylogenetics, one in which evolution is interpreted through three-dimensional form rather than linear code. The shift is not just technical; it changes how homology is detected, how gene families are defined, and reshapes our understanding of molecular evolution. Advances in deep learning–based structure prediction and the rapid expansion of structural databases provide access to evolutionary information that complements, and in some cases extends beyond, sequence-based inference.

Structural phylogenetics sits at the intersection of structural biology, biochemistry, evolutionary theory, and computational science, and repositions evolutionary biology within multidisciplinary and translational research. Realising its potential requires methodological innovation, scalable computational approaches, and closer integration between disciplines that have traditionally operated in parallel.

At a time when evolutionary biology must articulate its relevance in an increasingly outcome-driven research landscape, structural phylogenetics provides a bridge between fundamental theory and translational research. Drawing on our work on bacterial motility and structure–function relationships, I will illustrate how we can address fundamental questions while opening pathways toward applied research in the context of Aotearoa New Zealand.

Beyond Phylogenetic Inference from Structure

Charlie Carter, University of North Carolina at Chapel Hill, USA

Interpreting triplet codons along genes as strings of amino acids boosted their latent biological activity by about a billion-fold. We try to unravel how nature first invented genetic coding by discovering how to do that. Aminoacyl-tRNA synthetases (synthetases) are enzymes that translate the code by forming a covalent bond between specific amino acids and RNA substrates carrying anticodons. There are two distinct synthetase superfamilies. Each Class manages about ten of the twenty amino acids. Class 1 and 2 synthetase ancestries may be coupled. There is mounting evidence that at one time they were encoded on opposite strands of the same ancestral gene(s). The reflexive nature of synthetase genes is a source of fascination. Their translated products must enforce the coding rules by which they, themselves were assembled. Thus, their sequences are a rich source of data relevant to our studies, as are inferences based in structural alignments. Such inferences are necessary, but not sufficient to enable us to reconstruct their ancestry. We try to go beyond those inferences. To that end, we created ways to express and assay two different ancestral synthetase models. Protozymes have ~50 amino acids. They retain binding sites for amino acids and ATP and accelerate amino acid activation by a million-fold. Urzymes are a bit larger (90-130 amino acids) and can both activate amino acids and transfer them to cognate tRNAs and minihelices. These experiments furnish rates and specificities for the ancestral models. We first hope to identify convergence of sequences reconstructed from phylogenetic trees and those of our in vitro designs. Second, we hope to define a set of combinatorial gene libraries whose translated products have complementary specificity spectra capable of translating the combinatorial libraries of their own genes (Carter, et al. 2025). Recently, we also have found a protocol that produces similar constructs from full-length plasmids in vivo (Tang and Carter 2025). These new models open new possibilities moving us in that direction.

References

Carter CW, Jr, Tang GQ, Patra SK, Betts L, Dieckhaus H, Kuhlman B, Douglas J, Wills PR, Bouckaert R, Popovic M, Ditzler M. 2025. Structural Enzymology, Phylogenetics, Differentiation, and Symbolic Reflexivity at the Dawn of Biology. Genome Biology and Evolution 17: evaf095.

Tang GQ, Carter CW, Jr. 2025. Escherichia coli deletes in vivo the same domains from a double-mutant leucyl-tRNA synthetase gene that were deleted in vitro to make the LeuAC urzyme. BioRxiv doi: https://doi.org/10.1101/2025.04.05.647346.

Mycovirome Online Platform (MVoP): A collaborative online tool for mycovirus discovery.

Cinthy Lorena Jimenez Silva, The University of Auckland, New Zealand

To advance understanding of mycovirus biology, we need a fungal host that serves as a model system. Mycoviruses exhibit high diversity and can influence the phenotype of their fungal hosts, yet they remain underexplored in both genomic and functional studies. Similarly, fungi are highly diverse with many causing severe symptoms on a wide range of hosts, including crop and tāonga plants. We propose Botrytis cinerea, a pathogen of 1400+ plant species that already has excellent genetic/biological resources, as a central model for global mycovirus research, enabling systematic investigation of virome diversity, epidemiology, evolution and host–virus interactions. We introduce the Mycovirome Online Platform (MVoP), an interactive Wikipedia-like tool that provides a curated database, and virus discovery pipelines to support collaborative research. This platform systematizes existing knowledge about the mycoviruses that have been characterize in B. cinerea by showcasing their geographic and temporal distribution. It also details their taxonomic diversity of selected mycovirus sequences including their phylogenetic relationships. Through its graphical user interface, MVoP facilitates a seamless exploration between mycovirus diversity, sequences, protein structure, and phylogenies, providing a friendly introduction to the material for non-experts complemented by insightful resources for experts. We outline the essential steps for developing B. cinerea as a model and invite the community to engage with tools, data, and workflows that facilitate virome characterization and knowledge sharing. Pipelines, interactive plots and curated multiple sequence alignments can be extracted for downstream analyses. Establishing a unified model will allow researchers to address complex questions that cannot be resolved through fragmented approaches. As a major plant pathogen, insights from B. cinerea have immediate practical relevance and can be translated to other fungal systems, enriching our understanding of mycovirus-driven ecology and evolution. This knowledge is a crucial step toward bioprotection strategies targeting agroecosystems and native flora.

Charting the protein universe: Deep evolutionary signals in sequence-structure space

Claudia Alvarez Carreno, University College London, UK

The protein universe is vast and ancient. The first protein folds originated and began to diversify more than 3.5 billion years ago; consequently, the deep evolutionary relationships among folds remain difficult to resolve. Although some regions of protein space are well characterized, much of it is still largely uncharted. The AlphaFold Database (AFDB) has transformed access to structural information by providing predicted models for over 214 million sequences. At the same time, advances in structural bioinformatics and computational molecular evolution now allow us to organize and interpret this diversity at unprecedented scale.

