29th Annual New Zealand Phylogenomics Meeting

Tuesday 10 Feb — Friday 13 Feb 2026

Onetangi Community Hall,

Waiheke Island, New Zealand

Programme & Abstracts

Organised by David Welch and Walter Xie (University of Auckland)

Programme

All talks, lunches and morning/afternoon teas are in the Onetangi Community Hall.

Tuesday 10 February

12:00 pm	**Arrival and Lunch**
1:35 pm	**Welcome**
1:40 pm	Simone Linz *Order-dependent dissimilarity measures between phylogenetic trees*
2:00 pm	Sophie Yang *Improved Phylogenetic Posterior Estimation through Regularised Conditional Clade Distributions*
2:20 pm	Shelby Cox *Tropical phylogenetics*
2:40 pm	David Swofford *Divergence time estimation under the multispecies coalescent model using MAP/composite likelihood: Extension to general models*
3:00 pm	**Afternoon Tea**
3:30 pm	Matthew Fullmer *Adding 3Di characters to amino acid datasets can improve deep resolution, but the effect is weaker in shorter and alpha-helical proteins*
3:50 pm	Nick Matzke *Validating the use of 3Di protein structure characters using LECA mitochondrial proteins*
4:10 pm	Alexei Drummond *Tractable Time-Tree Distributions*

Wednesday 11 February

9:30 am	Daniel Huson *Displacement optimized (DO) layout of phylogenetic networks and tanglegrams*
9:50 am	Mike Steel *A dichotomy law for certain classes of phylogenetic network*
10:10 am	Takatora Suzuki *Which Phylogenetic Networks are Level-k Networks with Additional Arcs? Struc- ture and Algorithms*
10:30 am	**Morning Tea**
11:10 am	Charles Semple *Counting trees in phylogenetic networks*
11:30 am	Russell Gray *Kava consumption and the rise of sociopolitical complexity in Oceania*
11:50 am	Scott C. Schmidler *Approximation Algorithms for Evolutionary Models with Site-Dependence*
12:10 pm	**Lunch**
2:00 pm	Andrea Maria Grecu *MutSimABC: Jointly estimating somatic mutation rates and meristem dynamics in long-lived trees*
2:20 pm	Marcus Overwater *The Calibrated Coalescent Point Process for Bayesian Molecular Clock Dating*
2:40 pm	Walter Xie *BEAST 2.8 - toward BEAST 3 a next generation Bayesian phylogenetic platform*
3:00 pm	**Afternoon Tea**
3:30 pm	Yuan Xu *Mixture Tree Likelihood*
3:50 pm	David Bryant *A Balanced approach to Likelihood*
4:10 pm	Jotun Hein *Recombinations: Algorithms and Recoverability*
6:30 pm	**Conference Dinner** Ki Māha at 1 Fourth Ave, Onetangi Beach

Thursday 12 February

9:30 am	Jasmine Saghafifar *Exploring Time-Nonreversibility and Site-Mixtures in Virus Phylogenetics with ABySS*
9:50 am	Remco Bouckaert *PhyloPotts models for phylogenetics with dependent sites*
10:10 am	Sinuo He *Dissecting Phylogenetic Instability in a Bayesian Framework*
10:30 am	**Morning Tea**
11:10 am	Natalia Mrnjavac *The primordial assembly of microbial metabolism*
11:30 am	Peter Wills *Phylogenetics and evolutionary causation*
11:50 am	Sishuo Wang *Host phylogeny-aware principal coordinate analysis (PCoA) reveals new patterns of microbiome evolution*
12:10 pm	**Lunch**
2:00 pm	Chris Simon *Evolution of 13- and 17-year periodical cicadas and how they decide when to emerge*
2:20 pm	Suyeon Kim *Evolution of Plasmid-carried Antibiotic Resistance Genes in a Fluctuating Envi- ronment*
2:40 pm	Yogapriya Subramaniyan *Structural Phylogeny of Flagellar FliF with Sporulation and T3SS homologs*
3:00 pm	**Afternoon Tea**
3:30 pm	Ming Yang *Copy number inference for single-cell cancer phylogenies*
3:50 pm	Eva Li *Assessing Bayesian Phylogenetic Inference on Single-Cell DNA Sequencing Tumour Data*

Friday 13 February

9:30 am	Emma Holvast *Inferring phylogeny using 3D morphometric data: a case study of bandicoot (Marsupialia: Peramelemorphia) postcrania*
9:50 am	Isaac Stead *“Total evidence” in language phylogenies: combining lexical data and sporadic sound change*
10:10 am	Yao Xiao *Joint inference of single nucleotide variants and cell phylogenies from single-cell DNA read counts*
10:30 am	**Morning Tea**
11:10 am	Mareike Fischer *Metaconcepts of rooted tree balance*
11:30 am	Lars Berling *Extended Clades and CCDS*
11:50 am	Sergey Bocharov *Typical edges in trees reconstructed from birth-death processes with Bernoulli killing*
12:10	**Conference ends (no lunch served)**

Abstracts (ordered by last name)

Lars Berling

(Simon Fraser University, lars_berling@sfu.ca)

