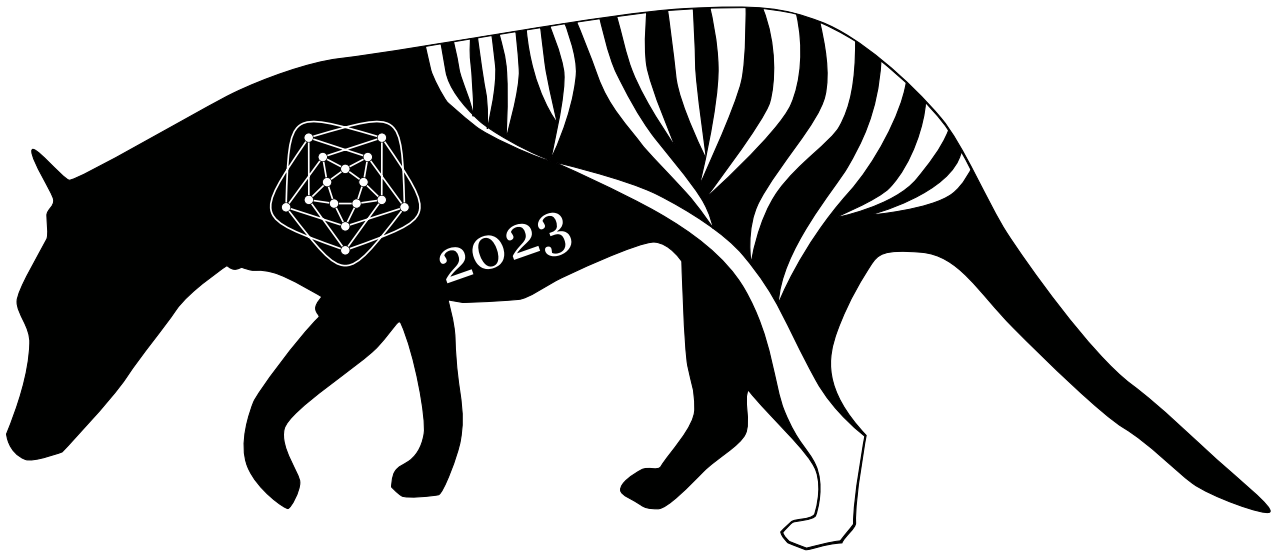
Phylomania 15
Nearest Neighbours

Hobart
November 22-24
2023



Conference Programme

Session Guide

		9:00-9:20	Acknowledgement of Country and Opening	
WEDNESDAY	Deep Time	9:20-10:00	A Roger^{P}	New phylogenetic models of amino acid sequence evolution for probing the relationships between prokaryotes and eukaryotes.
		10:00-10:20	Y Lang^{P}	How far into the past can we see?
		10:20-10:40	J Douglas^{P}	Bayesian phylokinetics: ancestral protein reconstruction and functional characterisation in a Bayesian framework
		10:40-11:10	Morning Tea	
	Macroevolution	11:10-11:30	S Blomberg^{P}	Is a general mathematical theory of macroevolution possible?
		11:30-11:50	A Butryn^{P}	Developing a novel statistical model of morphological evolution that accounts for unobserved but observable character states
		11:50-12:10	G Jordan^{P}	Trees (phylogenetic) and leaves (green not graph)
		12:10-1:30	Lunch	
	Algorithms	1:30-1:50	M Charleston^{P}	The Phylogenetic Utility of Shared k-mers (PHUSK)
		1:50-2:10	Y-B Chan^{P}	A probabilistic algorithm for gene-species reconciliation with segmental duplications
		2:10-2:30	W He^{P}	The accuracy of species tree inference under gene tree dependence
		2:30-2:50	S Tule^{P}	Optimal Phylogenetic Reconstruction of Insertion and Deletion Events
		2:50-3:20	Afternoon Tea	
	Networks	3:20-3:40	A Francis^{P}	A new way to encode phylogenetic network classes
		3:40-4:00	H Banos^{O}	TINNIiK: An algorithm to Infer the Tree of Blobs of a Species Network Under the Coalescent
		4:00-4:20	K Wicke^{O}	The weighted total cophenetic index: A novel balance index for phylogenetic networks
		4:20-4:40	Y Murakami^{O}	Proximity measures for phylogenetic networks
THURSDAY	Algebra	8:40-9:00	R Homs Pons^{O}	A novel algebraic approach to time-reversible evolutionary models
		9:00-9:20	R Yoshida^{O}	Tropical Neural Networks for classifying phylogenetic trees
		9:20-9:40	J Sumner^{P}	Intertwiners for phylogenetic branching operations
		9:40-10:00	J Stevenson^{P}	Rearrangement distances on circular genomes
		10:00-10:30	Morning Tea	
	Models and Algorithms	10:30-10:50	A Soewongsono^{O}	A Diffusion-Based Approach for Simulating Forward-in-Time State-Dependent Speciation and Extinction Dynamics
		10:50-11:10	M O'Reilly^{O}	Markovian Binary Trees: Likelihood of Reconstructed Species Trees
		11:10-11:30	C Elgert^{P}	From scRNA-seq count matrix to phylogenetic inference
		11:30-11:50	R McArthur^{P}	Spectral Cluster Supertree: A fast and accurate supertree method for rooted trees
		12:00-1:20pm	Lunch	Poster Session