In this talk, I present an analysis of a map of sequence relationships, coupled with phylogenetic context, to explore the evolutionary history of protein folds. This map was generated from pairwise comparisons of profile Hidden Markov Models (HMMs) derived from The Encyclopedia of Domains (TED). TED catalogues 365 million putative domains from AFDB v4, with more than 251 million mapped to the CATH classification of protein structures. I show how HHM–HHM comparisons reveal signals of deep ancestry within and across CATH superfamilies. Together, these analyses provide a framework for reconstructing how folds emerged and diversified at the earliest stages of protein evolution.

Bayesian Phylogenetic Inference of Protein Structure Evolution Using Angular Diffusion Model

Clementine Yan, The University of Auckland, New Zealand

Protein structure is more conserved than amino acid sequence over long evolutionary timescales, yet most phylogenetic methods rely solely on sequence data. The angular diffusion model proposed by Golden et al. (2017) uses dihedral angles (ϕ and ψ) to represent protein backbone structure and detect structural dependencies between neighbouring Cα atoms. However, this model has not been integrated into a phylogenetic framework. Here, we implement the angular diffusion model for phylogenetic inference by incorporating it into the Bayesian phylogenetics software BEAST2. We extend the tree likelihood function to compute the pseudo-transition probability density described by Golden et al. We validate our implementation using simulated datasets and evaluate its performance on a ferritin dataset. By incorporating structural dependencies, this approach captures more realistic evolutionary trajectories and provides improved estimates of evolutionary parameters compared to sequence-only methods. Our implementation enables researchers to leverage protein structural information for more accurate phylogenetic reconstruction.

FoldTree2: Integrative Sequence–Structure Phylogenetics with GNN-Derived Structural Alphabets

David Moi, UNIL Department of Computational Biology, Switzerland

Inferring deep protein relationships requires models that reflect how evolution optimizes the folded object, not just its sequence. Our earlier framework, FoldTree, established a topology-based benchmark to compare structure- and sequence-based phylogenetics and showed that phylogenies incorporating local structural information outperform sequence-only approaches for remote homologs and surprisingly sharpen resolution among close relatives, enabling insights such as the diversification of RRNPPA quorum-sensing receptors across bacteria, plasmids, and phages. Here we introduce FoldTree2, a new generation of structural alphabets learned with graph neural networks (GNNs). Proteins are represented as residue-level graphs with physicochemical and geometric node features and edges encoding proximity, hydrogen bonds, and long-range coupling. A vector-quantized autoencoder maps structures to a compact N-token alphabet and can reconstruct contact patterns and amino acids from strings of discrete tokens alone, demonstrating that the alphabet captures both local environments and global constraints. We have built tools to use this new representation for alignment and maximum-likelihood phylogenetic inference and have quantified its performance using a new set of benchmarks designed to measure the information content of the different representations of the protein object and the quality of ancestral state reconstruction with regards to foldability alongside our previous benchmarks on tree topology and ultrametricity.

Hook, Line, and Swimmer: A look into the bacterial flagellar hook and filament capping proteins

Divyangi Pandit, University of New South Wales, Australia

Bacterial motility is powered by the bacterial flagellar motor (BFM), a molecular nanomachine comprising of a rotor with stators, a flexible hook, and the extracellular flagellar filament capped by the filament cap protein. Torque is generated in the basal body which is partially embedded in the cytoplasmic membrane and transferred through the hook to the filament which rotates to generate thrust and propel the bacteria forward in liquid media. As such the interactions between these proteins are integral to ensuring functional motility. The hook is comprised of multiple subunits of the FlgE protein which forms a flexible, yet stable, helical structure acting as a universal joint. The filament cap protein is composed of FliD subunits which can arrange themselves into pentamers or hexamers depending on bacterial species. To understand the evolutionary context of these proteins, we can create ancestral sequence reconstructions (ASR) of Escherichia coli FlgE and FliD proteins where modern extant amino acid sequences are used to calculate ancestral sequences at particular nodes of interest.

Lost in translation: the sequence verification challenge in studies of protein evolution by ancestral sequence reconstruction

Elizabeth M. J. Gillam, The University of Queensland, Australia

Ancestral sequence reconstruction (ASR) is a computational approach that leverages sequence data from the extant members of protein families to infer how they have evolved from ancient precursors. Ancestral proteins at key evolutionary stages can then be resurrected, i.e., synthesised and characterised by biochemical and biophysical approaches, to trace the structural and functional changes that have occurred through the evolutionary history of the protein family. The expansion of genome sequencing over the last two decades has led to a massive increase in the availability of extant protein sequences for many protein superfamilies, providing abundant sequence information as input for ASR. However, databases are rife with sequence errors resulting from imperfect interpretation of start and stop sites and intron/exon boundaries. Effective sequence curation relies on knowing how much variation can be tolerated by a protein fold. However, it is not often possible to discriminate erroneous sequences from bona fide sequence diversity since only a small proportion of the available sequences have been expressed and proven to encode structurally intact, functional proteins. This presentation will explore the importance of using biochemical and evolutionary insight to eliminate poor quality sequences prior to ASR, with examples drawn from reconstructions of several enzyme families and subfamilies involved in secondary metabolism in animals, plants and bacteria.

Mosaic Evolution of Zinc Selectivity: Decoupling Chemical Optimization from Structural Constraints in Bacterial Solute-Binding Proteins

Evelyne Deplazes, The University of Queensland, Australia

According to the laws of chemistry, Zinc naturally binds to proteins much more strongly than Manganese. To overcome this, bacteria use special pocket designs that fit Manganese but attempt to block Zinc. However, how some of these proteins evolved to switch back to carrying Zinc after they had already perfected the "harder" task of binding Manganese remains poorly understood.
We used a combination of evolutionary mapping and machine learning to study how these proteins change over time. By training a computer model on thousands of protein structures, we created a "Zinc Score" that measures how likely a protein is to prefer Zinc based on its shape and chemical makeup.
Our early results show a mosaic pattern of evolution: the local chemistry of the binding site can adapt to a new metal much faster than the overall shape of the protein changes. This suggests the protein's outer structure acts as a scaffold that limits how the protein can evolve, even while the inner pocket is being fine-tuned. Overall, "Zinc Score" can facilitate the design of new proteins to grab specific metals by changing just a few chemical parts rather than rebuilding the entire protein structure.