Extended Clades and CCDS

The inference of phylogenetic trees is typically accompanied by uncertainty quantification, for example through sampling from a posterior distribution or by generating bootstrap replicates. However, how best to summarize and communicate the uncertainty contained in such samples remains poorly understood and is often treated only superficially. This uncertainty is commonly summarized using a single representative tree, which serves as a backbone onto which additional measures of uncertainty are mapped. Methods to create these representative trees have been restricted to “plain tree topologies”: trees encoding only rela- tionships among taxa and no additional structure or information. Inference of more sophisticated models produce annotated tree objects These have additional information representing nodes’ locations in the case of phylogeography, host information when inferring transmission trees, or sampled ancestor status when incorporating fossil information. Nevertheless, these enriched representations are often reduced to a single representative tree, typically using methods developed for plain topological trees and without accounting for the resulting methodological mismatch. Here, we introduce the concept of an extended clade and investigate an extension of the conditional clade distribution model. We show the use case of discrete trait phylogeog- raphy where we illustrate that the CCD method can resolve erroneous results that arise using a standard summary tree and overlaying the geographic annotation.

Sergey Bocharov

(Xi’an Jiaotong-Liverpool University, Sergey.Bocharov@xjtlu.edu.cn)

Typical edges in trees reconstructed from birth-death processes with Bernoulli killing

Consider a birth-death process run up to some time t. At time t every particle alive (if there are any) is killed with a probability p independently of all other particles. We are interested in the tree reconstructed from the surviving particles (if there are any). We shall discuss the limiting distribution of the lengths of edges chosen uniformly at random from all pendant and all interior edges as t tends to infinity.

Remco Bouckaert

(University of Auckland, r.bouckaert@auckland.ac.nz)

PhyloPotts models for phylogenetics with dependent sites

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.

David Bryant

(University of Otago, david.bryant@otago.ac.nz)

A Balanced approach to Likelihood

At last year's phylogenomics conference I talked about a new algorithm for computing the likelihood of a tree which could potentially reduce computation time from O(n) time per site to O(log n), at least in the worst case. In theory. This year I'll talk about how that works in practice. It is all about balance.

Shelby Cox

(Max Planck Institute for Math in the Sciences, shelby.cox@mis.mpg.de)

Tropical phylogenetics

Phylogenetic reconstruction often yields many competing trees, due to both biological and methodological variation. The space of phylogenetic trees is not convex, which complicates attempts to compare these recon- structions. By embedding trees in the tropically convex space trop M_0,n, we obtain a geometric framework for analyzing collections of trees on the same set of leaves. In this talk, I'll explain how tropical convexity enables the computation of weighted median trees, and outline future directions in tropical phylogenetics.

Alexei Drummond

(University of Auckland, a.drummond@auckland.ac.nz)

Tractable Time-Tree Distributions

Bayesian phylogenetic inference produces posterior distributions over time-trees, but compact, tractable parametric representations remain elusive. We compare two construction-based parameterisations for time- tree distributions: the Height-Ratio approach, which builds trees from root to tips by sampling node positions as proportions of the interval to the root, and the Shorter-Branch approach, which builds trees from tips to root by sampling the shorter branch length at each divergence. Height-Ratio accumulates error towards the tips but yields accurate node height marginals; Shorter-Branch accumulates error towards the root, producing overly wide tree height distributions. We evaluate variants using Beta, LogNormal, squashed LogNormal and scaled Beta distributions, with and without explicit dependencies on parent branch, node height, and tree height. Validation on simulated Yule trees (50–200 taxa) and 22 empirical PhyloData posteriors reveals a dilemma: Height-Ratio models achieve superior likelihood and point estimates, while Shorter-Branch models produce better Cramér–von Mises goodness-of-fit rankings. We discuss the trade-offs and implications for selecting tractable time-tree distributions.

Mareike Fischer

(University of Greifswald, mareike.fischer@uni-greifswald.de)

Metaconcepts of rooted tree balance

Measures of tree balance are widely used in areas such as mathematical phylogenetics and theoretical com- puter science to quantify structural properties of (rooted binary) trees. Traditionally, tree balance is de- scribed by a single numerical index, with many different balance and imbalance indices proposed in the literature. Most of these indices rely on closely related aspects of tree shape, including clade sizes, balance values, or leaf depths. In my talk, I will introduce a unifying framework for tree (im)balance measures based on metaconcepts of tree balance. A metaconcept is defined as a function Φ_f that aggregates a chosen tree shape characteristic captured by a function f. This perspective reveals that many classical indices arise as special cases of a common underlying construction. Beyond unification, metaconcepts generate new balance measures and enable the systematic study of entire families of indices. This approach provides a deep struc- tural insights into similarities and differences between balance measures. I will conclude my talk with an outlook on future research.