	Applications	1:20-1:40	A Macdonald^{P}	Can phylogenetics map traits to genes? Developing PhyloG2P Methods
		1:40-2:00	L Yates^{P}	Phylo-transcriptomics: a new method for analysis of gene-expression data across multiple related species.
		2:00-2:20	M Fourment^{P}	torchtree: phylogenetic inference using gradient-based algorithms
		2:20-2:40	B Halliwell^{P}	Multi-Response Phylogenetic Mixed Models: New Horizons for Conservative Trait Correlation
		2:40-3:10	Afternoon Tea	
	Eclectica	3:10-3:30	C L Jimenez Silva^{P}	Bayesian Assessment of Marsupial Phylogenomics and Calibrations Consistency
		3:30-3:50	H Nigus^{P}	Simulation of conservation management strategies: Trade-offs between genetic diversity and extinction probability
		3:50-4:10	R C Appaw^{P}	What makes a spatial network? Leveraging machine learning for the robust prediction and classification of networks
		evening	Conference Dinner	
FRIDAY	Models and Selection	9:20-9:40	N Fountain-Jones^{P}	Antivirals can supercharge SARS-CoV-2 evolution in immunocompromised patients
		9:40-10:00	J Mitchell^{P}	Convergence-Divergence Models: Alternatives to Phylogenetic Networks
		10:00-10:20	B Holland^{P}	Are AIC and BIC fit for purpose with complex phylogenetic models?
		10:20-10:50	Morning Tea	
		10:50-11:50	Panel discussion^{P,O}	
		11:50-12:00	Awards and Closing Remarks	
		12:00-1:00	Lunch	
		afternoon	Excursion	

^{P} $\Rightarrow$ In Person; ^{O} $\Rightarrow$ Online; * $\Rightarrow$ Student Presentation

Abstracts

What makes a spatial network? Leveraging machine learning for the robust prediction and classification of networks (Student presentation)

Raima Carol Appaw (He), UTAS-Maths/Physics Department
raima.appaw@utas.edu.au

Eclectica

The ability to simulate realistic networks based on empirical data is an important task across scientific disciplines, from epidemiology to computer science. Often simulation approaches involve selecting a suitable generative model such as Erdös-Rényi or small-world. However, there are few tools available to quantify if a particular generative model is suitable for a given network. We utilize novel advances in interpretable machine learning to interrogate our model and map how network features interact to shape these predictions.
Moreover, we extend this classification framework to empirical network data spanning various animal taxonomic classes, shedding light on the social dynamics of different animal species. Our study underscores the significance of specific network attributes, their interdependencies, and their role in distinguishing network generative models and comprehending real-world networks