Cross-phyla protein annotation by structural prediction and alignment

Fabian Ruperti, The University of Queensland, Australia

Protein annotation is a major goal in molecular biology, yet experimentally determined knowledge is typically limited to a few model organisms. In non-model species, the sequence-based prediction of gene orthology can be used to infer protein identity; however, this approach loses predictive power at longer evolutionary distances. Here we propose a workflow for protein annotation using structural similarity, exploiting the fact that similar protein structures often reflect homology and are more conserved than protein sequences.

We propose a workflow of openly available tools for the functional annotation of proteins via structural similarity (MorF: MorphologFinder) and use it to annotate the complete proteome of a sponge. Sponges are highly relevant for inferring the early history of animals, yet their proteomes remain sparsely annotated. MorF accurately predicts the functions of proteins with known homology in >90% of cases and annotates an additional 50% of the proteome beyond standard sequence-based methods. We uncover new functions for sponge cell types, including extensive FGF, TGF, and Ephrin signaling in sponge epithelia, and redox metabolism and control in myopeptidocytes. Notably, we also annotate genes specific to the enigmatic sponge mesocytes, proposing they function to digest cell walls.

Our work demonstrates that structural similarity is a powerful approach that complements and extends sequence similarity searches to identify homologous proteins over long evolutionary distances. We anticipate this will be a powerful approach that boosts discovery in numerous -omics datasets, especially for non-model organisms.

Genome-wide analysis of protein intrinsic disorder in *Drosophila melanogaster*

Hassan Sibroe Abdulla Daanaa, National Institute of Genetics, Japan

Addressing the relevance of protein structure in function and evolution remains challenging. One established paradigm posits that function necessitates a well-folded structure. A debated paradigm, termed disorder-function, suggests that function can also exist without a folded structure. Efforts to test disorder-function focus on intrinsic disorder, which commonly refers to highly flexible peptide segments that presumably lack structure under physiological conditions. Computational predictions of intrinsic disorder at the residue-level have enabled large-scale analysis in many organisms, yet important challenges remain. The predictions can be conflicting among approaches and the prevalence of intrinsic disorder along protein sequences has been examined in only relatively narrow scopes. I address these challenges mainly in *D. melanogaster* by leveraging various publicly available data (e.g., genomics) and multiple intrinsic disorder prediction programs. I present evidence that intrinsic disorder is unevenly distributed along protein sequences, in a manner associated with amino acid composition, tertiary structure and molecular interactions (e.g., protein-binding), suggesting functional significance. Furthermore, estimates of protein evolutionary rates at disordered residues appear generally similar along proteins, and are higher than those rates at non-disordered residues, hinting at weaker sequence constraints of intrinsic disorder. I observe that biases in disorder prediction can be a factor in the overall patterns, but do not appear to offer a sufficient explanation. Findings in this study strengthen support for the disorder-function paradigm and offer new insights into the possible locations and primary roles of disorder, giving impetus for experimental inquiry.

A new information criterion favors mixture models over partition models and influences partition model selection.

Huaiyan Ren, Australian National University, Australia

In phylogenetics, partition and mixture models are widely employed to analyse large and heterogeneous genomic datasets using Maximum Likelihood methods. A common question in the field is: Which model better fits the data? Traditional model selection methods based on information criteria tend to show undesirable biases for such comparisons. Susko et. al. (2025) introduced a new measure, marginal Akaike information criterion (mAIC), which adjusts the likelihood of partition models in a way that is comparable to that of mixture models.

We here introduce an efficient implementation of mAIC in IQ-TREE. An extensive benchmark on a diverse set of empirical datasets demonstrates that mixture models are generally favoured. We also applied mAIC to PartitionFinder, an algorithm for identifying optimal partition schemes, and found that mAIC supports substantially different schemes compared with the traditional AIC, calling into question the use of AIC for model selection.

A Tale of Tails: Investigating the Evolution and Adaptation of the Bacterial Flagellar Filament

Jacob Scadden, University of New South Wales, Australia

Motility and chemotaxis are major adaptations that have allowed microbial life to expand to nearly every environment on Earth. One of the major mechanisms of bacterial motility utilises the bacterial flagellar motor (BFM) to drive movement. Understanding the evolution of the flagellar filament protein, flagellin, is important in elucidating the diversity observed within this protein and its adaption in various environments to provide motility. We generated multiple phylogenetic trees using representative flagellin sequences to generate ancestral sequence reconstructions (ASR) of "ancient" flagellin proteins. Candidates were cloned into Escherichia coli K-12 and Salmonella enterica serovar Typhimurium strains which had the flagellin gene(s) deleted. All ASR candidates did not generate filaments in E. coli, however, when expressed in S. Typhimurium motility was observed when some "ancestral" flagellins were expressed. Additionally, to test the ability of the surface exposed domains of flagellin to be exchanged, we designed flagellin chimeras and expressed these in E. coli ΔfliC. Two of the eleven outer domain chimeras, as well as outer domain deleted flagellin variants, were able to form filaments and provide motility. This study showed that the outer domains are not necessary for motility in E. coli K-12 and that rational engineering of outer domains can provide motility in addition to identifying key residues that are maintained in flagellar filaments.