Matthew Fullmer

(University of Auckland, m.fullmer@auckland.ac.nz)

Adding 3Di characters to amino acid datasets can improve deep resolution, but the effect is weaker in shorter and alpha-helical proteins

The recent introduction of Foldseek's 3Di character alphabet to encode 3D protein structure has opened up new possibilities for structural phylogenetics. These characters, like protein structure, are more conserved than amino acids, raising the possibility of better resolution of very deep branches on the tree of life. As 3Di characters have a 20-letter alphabet, they are readily treatable with off-the-shelf algorithms for model-based phylogenetic inference and related methods such as bootstrapping. However, it remains to be seen if 3Di phylogenies are broadly more resolved than sequence-based phylogenies. We present data from superfamilies spanning all combinations of alpha-helix, beta-sheet, and mixed superfamilies showing that 3Di combines with sequence to produce better resolved phylogenies than either sequence or 3Di alone. We also show that information-theoretic measures, applied to superfamily alignments, significantly correlate with resolution in phylogenies derived from these alignments. Further, we identify the proportions and combination of alpha helices and beta sheets in proteins as a major driver in reducing the information carried by 3Di character alignments, explaining the relatively poor performance of 3Di characters on superfamilies with highly-conserved structure but high alpha helical content. Our results provide encouragement for the further use of 3Di to address challenging questions in deep history, but also sound a note of caution about which proteins it is most suitable for.

Russell Gray

(Max Planck Institute for Evolutionary Anthropology, rd.gray@auckland.ac.nz)

Kava consumption and the rise of sociopolitical complexity in Oceania

Humans have been using psychoactive substances for millennia, despite their potential negative health and social consequences. According to some scholars, our craving for mind-altering drugs is an evolutionary mistake – a hijacking of our reward system. In contrast, the "drunk hypothesis" argues that intoxication has been adaptive and essential for the rise of large-scale societies, because it promotes social bonding, increases cooperation, alleviates stress and enhances human creativity. Here we test this hypothesis using the example of kava, a traditional Pacific beverage with a range of psychoactive effects, made from the root of Piper methysticum. Our analysis of 83 Oceanic-speaking societies shows a positive relationship between traditional kava consumption and both political complexity and social stratification. However, the results are not robust to controls for non-independence. Moreover, we found no evidence of coevolution between kava drinking and either of the two sociopolitical traits, even after controlling for spatial non-independence. Despite the cultural significance of kava in many Pacific societies, our results suggest that its consumption was unlikely to have been a major driver of sociopolitical complexity, underscoring the importance of controlling for non-independence in cross-cultural studies.

Andrea Maria Grecu

(The University of Auckland, amgstar86@gmail.com)

MutSimABC: Jointly estimating somatic mutation rates and meristem dynamics in long-lived trees

Long-lived trees accumulate somatic mutations over centuries, forming genetic mosaics in which branches carry distinct genotypes. Accurately estimating mutation rates in these organisms is essential for understand- ing somatic evolution, yet current approaches face a fundamental challenge: meristem dynamics complicate the relationship between observed genetic variation and underlying mutation processes. Recent phyloge- nomic approaches leverage tree topology to estimate mutation rates by assuming genetic variation follows physical branching structure. However, meristem elongation and branching events redistribute mutations in ways that decouple genetic patterns from tree topology, making accurate somatic mutation rate infer- ence difficult. We developed MutSimABC, an Approximate Bayesian Computation (ABC) framework that uses simulation-based inference to jointly estimate mutation rate, meristem elongation dynamics (StD), and branching bias. Building on the Tomimoto and Satake (2023) mechanistic models, MutSimABC compares observed mutation distributions to simulated data without requiring topological assumptions or likelihood calculations. Validation using 169 simulated datasets demonstrated robust parameter recovery (99% recovery) and application to Eucalyptus melliodora genomic data yielded mutation rate estimates of 2.3–1.0 × 10 ¹ per site per year and revealed partially stochastic meristem behavior. Our approach enables mechanistic investigation of somatic evolution in plants and is generalizable to any long-lived plant with hierarchical branching.

Sinuo He

(University of Auckland, she922@aucklanduni.ac.nz)

Dissecting Phylogenetic Instability in a Bayesian Framework

Phylogenetic inference often needs to add new taxa into an existing analysis. However, adding a taxon is not simply adding a new leaf to a tree. A taxon may show uncertainty only in its own placement, or it may change the inferred relationships among other taxa. Existing definitions of rogue taxa mainly focus on placement uncertainty, but do not fully explain whether phylogenetic inference remains stable after a taxon is added or removed. In this study, we investigate taxon effects on phylogenetic stability under a Bayesian framework, where stability is defined as whether relationships among the remaining taxa change after a taxon is added or removed. For each dataset, we perform one full MCMC analysis and a series of leave-one-out analyses. We introduce three complementary entropy-based measures to describe a taxon's contribution to overall uncertainty, its placement uncertainty, and whether it induces structural differences between posterior distributions. Based on these measures, we construct a taxon classification framework distinguishing classic rogues, local disruptors, ambiguous neutral taxa, and stabilising taxa. This framework provides a unified perspective for dataset curation, sampling strategies, online phylogenetic inference, and for distinguishing placement uncertainty from true phylogenetic conflict.