Easy come, easier go: mapping the loss of flagellar motility across the tree of life

Jamiema Sara Philip, UNSW, Australia

The bacterial flagellum is a macromolecular complex that rotates at high speeds to propel bacteria through different environments. Gene clusters encoding flagellum components can contain up to approximately 50 protein-coding genes. However, the variation in the number and composition of flagellar proteins, coupled with their rapid divergence in bacterial species, has largely obscured the origins of flagellar evolution. Quantifying the proteins present in flagellar motors from a range of species delivers insight into how motility has changed throughout history and provides a platform to develop genome-based predictions on whether a species is likely to be motile. Traditional sequence-based methods struggle to identify flagellar proteins due to low homology. To overcome this, we employed large-scale sequence and structural comparisons to investigate flagellar protein distribution across diverse bacteria. Using these data, we developed a high-accuracy (95%) classifier that predicts bacterial motility. By mapping flagellar motility onto the GTDB bacterial tree of life, we then confirmed that the last common bacterial ancestor was motile, with flagellar loss four-fold higher than the rate of gain. We also identified a strong correlation between the presence of filament homologs and bacterial motility. By integrating structural homology with sequence homology searches, we refined the detection of distant flagellar homologues, mitigating false positives and shedding light on previously unexplored species. This holistic approach indicates the significance of true data integration in evolutionary genomics. Overall, our study provides key insights into the evolutionary dynamics of flagellar motility, improves genome-based predictions and provides a framework for further ecological studies on flagellar diversity and adaptation.

Exploring Time-Nonreversibility and Site-Mixtures in Virus Phylogenetics with ABySS

Jasmine Saghafifar, University of Auckland, New Zealand

Modern structure prediction tools like AlphaFold mean we now have more protein structures to study than ever, which corresponds with the rising importance of structural phylogenetic methods that can compete with sequence methods. Phylogenetic inference of protein evolution traditionally involves the use of a fixed time-reversible substitution model informed by empirical data (favoured for simplicity and small parameter-space), and optionally, manual partitioning by structural domains. However, these norms were built when computational intensity was a more pressing concern; with the ever-increasing speeds of processors, we may investigate under-explored approaches to modelling these processes in the hopes of extracting as much from data as possible for more realistic and effective inference.

ABySS (Advanced Bayesian Site and Substitution) is a new software package with time-nonreversible substitution and a combined site-mixture and model averaging approach (coined as Hmix). In this talk, I will demonstrate (1) the models' implementation into a Bayesian framework (BEAST2), (2) its validity through well-calibrated simulation studies, and (3) its performance compared to traditional methods, using protein sequence and structural data. I will discuss whether virus protein structural evolution modelled under time-nonreversible substitution would outperform time-reversible substitution models, and that site-mixture or Hmix will prove an effective tool for substitution model selection and comparison, including bypassing the need for manual partitioning methods.

What can we learn from ancient (protein) history?

Joanna Masel, University of Arizona, USA

We assigned dates of origin not to whole proteins, but to their modular domains, as annotated by the Pfam database as domains (HMM profiles) and clans (sets of HMM profiles believed to be evolutionarily related albeit highly diverged). The genetic code of 20+ amino acids is believed to have been completed around the time of the Last Universal Common Ancestor (LUCA). We therefore inferred which amino acids were added first to the genetic code on the basis of their enrichment in ancestrally reconstructed clans dating back to LUCA, relative to ancient but post-LUCA clans. Smaller amino acids were recruited first, with different curves for amino acids activated by Class I vs. Class II amino acyl tRNA synthetases. The most ancient clans, which had already diverged pre-LUCA, used more aromatic amino acids, even tryptophan, which is the largest and last-added of the standard 20 amino acids. Clans that diverged prior to the Last Bacterial Common Ancestor, but which were not in LUCA, used far more tryptophan than similarly ancient archaeal clans. Tryptophan's synthetase emerged post-LUCA in archaea and was transferred to bacteria, suggesting that ancient bacteria once had a different mechanism for incorporating tryptophan. In the process of resolving tryptophan's origin, we detected an ancient recombination event between versions of tyrosine's synthetase on the stem bacterial vs. stem archaeal branches. Patterns of metal usage have also changed over time, in ways that reflect both supply under changing oxygenation conditions, and functional demand.

The Promise and Challenges of Viral Protein Structural Phylogenetics Across Deep Time

Jonathon Mifsud, KU Leuven, Belgium

The rapid rate of virus evolution leaves us with little genomic traces over deep evolutionary timescales, complicating the reconstruction of deep viral evolutionary history. These challenges, however, make viruses a powerful system for testing recent advances in protein structure prediction and phylogenetics. In this talk, I will discuss the promise and pitfalls of applying protein structural prediction and phylogenetics to viral proteins and how this approach may help address fundamental questions in virus evolution. I will then present findings from studies of Flaviviridae viruses, showing how protein structure prediction combined with phylogenetics reveals previously unrecognized evolutionary connections underlying viral membrane fusion mechanisms and virus–host evolution. Finally, I will introduce preliminary work on integrating viral protein structural data into time-resolved Bayesian phylogenetics.

Stochastic mapping and ancestral gapping: reconstructing the evolutionary pathways in protein sequences, structures and indels

Jordan Douglas, Australian National University, Australia

Ancestral sequences and substitution histories are usually averaged out in phylogenetic inference and are therefore not reported to the user. However, they can be recovered through ancestral sequence reconstruction (ASR) and stochastic mapping. In this talk I will present a Bayesian phylogenetic implementation of an ASR and stochastic mapping. I will show how this approach can be used to infer the evolutionary history of insertions and deletions, which is especially important for studying ancestral protein sequences. I will also show how the history of amino acid and 3Di character changes can be reconstructed using this approach.

Douglas, J., & Bromham, L. (2025). Reconstructing substitution histories on phylogenies, with accuracy, precision, and coverage. bioRxiv, 2025-12.