Jotun Hein

(University of Oxford, hein@stats.ox.ac.uk)

Recombinations: Algorithms and Recoverability

The evolutionary process of genetic recombination has the potential to rapidly change the properties of a viral pathogen, and its presence is a crucial factor to consider in the development of treatments and vaccines. It can also significantly affect the results of phylogenetic analyses and the inference of evolutionary rates. The detection of recombination from samples of sequencing data is a very challenging problem and is further complicated for SARS-CoV-2 by its relatively slow accumulation of genetic diversity. The extent to which recombination is ongoing for SARS-CoV-2 is not yet resolved. To address this, we use a parsimony-based method to reconstruct possible genealogical histories for samples of SARS-CoV-2 sequences, which enables us to pinpoint specific recombination events that could have generated the data. We propose a statistical framework for disentangling the effects of recurrent mutation from recombination in the history of a sample, and hence provide a way of estimating the probability that ongoing recombination is present. We apply this to samples of sequencing data collected in England and South Africa and find evidence of ongoing recombination. We investigate the probability of recovering the true topology of ancestral recombination graphs (ARGs) under the coalescent with recombination and gene conversion. We explore how sample size and mutation rate affect the inherent uncertainty in reconstructed ARGs; this sheds light on the theoretical limitations of ARG reconstruction methods.

Emma Holvast

(Australian National University, emma.holvast@anu.edu.au)

Inferring phylogeny using 3D morphometric data: a case study of bandicoot (Marsupialia: Peramelemorphia) postcrania

Isolating phylogenetic signal from morphological data is crucial for accurately merging fossils into the tree of life and for calibrating molecular dating. However, subjective character definition is a major limitation which can introduce biases that mislead phylogenetic inferences and divergence time estimation. Here, we present the results of a systematic review describing the current state of the field in using quantitative (e.g., geometric morphometric; GMM) data for phylogenetic reconstruction and assessing the efficacy of these data compared to traditional discrete characters. We also present an application of novel methodology (Celik et al., in review) to discretise the shape variation in a GMM-based landmark dataset extracted from 3D scans of bandicoot humeri, subdividing the landmarks into subregions and coding each of these into character states using clustering (UPGMA, K-Means) methods. Phylogenies inferred under maximum parsimony based on k-means clustering for shape covariation with size (allometric signal) generally succeeded in grouping species at the genus level, with moderate support. Our case study demonstrates promise for improving objectivity and leveraging quantitative (i.e., GMM) data in morphological phylogenetics moving forward. Joint work with Mélina Celik, Matthew Phillips (both QUT) and Laura Wilson (ANU).

Daniel Huson

(University of Tuebingen, daniel.huson@uni-tuebingen.de)

Displacement optimized (DO) layout of phylogenetic networks and tanglegrams

The layout of phylogenetic networks and tanglegrams plays a crucial role in their interpretation, yet most existing drawing methods rely on heuristics that are difficult to compare, tune, or objectively evaluate. In this talk, I introduce Displacement Optimization (DO) as a unified, objective-driven framework for visualizing both phylogenetic networks and tanglegrams. The central idea is to quantify and minimize two intuitive contributors to visual complexity: reticulate displacement, which captures the added complexity caused when reticulation edges span large distances in a network, and taxon displacement, which measures inconsistencies in the relative ordering of taxa across trees or networks in a tanglegram. I will demonstrate the use of DO layout on several previously published phylogenetic networks and tanglegrams, showing how the method yields clearer and more faithful visualizations. Comparative analyses highlight that DO layout performs favorably against two widely used tools—the cophylo R package for trees and the Dendroscope program for networks—offering improved interpretability and a principled basis for assessing layout quality.

Suyeon Kim

(University of Auckland, skim404@aucklanduni.ac.nz)

Evolution of Plasmid-carried Antibiotic Resistance Genes in a Fluctuating Environment

Antibiotic resistance in bacteria is a global health concern. Theoretically, antibiotic resistance genes (ARGs) should be more prevalent on bacterial chromosomes than on plasmids, because plasmids impose fitness costs on their hosts. However, in nature, ARGs are more prevalent on plasmids, suggesting that plasmid-carried ARGs confer adaptive advantages that outweigh the costs. We hypothesise that this advantage arises from phenotypic switching: cells with plasmid-carried ARGs can rapidly gain or lose plasmids to switch between resistant and sensitive phenotypes, which is advantageous in environments where antibiotic presence fluctu- ates. In contrast, cells gain or lose chromosome-carried ARGs more slowly, which limits phenotypic switching and renders these cells disadvantageous in fluctuating environments. To test this switching hypothesis, we developed an ordinary differential equation model of bacterial population dynamics and competed cells with plasmid-carried versus chromosome-carried ARGs under fluctuating environments. Consistent with our hy- pothesis, the model showed that plasmid-carried ARGs are advantageous when antibiotics are infrequently present, whereas chromosome-carried ARGs are advantageous when antibiotics are almost always present. Unexpectedly, in intermediate fluctuating environments, both resistance types stably coexisted without com- petitive exclusion. This result suggests that the genomic location of ARGs is determined not only by fitness differences between resistance types but also likely by cell-cell interactions.

Eva Li

(University of Auckland, yil218@aucklanduni.ac.nz)

Assessing Bayesian Phylogenetic Inference on Single-Cell DNA Sequencing Tumour Data

Single-cell DNA sequencing (SCS) provides valuable insights into tumour heterogeneity and enables the reconstruction of tumour cell phylogenies; however, its high error rates present significant challenges. In this study, we aim to analyse the posterior distributions generated by Bayesian phylodynamic inference for SCS tumour data. Our work involves three experimental setups that use the same analysis pipeline but different simulation models: (1) a simple tree under a strict clock model, (2) a calibrated tree under a local clock model, and (3) a more complex tree under a local clock model. The key steps in our pipeline include: (i) simulating trees and sequence alignments using implemented LinguaPhylo (LPhy) models, (ii) evaluating how different variant calling parameters mitigate sequencing noise, and (iii) comparing different variant calling tools and result-mapping strategies for downstream analyses. This simulation pipeline is designed to generate realistic sequences, identify optimal variant-calling parameters to minimise noise, and determine more effective approaches for mapping variant called alignment with candidate sites, thereby improving the accuracy of tumour cell phylogenetic inference. Future work will extend this pipeline by applying the results to estimate phylogenetic features through Bayesian analysis while assessing the performance of error models.

Simone Linz

(University of Auckland, s.linz@auckland.ac.nz)

Order-dependent dissimilarity measures between phylogenetic trees

Several new vector representations for rooted phylogenetic trees were recently introduced that are based on imposing an ordering on their leaves. Hence, for a fixed tree and two different orderings on its leaves, the resulting vectors are typically different. Advantageously, comparing the vectors of two rooted phylogenetic trees leads to polynomial-time computable dissimilarity measure between them. In this talk, we review the three new vector representations HOP, ordered leaf attachment, and Phylo2Vec. We then compare them to several dissimilarity measures that are based on agreement forests and, disadvantageously, NP-hard to compute. This is joint work with Katherine St. John, Charles Semple, and Kristina Wicke.

Nick Matzke

(University of Auckland, n.matzke@auckland.ac.nz)

Validating the use of 3Di protein structure characters using LECA mitochondrial proteins

Three-dimensional protein structure characters from the FoldSeek program, known as 3Di characters, are more conserved than amino acid characters, but can be aligned and used in the same phylogenetics inference programs as amino acids. While previous work suggests that 3Di characters can improve resolution for ancient phylogenetic relationships at gigayear timescales, proving that accuracy is improved is very difficult, as we almost never have access to the true phylogeny at such timescales. We propose a test using mitochondrial proteins found in the Last Eukaryotic Common Ancestor (LECA). While the true phylogeny is unknown, it is plausible that these proteins share a common history. Therefore, if 3Di characters improve accuracy, the phylogenies of individual proteins should increase congruence with each other and with the all-proteins mitochondrial tree. We find significant evidence for this effect, and show that the improvement is stronger for less-conserved proteins.

Natalia Mrnjavac

(University of Auckland, N.Mrnjavac@hhu.de)

The primordial assembly of microbial metabolism

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, we get a glimpse into the primordial assembly of metabolism.

Marcus Overwater

(ETH Zurich, marcus.overwater@bsse.ethz.ch)

The Calibrated Coalescent Point Process for Bayesian Molecular Clock Dating

Bayesian molecular clock dating involves jointly inferring phylogenetic trees and divergence times from molecular sequences. When data consists of contemporaneous sequences, identifying absolute divergence times requires “calibrating” the molecular clock. This involves placing informative priors, usually derived from fossil data, on the ages of specific clades. However, two theoretical problems arise. First, calibration priors must be structurally consistent with the tree topology, ensuring that parent nodes are always older than their descendants. Second, the tree prior interacts with the calibration priors; without proper conditioning, the tree prior can distort the intended distribution of calibration ages. In this talk, I present a solution to the first problem using a generalized Dirichlet tree distribution to define calibration priors that are consistent with both fossil information and tree topology. For the second problem, I use the “Coalescent Point Process” representation of a general class of Birth-Death processes to efficiently compute the marginal probability of the ages of the calibration nodes. This results in the Calibrated Coalescent Point Process, a tree prior explicitly conditioned on the ages of calibration nodes.

Jasmine Saghafifar

(University of Auckland, jsag090@aucklanduni.ac.nz)

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

Modern genomic sequencing technologies mean we are now processing more sequences than ever, which corresponds with the rising importance of phylogenetic methods – for example, in contexts of epidemiology and public health during COVID-19 with contact tracing. Phylogenetic inference of nucleotide sequence evolution traditionally involves the use of a time-reversible substitution model (favoured for simplicity and small parameter-space), and optionally, manual partitioning by codon positions or genes. 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 substitution processes in the hopes of more realistic and effective inference. ABySS (Advanced Bayesian Site and Substitution) is a new software package with time-nonreversible substitution and site-mixture approaches. 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 empirical sequence data. I will discuss whether virus evolution modelled under time-nonreversible substitution would outperform time- reversible substitution models, and whether site-mixture can serve as an effective tool for substitution model selection and comparison, including bypassing the need for manual partitioning methods.

Scott C. Schmidler

(Duke University, Scott.schmidler@duke.edu)

Approximation Algorithms for Evolutionary Models with Site-Dependence

A major challenge in vaccine design is the elicitation of broadly-neutralizing antibodies (bnAbs) that protect against viral escape. Stochastic models of affinity maturation play a critical role in understanding barriers to success and guiding novel design strategies. A key difficulty is phylogenetic inference in the presence of site-dependent evolutionary processes, a long-standing challenge in computational biology. We describe a sequential Monte Carlo algorithm for approximating marginal sequence likelihoods under site-dependence and provide finite sample bounds on the approximation error. We extend this to a combined data-augmentation and importance sampling scheme for approximating posterior probabilities of phylogenetic trees and recon- structing ancestral sequences under site-dependent models of molecular evolution. We apply this approach to reconstruct maturation lineages from high-throughput B cell receptor repertoire sequencing data, and demonstrate the impact of accounting for context-dependence on reconstruction accuracy. We then apply our models to the problem of choosing targets for the design of boosting immunogens aimed at elicitation of bnAbs for HIV.