Expanding kinetoplastid genome annotation through protein structure comparison

Juan Manuel Trinidad-Barnech, Universidad de la República, Uruguay

Kinetoplastids belong to the Discoba supergroup, an early divergent eukaryotic clade. Although the amount of genomic information on these parasites has grown substantially, assigning gene functions through traditional sequence-based homology methods remains challenging. Recently, significant advancements have been made in in-silico protein structure prediction and algorithms for rapid and precise large-scale protein structure comparisons. In this work, we developed a protein structure-based homology search pipeline (ASC, Annotation by Structural Comparisons) and applied it to transfer biological information to all kinetoplastid proteins available in TriTrypDB, the reference database for this lineage. Our pipeline enabled the assignment of structural similarity to a substantial portion of kinetoplastid proteins, improving current knowledge through annotation transfer. Additionally, we identified structural homologs for representatives of 6,700 uncharacterized proteins across 33 kinetoplastid species, proteins that could not be annotated using existing sequence-based tools and databases. As a result, this approach allowed us to infer potential biological information for a considerable number of kinetoplastid proteins. Among these, we identified structural homologs to ubiquitous eukaryotic proteins that are challenging to detect in kinetoplastid genomes through standard genome annotation pipelines. The results (KASC, Kinetoplastid Annotation by Structural Comparison) are openly accessible to the community at kasc.fcien.edu.uy through a user-friendly, gene-by-gene interface that enables visual inspection of the data.

Bridging Protein Structure Analysis and Phylogenetics with Mass Spectrometry

Kevin Downard, Prince of Wales Clinical Research Sciences, UNSW, Australia

Protein structures have been interrogated and analysed with mass spectrometry for several decades. From charge distribution analysis, to ion mobility, and a multitude of labelling methods (eg. HX, RP, XL and LP), mass spectrometry has evolved from a primary, to tertiary and quaternary structural detection approach.

We have developed a means with which to construct protein phylogenies using mass spectral data that provides a bridge to protein structures. Known as “phylonumerics”, the approach avoids the need to generate sequence data or perform an alignment. Mass spectral datasets are input directly into a purpose built algorithm to chart the evolution of a protein across divergent species through a pairwise comparison of peptide masses produced from their digestion.

Homologous proteins, featuring a set of more common peptide masses within their mass maps, are grouped together on the mass tree, while unrelated proteins are positioned further apart. The algorithm also calculates all possible single point substitutions, and displays these on the tree, in the same tree building step. Such substitutions may be associated with mutations that impact upon a protein’s structure.

In this presentation, we demonstrate how these structure altering mutations can be tracked along interconnecting branches of a mass tree, to follow their consequences for a protein's evolution, by assessing which structural perturbations achieve a functional gain. Results for model proteins and those important to the evolution of respiratory viruses will be presented.

Estimating evolutionary time from protein structure: what do we need to make it work?

Lindell Bromham, Australian National University, Australia

Molecular clocks – evolutionary dates based on comparison of molecules between species – have revolutionised biology and provided new ways of answering long standing questions, from recent viral origins to the beginnings of cellular life. Molecular clocks have also changed evolutionary theory – the discovery of surprisingly clock-like evolution of proteins was one of the foundational observations on which neutral theory was based. An early conclusion was that DNA sequences evolve according to generation time, but proteins according to absolute time, due to the interplay between mutation rate, population size and selection. But the theoretical basis of protein clocks was based on a number of assumptions – that most amino acid substitutions are neutral, that mutation rate is predictable from generation time, and that generation time is inversely related to population size. In contrast to the neat predictions of theory, empirical studies show that rate of change in proteins is more complicated, shaped by life history trade-offs and selection on mutation rate. Can molecular dates based on protein structure overcome some of the problems associated with sequence-based molecular clocks? Whether structural variants change in a clock-like manner depends on how the rate of origin of structural variants varies across the genome, between individuals, among lineages, and over time, as well as the distribution of fitness effects for changes in structure, and the size and structure of the populations in which they occur. If we can address these issues, then molecular clocks based on protein structure might open up new possibilities for dating ancient evolutionary events.

More Than One Way to Lose a Gene: Distinct Evolutionary Fates of NOX5 in Rodents

Marco Antonio Chávez Tinoco, Universidad Nacional Autonoma de México, México

Nox5 is an essential NADPH oxidase in human sperm, where it produces reactive oxygen species (ROS), which are key to the capacitation process through its interaction with the HV1 ion channel. However, NOX5 is absent in mice, posing an evolutionary puzzle given its importance in humans. This work aims to determine the adaptive causes and consequences of this gene loss. By combining comparative phylogeny to identify the divergence of NOX5 and its correlation with HV1, as well as transcriptomics (RNA-seq) and functional assays—such as ROS measurement and subcellular localization—we seek to uncover compensatory mechanisms in mice. Our results will provide clues on how evolution reshapes redox signaling networks in mammalian reproduction, with implications for infertility studies and evolutionary biology.

Exploring the Protein Universe with Multimers, Motifs & Interfaces

Martin Steinegger, Seoul National University, South Korea

The rapid rise of highly accurate structure predictors has transformed our ability to explore proteins at a global scale. With billions of predicted structures available, new opportunities emerge to study not only monomers, but also multimer organization, conserved structural motifs, and protein–protein interfaces. In this talk, I will present our fast computational methods Foldseek-Multimer, Folddisco, and Foldseek-Interface, which enable large-scale search, comparison, and clustering across these structural layers. By integrating multimeric assemblies, structural motifs, and interaction interfaces, we can systematically map functional space, annotate previously uncharacterised proteins, and gain new insights into protein evolution and biological mechanisms.