Charles Semple

(University of Canterbury, charles.semple@canterbury.ac.nz)

Counting trees in phylogenetic networks

The evolution of an individual gene is typically assumed to be tree-like, and so a phylogenetic network is often viewed as an amalgamation of gene trees. With this viewpoint, a natural question is the following: given a phylogenetic network N, how many (distinct) gene trees are embedded in N? If N has no reticulations, the answer is easy. But what if N has many reticulations? The question now is potentially much more challenging. In this talk we explore this question and discuss a recent result. This is joint work with Kristina Wicke.

Chris Simon

(University of Connecticut, chris.simon@uconn.edu)

Evolution of 13- and 17-year periodical cicadas and how they decide when to emerge

The 13- and 17-year periodical cicadas (Hemiptera: Cicadidae: Magicicada) are popular model organisms in evolution and ecology because they predictably emerge in large numbers, have a complex biogeography shaped by both spatial and temporal isolation, and include three largely sympatric, parallel species groups that are evolutionary replicates. Magicicada are also relatively easy to capture and manipulate and their spectacular, synchronized mass emergences facilitate outreach and citizen science opportunities. Over 2000 Magicicada studies have provided insights on habitat selection, resource partitioning/pulses, patchiness, climate change, habitat fragmentation, extinction, predator satiation, masting, competition, the evolution of periodicity, the Allee effect, lek mating systems, life history theory, selection for fecundity, latitudinal trends in body size, reproductive character displacement, allochronic and parallel speciation, speciation with gene flow, the evolution of endosymbionts and parasites, and more. I will highlight exciting new developments related to allochronic and parallel speciation, the evolution of 13- and 17-year periodical life cycles, and a field test of Sota's critical-weight/4-year-gate hypothesis where we studied growth and gene expression in 11-16-year-old last instar nymphs of two Magicicada species at multiple locations. Our results support this hypothesis.

Isaac Stead

(Max Planck Institute for Evolutionary Anthropology, isaac_stead@eva.mpg.de)

"Total evidence" in language phylogenies: combining lexical data and sporadic sound change

Mike Steel

(University of Canterbury, mathmomike@gmail.com)

A dichotomy law for certain classes of phylogenetic networks

Phylogenetic trees and networks are directed acyclic graphs that represent evolutionary relationships between species in biology and related fields. Numerous classes of phylogenetic networks have been defined and investigated. If a network in some class C is restricted to a subset of its leaves, then the resulting induced network may lie outside C; by contrast, certain other classes of networks are 'closed' when we restrict to subsets of leaves (e.g. the class of phylogenetic trees). It turns out that any closed subclass of the class of phylogenetic trees is either all trees or a vanishingly small proportion of them (as the number of leaves grows). In this talk, we use asymptotic enumeration techniques to explore whether this dichotomy phenomenon holds for more general classes of phylogenetic networks. This is joint work with Michael Fuchs.

Yogapriya Subramaniyan

(University of Auckland, ysub944@aucklanduni.ac.nz)

Structural Phylogeny of Flagellar FliF with Sporulation and T3SS homologs

FliF is a transmembrane protein forming the scaffolding platform for the assembly of the bacterial 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. 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, also forms an inner membrane ring scaffolding for Type 3 export and has two RBM domains homologous to FliF. 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.

Takatora Suzuki

(Waseda University, takatora.szk@fuji.waseda.jp)

Which Phylogenetic Networks are Level-k Networks with Additional Arcs? Structure and Algorithms

Phylogenetic networks represent reticulate evolution. Extracting their subgraphs is a well-studied approach to analyzing complex networks. Specifically, Francis and Steel (2015) introduced tree-based networks and their support trees. Hayamizu (2022) subsequently established a structure theorem, yielding optimal algo- rithms for various computational problems on support trees of rooted almost-binary networks. However, since networks are not necessarily tree-based, it is meaningful to consider support networks beyond trees. We present a generalization of the structure theorem to extend the theoretical framework to support net- works. A key result is a direct-product characterization of each of the three sets: all, minimal, and minimum support networks, for a given network. These characterizations yield optimal algorithms for counting and generating the support networks of each type. We also examine two optimization problems on the set of support networks: finding a support network with the fewest reticulations is linear-time solvable, whereas finding one with the minimum level is NP-hard. We address the latter with exact and heuristic algorithms that exhaustively search the sets of minimal and minimum support networks, respectively. Although both run in exponential time, they are practical for a wide range of reticulation numbers. Joint work with Momoko Hayamizu.

David Swofford

(Florida Museum of Natural History, dave.swofford@gmail.com)

Divergence time estimation under the multispecies coalescent model using MAP/composite likelihood : Ex- tension to general models

I present a new approach to the approximation of expected site-pattern frequencies under the Multispecies Coalescent model and its application to composite-likelihood/MAP estimation of divergence times. This permits the quartet-based method of Peng, Swofford, and Kubatko (2022, Bioinformatics 38:5182-5190) to be extended beyond the Jukes-Cantor model to any stochastic DNA substitution model. The new method works by computing the probability of all possible coalescent histories for each 4-tip gene tree, as well as the probabilities of the gene trees themselves. Expected branch lengths are obtained as averages of the branch lengths for each history, weighted by the probabilities of those histories. These branch lengths are used to obtain the probability of each of the 256 possible site patterns using any appropriate stochastic substitution model. Preliminary simulations show that (1) for the JC model, our method yields expected site-pattern probabilities almost identical to those obtained by analytical integration, and (2) estimated divergence times are dramatically less biased than those obtained from an incorrect substitution model. Furthermore, estimation of node ages using our method is much faster than Bayesian MCMC-based approaches such as BPP and StarBEAST. (joint work with Laura Kubatko)

Sishuo Wang

(The Chinese University of Hong Kong, sishuowang@hotmail.ca)

Host phylogeny-aware principal coordinate analysis (PCoA) reveals new patterns of microbiome evolution

While Principal Coordinate Analysis (PCoA) is a widely used dimensionality reduction method to analyze the (gut or skin) microbiota composition of the host, traditional applications overlook the influence of host phylogeny, where related hosts sharing a common evolutionary history likely exhibit more similar microbial communities. This violates the fundamental assumptions of independence inherent in standard statistical interpretations of PCoA. Through simulations, we show that standard PCoA tends to overestimate the variance explained by leading principal components, leading to spurious associations between microbiome and host traits driven by shared evolutionary history. Further, we introduce a novel, phylogeny-aware PCoA method that significantly improves estimation accuracy. Robust across diverse simulated parameters, this method reveals considerable alterations in results when applied to various animal microbiome data sets, compared to traditional PCoA approaches to microbiota composition analysis. Thus, results obtained from traditional PCoA analyses of microbiome data across (animal) hosts, which often neglect their phylogenetic non-independence, may need careful re-evaluation. This is a collaborative work with Youhua Chen and Qi Xiao.

Peter Wills

(University of Auckland, p.wills@auckland.ac.nz)

Phylogenetics and evolutionary causation

The supplementation of genetic sequences with other data characteristic of individual taxa, their components or particular properties, has significantly improved both the methodology of, and results obtained from, phylogenetic tree-building. For example, incorporation of protein structural data can improve the alignment of sequence data, leading to a more accurate representation of the elementary changes that potentially occurred along individual tree branches. That is not surprising, in that genetic mutation is not cause selection , it merely provides opportunity. The advantage gained from the functionality of the modified trait operating within the taxon's environment is the cause of selection. Thus, the better we can correlate trait functionality with sequence data, the better we will be able to deduce phylogenetic relationships. I will discuss what these observations teach us about using phylogenetics to uncover the evolutionary development of biological functionality, paying special attention, as usual, to emergence of genetic coding during the origin of life.

Ming Yang

(University of Auckland, myan486@aucklanduni.ac.nz)

Copy number inference for single-cell cancer phylogenies

Tumours evolve through the accumulation of genetic alterations, including single-nucleotide variants (SNVs) and copy-number variants (CNVs), generating substantial intra-tumour heterogeneity. Single-cell sequencing enables this heterogeneity to be studied at high resolution, but robust phylogenetic inference requires models that account for distinct evolutionary processes and technical errors across data types. Here, we implement the Nested birth–death (NestedBD) model for CNV evolution within the LPhy–BEAST workflow, enabling Bayesian phylogenetic inference directly from single-cell copy-number profiles using a unified specification for simulation and inference. Through well-calibrated simulation studies, we show that NestedBD accurately recovers tree topologies, branch lengths, and evolutionary parameters, validating its suitability for CNV- based analyses. We then combine CNV-based inference with SNV data modelled using the GT16 model to explore joint SNV–CNV phylogenetic inference. This framework allows us to assess how integrating complementary genomic signals influences phylogenetic reconstruction relative to single-data-type analyses. Finally, we apply these methods to single-cell colorectal cancer data to illustrate how different classes of genomic alterations contribute to reconstructing tumour evolutionary history. We anticipate that these developments will support efforts to characterise tumour evolution and advance our understanding of cancer heterogeneity at single-cell resolution.

Yao Xiao

(University of Auckland, yxia415@aucklanduni.ac.nz)

Joint inference of single nucleotide variants and cell phylogenies from single-cell DNA read counts

With the development of single-cell sequencing technology, phylogenetics can be applied to cell and devel- opmental biology. For example, during development, cell variations can be viewed as markers of evolution. Consequently, evolutionary models offer new perspectives for understanding somatic development and cancer evolution. However, most existing methods handle data filtering, variant calling, and phylogenetic inference separately. This workflow does not fully leverage the information in the data, potentially leading to biased results. To address these issues, we propose a Bayesian inference–based model that simultaneously performs variant calling and phylogenetic tree inference directly from single-cell DNA sequencing read counts. Our model has the following features: it considers all 16 possible diploid genotypes and accounts for both sequenc- ing errors and allele dropout that may occur during sequencing. Our research aims to provide a powerful new tool for studying cancer evolution and developmental biology.