Structures of the YenTc holotoxin prepore and pore reveal the evolutionary basis for chitinase incorporation into ABC toxins

Michael Landsberg, The University of Queensland, Australia

ABC toxins are toxin-translocating, pore-forming proteins found in a wide range of insecticidal bacteria and some mammalian pathogens. They are comprised of three major components which initially assemble into an soluble, oligomeric pre-pore before being triggered by an as yet undefined physiological stimulus to form membrane-penetrating pores. The Yersinia entomophaga toxin complex (YenTc) was the first example of an ABC toxin to have its component structures determined and consequently became a prototypical example of this class of bacterial toxins. Studies conducted in parallel and subsequently on other family members, however, suggest the existence of distinct subclasses with divergent molecular architectures and potentially, different modes of action. We recently used cryo-EM in conjunction with Alphafold2-assisted structural modelling of flexible domains to define the structural mechanism via which enzymatically-active chitinase subunits are incorporated into the holotoxin structure of YenTc. We showed using phylogenetic analyses that this is a defining feature associated with a subclass of ABC toxins that has evolved relatively recently. Improved cryo-EM structures of the YenTc pre-pore and pore also pinpoint structural differences which influence its mechanism of pore-formation and appear to be unique to this subclass. Expanding our phylogenetic analysis provides evidence that at least three major clades of ABC toxins exist, one of which has not been defined previously. We suggest that investigation of these may reveal further, as yet unappreciated diversity in function and mechanism of ABC toxins. Our findings therefore enhance our understanding of the structural diversity that defines distinct ABC toxin subclasses and provide compelling avenues for future research.

From Molecular Machines to Evolutionary Blueprints

Mihnea Bostina, University of Otago, New Zealand

The microscopic world harbors some of nature's most intricate molecular machines, bacteriophages, viruses that infect bacteria and hold revolutionary potential for combating antibiotic resistance. Bacteriophages with long contractile tails (myophages) are among the most complex biological assemblies known, comprising thousands of protein subunits derived from over twenty distinct gene products. These remarkable nanomachines store all the energy required for infection, maintaining a stable structure in the environment until host recognition triggers tail contraction and subsequent genome ejection.

We present our recent work on parallel cryo-EM reconstructions of two myophages to build a structural phylogenetic framework for tail evolution. The Pectobacterium phage ΦTE, a flagellotropic myophage, features radially extended tail fibers, a compact baseplate hub, and a bouquet-hinge adaptor that propagates a signaling cascade from flagellar contact to sheath contraction. The Escherichia phage Bas63 displays tail fibers aligned parallel to the tail axis, a novel fibre platform, and an extended rope-domain collar enabling modular baseplate assembly and puncture.

Using cutting-edge cryo-EM, AI-driven modeling, and complementary methods, we resolved over 128 unique protein chains. In total, the two structural assemblies comprise over 3,000 subunits, representing the largest near-complete atomic models of phage tails to date.

Structural conservation mapping reveals a bimodal pattern: a core machinery (portal, adaptor, tube, sheath, tape measure, hub) that is nearly 100% conserved, and host-interacting components (decoration proteins and tail fibers) that diverge dramatically—driving receptor specificity and host-range adaptation. Comparing the atomic architectures of Bas63 and ΦTE within a cross-species phylogenetics, we show how structural modularity balances functional conservation with adaptive divergence. These two structures provide snapshots of phage tail evolution and offer blueprints for precision phage engineering.

References

1. Hodgkinson-Bean J, Ayala R, Jayawardena N, Rutter GL, Watson BNJ, Mayo-Muñoz D, Keal J, Fineran PC, Wolf M, Bostina M. 2025. Global structural survey of the flagellotropic myophage φTE infecting agricultural pathogen Pectobacterium atrosepticum. Nature Communications. https://doi.org/10.1038/s41467-025-58514-x
2. Hodgkinson-Bean L, Ayala R, McJarrow-Keller , Cassin L, Rutter GL, Crowe AJM, Wolf M, Bostina M. 2025. Cryo-EM Structure of Bacteriophage Bas63 Reveals Structural Conservation and Diversity in the Felixounavirus genus. Science Advances. https://doi.org/10.1126/sciadv.adx0790

From rocks to life: the primordial assembly of microbial metabolism

Natalia Mrnjavac, Henrich Heine University Düsseldorf, Germany

A cornerstone implication of Charles Darwin's work in the 1800s was that all life shares common ancestry. The last universal common ancestor (LUCA) is a hypothetical stage in the early evolution of life connecting prebiotic reactions at life's origin and cellular life. According to the two-domain tree of life, LUCA is the common ancestor of bacteria and archaea. In order to learn more about the transition from LUCA to the last bacterial (LBCA) and the last archaeal (LACA) common ancestors, we carried out a phylogeny-independent computational study focusing on core metabolic genes. What can we learn about the presence and evolution of core metabolic pathways that synthesize amino acids, cofactors and nucleotides in LUCA, LBCA and LACA? What was the role of the environment in this transition, and how does this align with current theories? By relying on the distribution of structurally enhanced protein families in order to circumvent the challenges of phylogenetic reconstructions, we get a glimpse into the primordial assembly of metabolism.

Protein structure characters in the light of phylogenetic systematics

Nicholas Matzke, University of Auckland, New Zealand

Protein structure characters have great potential for improving phylogenetic inference, especially for deep nodes where amino acid sequences are highly diverged. The combination of AlphaFold structure predictions and Foldseek's "3Di" structural alphabet makes it relatively easy to conduct model-based phylogenetic inference that includes a partition of slow-evolving 3Di characters. However, we show that even identical amino acid sequences can produce substantially different 3Di codes, depending on the source of structural model and whether inter-chain interactions are considered. We argue that such variability can be addressed with key concepts from traditional organism-based phylogenetic systematics: semaphoront, hypodigm, and character ascertainment method. To illustrate this, we develop an analogy between organismal development, taphonomy, and subsequent description and character coding by a systematist, and the process of protein synthesis, folding, and interaction and subsequent extraction, experimentation, and structural modeling by a biochemist. We conclude that differences in 3Di codes between semaphoronts and are not intrinsically a problem, but they do require that the researcher uses the same replicable method on all proteins in the phylogenetic analysis. The guiding principle should be to maximize the chance that character differences in the data matrix are the results of underlying evolutionary changes, rather than artefactual differences between proteins due to differences in the methods used for obtaining semaphoronts and coding characters.

Genetic coding represented as the first local example of intelligence in the cosmos

Peter Wills, Department of Physics, University of Auckland, New Zealand

Structural phylogenetic methods provide a more direct insight into the evolution of function than other tree-building exercises that rely solely on alignments of gene or protein sequences. Genetic sequence information provides a remarkably compact description for the synthesis of a protein, but the concomitant production of functionality is mediated through the enormously complex filter of protein folding. The ability to simulate that process has reinforced the truism that evolution conserves a protein's structure much more tightly than its amino acid sequence. Reliably preserving functionality across evolutionary timescales is assisted by the rigidity of genetic encoding and the mechanism of its expression (translation). Thus, changes in function are generally gradual, slower than the rate of mutation. Furthermore, potentially extreme effects of disruptive noise in the maintenance of functionality are moderated by degeneracy at two levels of the genotype-phenotype correlation: (i) redundancy in sequence encoding; and (ii) the tolerance of amino acid substitutions. However, Darwinian gradualism cannot explain its own existence: it fails to account for the emergence of the genotype-phenotype relationship that underpins it. What then were the mechanisms whereby information-processing (computation) came to dictate and control the chemistry that occurs in living systems? Taking the origin of genetic coding as an example, I will characterise a class of sharp, non-Darwinian evolutionary transitions that entail the progressive spatiotemporal emergence/invention of biologically intelligent "sorting" processes, especially systems that learn about the functionality of the structures they produce.

The Genome Taxonomy Database (GTDB): a potential use case for structural phylogenetics

Phil Hugenholtz, The University of Queensland, Australia

The GTDB is a genome-based taxonomy for Bacteria and Archaea that defines taxonomic groups according to concatenated marker gene reference trees (one for each domain) to produce a standardised taxonomy from species to domain. Therefore, the quality of the reference trees is critical to our ability to provide a complete and systematic taxonomy. *De novo* reference trees are inferred for each annual release and groups are mostly stable between releases. However, long branch attraction (LBA) artefacts in the trees introduce noise into the taxonomy via loss of monophyletic groups and/or changes in relative evolutionary divergence, which is used to temporally standardise the GTDB taxonomy. Also, ancient relationships at the phylum level are often poorly resolved resulting in loss of potentially useful taxonomic information. Therefore, the GTDB may present a useful case study for structural phylogenetics if structure-based trees are more robust (in particular to LBA artefacts) and provide greater resolution of deep relationships than their sequence-based counterparts.

From Ions To Force: Structure and evolvability of the building blocks of ion-powered rotary motors

Pietro Ridone, University of New South Wales, Australia

Conserved ion-powered rotary motors (IRMs) drive bacterial flagellar motility by generating torque via ion transport. While IRMs have been structurally characterized in systems like Tol/Ton/Exb/Gld and Zor, functional stator chimeras (e.g., PotB from PomB/MotB) have mostly been designed arbitrarily and limited to closely related systems. Here, we engineered and tested chimeras of MotAB, PomAPotB, and ExbBD to explore cross-compatibility, generating motile strains with B-subunit chimeras. Directed evolution further enhanced motility, with genome sequencing revealing adaptive changes in the A-subunit and B-subunit peptidoglycan-binding domain. These findings underscore the complexity of IRM architecture and the challenges of rational chimera design.

PhyloPotts models for phylogenetics with dependent sites

Remco Bouckaert, University of Auckland, New Zealand

Standard phylogenetic methods based on amino acid sequences assume independence between sites. This is computationally convenient and suits the widely used Felsenstein's peeling algorithm, but ignores any structural dependencies. Potts models have been shown to be very well suited to capture pairwise interactions between sites in an alignment. These models are pure statistical models that can be trained on an amino acid alignment. To train the Potts model, no knowledge about the secondary or tertiary structure of proteins represented by the sequences in the alignment is required. In this talk, we explore how Potts models can be used to extend standard phylogenetic models with more realistic representation of dependence between sites.

Evolution of Antibiotic Resistance and Structure of the RNA Polymerase β Subunit

Samuel Ingloff-Richards, Australian National University, Australia

Rapidly emerging antibiotic resistance is posing a threat to health worldwide. Rifampicin, an antibiotic often used to treat tuberculosis (caused by Mycobacterium tuberculosis) has been one such antibiotic with increasing instances of rifampicin resistance occurring over the last decades. There are 57 amino acid changes which can occur in the structural core of the β subunit of RNA polymerase (RNAP) known to confer rifampicin resistance, but only 8% of living bacteria carry these resistance mutations, suggesting either resistance mutations are costly by perhaps reducing the efficiency of the RNAP transcription complex or low benefit. However, all known living bacteria carry an rpoB gene that has at least 25 single base changes across 19 sites predicted to confer rifampicin resistance. One possible explanation for this surprising observation is that evolution favours bacteria with rpoB genes that have a higher chance of mutating to produce a resistant phenotype. I will test if the rpoB sequences across a densely sampled phylogeny of living bacteria tend to have more possible single base changes conferring resistance than sequences sampled from their synonymous sequence spaces, and discuss how to use phylogenies to account for the non-independent sampling of sequences and structures. My investigation into rifampicin resistance will contribute to understanding the evolution of antibiotic resistance, and how selection may act without the need to perturb the protein sequence and structure.