Walter Xie

(University of Auckland, walter@cs.auckland.ac.nz)

BEAST 2.8 - toward BEAST 3 a next generation Bayesian phylogenetic platform

BEAST 2.8 marks a transitional step toward BEAST 3, a next-generation Bayesian phylogenetic platform designed around the core principles of statistical correctness, computational efficiency, and extensibility. Building on a mature inference engine, the roadmap emphasises rigorous validation and testing, reliance on well-established third-party libraries, and improved default operators, such as Bactrian operators. BEAST 3 further aims to provide strong typing and full compatibility with PhyloSpec, which is not only a modelling language but also a growing ecosystem of interoperable components intended to bring the phylogenetic community together. The platform is designed to accommodate an expanding range of data types, spanning molecular, morphological, geographic, linguistic, and single-cell data, while also enabling alignments to be treated as random variables. Together, these developments, paired with the adoption of Java 25 to improve robustness and maintainability, position BEAST 3 as a reliable, scientist-centred platform that prioritizes usability, reproducibility, and accessibility.

Yuan Xu

(University of Auckland, yxu927@aucklanduni.ac.nz)

Mixture Tree Likelihood

Choosing an appropriate molecular clock model is a key determinant of phylogenetic time-scale and rate estimates, yet model comparison in Bayesian phylogenetics commonly relies on marginal-likelihood estima- tors (e.g., stepping-stone or path sampling) that require multiple long MCMC analyses. We present a mixture-likelihood framework that embeds competing clock models within a single Bayesian analysis, yield- ing posterior support for each model while jointly estimating trees, divergence times, and rate parameters. The method represents the likelihood as a finite mixture over clock-specific components and infers mixture weights as posterior model support, allowing strict, uncorrelated lognormal relaxed (UCLN), and a newly implemented autocorrelated relaxed clock to be evaluated simultaneously. In simulations, the generating clock model is reliably recovered: across 100 replicates, the true model is contained in the 95% posterior credible set over clock models in >90% of analyses. We further illustrate the approach on empirical datasets, where a single run provides model-fit rankings and clock-averaged posterior estimates, reducing reliance on repeated runs and separate marginal-likelihood calculations. By integrating model choice into the posterior, the framework quantifies uncertainty in clock assumptions and propagates it to downstream divergence-time inference. This framework offers an efficient route to clock-model selection and model averaging in BEAST2 and is readily extensible to additional evolutionary model classes.

Sophie Yang

(University of Auckland, zyan598@aucklanduni.ac.nz)

Improved Phylogenetic Posterior Estimation through Regularised Conditional Clade Distributions

Bayesian phylogenetic inference uses Markov chain Monte Carlo sampling to estimate the posterior distribu- tion of phylogenetic trees. Due to the high-dimensionality and non-Euclidean nature of tree space, charac- terising tree distributions is inherently difficult. Traditional approaches often summarise posterior samples into a single point estimate, thereby discarding much of the information contained in the full distribution. We introduce a new model called regularised conditional clade distribution (regCCD) that addresses this limitation by combining the strengths of existing CCD parameterisations. We show that regCCD provides a better point estimate than the non-regularised CCD and outperforms the CCD0 and CCD1 parameteri- sations in estimating the overall posterior distribution. We further demonstrate its application in assessing whether different phylogenetic models or datasets produce statistically distinguishable tree distributions.

Participants

Lars Berling	Simon Fraser University
Sergey Bocharov	Xi'an Jiaotong-Liverpool University
Remco Bouckaert	University of Auckland
David Bryant	University of Otago
Ian Connolly	University of Auckland
Shelby Cox	Max Planck Institute for Math in the Sciences
Alexei Drummond	University of Auckland
Mareike Fischer	University of Greifswald
Andrew Francis	University of New South Wales
Matt Fullmer	University of Auckland
Alexandra Gavryushkina	Adelaide University
Russell Gray	Max Planck Institute for Evolutionary Anthropology
Andrea-Maria Grecu	University of Auckland
Simon Greenhill	University of Auckland
Sinuo He	University of Auckland
Jotun Hein	University of Oxford
Emma Holvast	Australian National University
Daniel Huson	University of Tuebingen
Suyeon Kim	University of Auckland
Toby Koch	Colorado State University
Eva Li	University of Auckland
Simone Linz	University of Auckland
Nicholas Matzke	University of Auckland
Natalia Mrnjavac	Heinrich Heine University Düsseldorf
Masafumi Obara	University of Auckland
Marcus Overwater	ETH Zurich

Alethea Rea	Murdoch University
Jasmine Saghafifar	University of Auckland
Scott Schmidler	Duke University
Charles Semple	University of Canterbury
Christine Simon	University of Connecticut
Isaac Stead	Max Planck Institute for Evolutionary Anthropology
Mike Steel	University of Canterbury
Yogapriya Subramaniyan	University of Auckland
Takatora Suzuki	Waseda University
David Swofford	Florida Museum of Natural History
Nobuto Takeuchi	University of Auckland
Sishuo Wang	The Chinese University of Hong Kong
David Welch	University of Auckland
Peter Wills	University of Auckland
Yao Xiao	University of Auckland
Walter Xie	University of Auckland
Yuan Xu	University of Auckland
Ming Yang	University of Auckland
Sophie Yang	University of Auckland