Incorporating Flexibility and Site-Dependence into Bayesian Structural Phylogenetics

Scott C. Schmidler, Duke University, USA

Traditionally, evolutionary inference for homologous proteins relies primarily on stochastic models of amino acid sequence evolution. Bayesian structural phylogenetics methods supplement this by incorporating 3D structural information, and can significantly improve performance when sequences are highly diverged. Previously we have developed methods using RMSD-based alignment of the 3D structures. However, homologous proteins may sometimes acquire conformational changes during evolution, while others may exhibit conformational flexibility in solution, both of which are problematic for rigid-body comparisons. I will describe an expansion of our previously-developed model for protein structure evolution to accommodate backbone flexibility in alignment. This Bayesian flexible alignment model requires no prior knowledge of flexion points in the protein structure, can detect subtle flexion while avoiding overfitting, and is demonstrated to improve phylogenetic inference in a difficult-to-align set of proteins. I will also describe our recent work on phylogenetic inference under models with site-dependence, which can further improve realism of structural evolution models.

Leveraging structural predictions to understand virus glycoprotein evolution on the macro and micro-scale

Spyros Lytras, The Institute of Medical Science, University of Tokyo, Japan

Viral glycoproteins mediate membrane fusion, dictate host range and facilitate virus immune evasion. The proteins' important roles paired with the fast evolutionary rate of viruses lead to rapid accumulation of changes in their sequences, hindering homology detection over deep evolutionary time. Recent AI-based methods such as AlphaFold and ESMFold have enabled fast and accurate prediction of protein structures from sequence alone. On the macro-scale, we leverage these structural prediction methods at scale to infer the evolutionary histories of viral glycoproteins based on both sequence but also structural similarity between them. Applied to large virus taxa such as the Flaviviridae and the Coronaviridae, this approach allows us to uncover the wide diversity and complex evolutionary origins of these proteins. On the micro-scale, we can predict co-structures of glycoproteins together with host proteins they are expected to interact with, such as entry receptors. By assessing the prediction confidence of glycoprotein-receptor co-structures for closely related viruses and receptor orthologs we can infer the hosts that each virus strain can infect as well as the interacting residues controlling the ability to infect. These computational predictions are consistent with experimental validation of receptor-based virus infectivity, showcasing the usefulness of structural predictions in assessing the properties of novel viruses. Overall, structural predictions can be instrumental for uncovering both the deep scale evolution of virus glycoproteins, obscured by accumulated diversity, but also the phenotypic effects of even single mutations that disrupt protein-protein structural interactions.

Evolutionary trajectory and inhibition of ancestral metallo-β-lactamases

Ulban Adhikary, University of Queensland, Australia

Since the discovery of penicillin, the development of each new antibiotic drug has been quickly followed by the emergence of resistance to it. β-lactams account for approximately 65% of antibiotic prescriptions worldwide and consequently, the majority of pathogens on the WHO Bacterial Priority Pathogens list use β-lactamase enzymes as a primary mechanism of antibiotic resistance. Of particular concern are metallo-β-lactamases (MBLs), which have no known clinical inhibitors, yet confer resistance to four of the World Health Organization's top five critical priority pathogens.

However, the structural determinants of MBL activity remain a mystery due to their highly variable sequences and mechanistic flexibility - factors that complicate rational inhibitor design. In this study, we focus on the most clinically relevant MBL subgroups, B1 and B2, which - despite sharing a common ancestor - exhibit distinct catalytic and structural properties. Their divergent adaptations offer valuable insights for targeted inhibitor development.

Using ancestral sequence reconstruction, we generated ancestral MBL variants dating back to the divergence of these two subgroups. These enzymes serve as model representatives of their respective clades. Known mechanistic inhibitors, EDTA and L-captopril, produced inhibition constants comparable to those of extant enzymes. Additionally, we screened potential inhibitor compounds and a range of metal ions using a high-throughput plate assay to assess how structural differences influence MBL inhibition.

Structural Phylogeny of Flagellar FliF with Sporulation and T3SS homologs

Yogapriya Subramaniyan, University of Auckland, New Zealand

FliF is a transmembrane protein forming the scaffolding platform for the assembly of the bacterial flagellum. It forms a ring of ~34 subunits in the inner membrane, and serves as the pore through which the axial components are exported, as well as the base onto which the switch complex assembles (FliG, FliM, FliN), enabling rotating and direction-switching. Therefore, understanding the origins of FliF is important for further tracing the evolutionary origins of flagellum. A possible non-flagellar homolog of FliF was recently reported in the sporulation complex of Bacillus and relatives. The proteins SpoIIIAG, SpoIIIAH and SpoIIIAF form an inner membrane ring which serves as a scaffold onto which the other sporulation proteins assemble. The sporulation complex enables the translocation of substrate between the mother spore and the forespore. FliF has 3 ring-binding motif (RBM) domains, and each of the sporulation proteins has an RBM domain in their structure. In addition, the non-flagellar Type 3 Secretion System (NF-T3SS) protein, the SctJ group (including EscJ, PrgK, and others), also forms an inner membrane ring scaffolding for Type 3 export and has two RBM domains homologous to FliF. FliF/EscJ/PrgK and SpoIIIAF/G/H exhibit only barely detectable sequence similarity at the amino acid level. However, structural similarities seem to be more conserved. Hence, the structural analysis of the scaffold protein in these systems could shed more light on their evolution. In this study, we performed multiple sequence alignments of three-dimensional structural (3DI) protein characters and structural superpositions of scaffold proteins, and reconstructed phylogenetic evolution using both structural and amino acid data. Results suggest that SpoIIIA proteins can serve as an outgroup to help root FliF/SctJ trees, and that the SpoIIIAF/G/H structures can be aligned to specific parts of FliF, indicating possible evolutionary origin through duplication and fusion of smaller ring-forming proteins. Also, based on structural superposition and MSA analysis, we hypothesise that the ancestors of SpoIIIAH and SpoIIIAG would have fused to form the ancestors of FliF . Additionally, an outgroup containing an RBM, which is structurally homologous to RBM1, has been identified. This could be further analysed and used in further analysis.