TINNIiK: An algorithm to Infer the Tree of Blobs of a Species Network Under the Coalescent

Hector Banos (He/his/him), California State University San Bernardino
hector.banos@csusb.edu

Networks

Joint work with John A. Rhodes - University of Alaska Fairbanks, Elizabeth S. Allman - University of Alaska Fairbanks, Jonathan Mitchell - University of Tasmania

The tree of blobs of a species network shows only the tree-like aspects of relationships of taxa on a network. Building on the theoretical work on the identifiability of the tree of blobs from gene quartet distributions under the Network Multispecies Coalescent model, we develop, TINNiK, a fast statistically consistent algorithm for tree of blobs inference. TINNiK has been implemented in the MSCquartets R package. By using TINNiK to estimate the tree of blobs, empiricists then could determine subsets of taxa to explore where hybridization occur independently.

Is a general mathematical theory of macroevolution possible?

Simone Blomberg (she/her/hers), School of the Environment, University of Queensland
s.blomberg1@uq.edu.au

Macroevolution

Two of the main components of macroevolutionary theory are the study of diversification: the branching pattern of species evolution, and the study of disparity: how and why species look different. These components exist as almost entirely separate fields of study. Yet, a full theory of macroevolution should be able to account for both types of phenomena. I propose that measure-valued branching Markov processes (superprocesses) may be mathematical objects that could form the basis of such a theory, and hence have the potential to unify the field. I give a brief, amateur, introduction to superprocesses in the hope that some mathematicians in the audience might be interested enough to want to collaborate on building such a general mathematical theory.

Developing a novel statistical model of morphological evolution that accounts for unobserved but observable character states (Student presentation)

Alexander Butryn, Newcastle University - School of Mathematics, Statistics and Physics
a.butryn2@newcastle.ac.uk

Macroevolution

Phylogenetic methods and evolutionary models play an important role in palaeontology, allowing us to understand morphological character evolution and reconstruct the interrelationships of extinct organisms in deep time. The Mk model (Lewis, 2003) is extensively used in phylogenetic analyses of palaeontological datasets to model the evolution of discrete morphological characters, where transitions among character states follow a Markov process. This model is popular as it requires only a single parameter for a fixed tree, and holds a large amount of explanatory power while retaining statistical tractability.

Nevertheless, due to its simplicity, the Mk model fails to capture potentially important aspects of the evolutionary process, and contains hidden assumptions that may result in biased outputs of phylogenetic analyses. One such assumption is that the number of states that any particular character can transition between is limited by the number of states observed in the dataset. This assumption is unrealistic as it fails to consider the incomplete nature of the fossil record, biases in taxon sampling or the contingent nature of evolution. Consequently, the ubiquitous reliance on the Mk model without understanding the impact of model misspecification may pose a concern to the credibility of the results and conclusions of such studies.

We propose a new, alternative model of morphological character evolution, which we call the Mk' model, that improves on the Mk model by accounting for theoretically observable but empirically unobserved character states. This is achieved with the inclusion of a single parameter that governs the rate of transitions to an additional state that represents all unobserved states. Alongside model development, we also conduct phylogenetic analyses of empirical and simulated datasets under various models of evolution to examine the effects of, and the Mk model's robustness to, model misspecification regarding the number of character states.

A probabilistic algorithm for gene-species reconciliation with segmental duplications

Yao-ban Chan, School of Mathematics and Statistics / Melbourne Integrative Genomics, The University of Melbourne
yaoban@unimelb.edu.au

Algorithms

Joint work with Michael Charleston

The phylogenetic trees of species can be distinct from the trees of their genes, due to various evolutionary processes that affect genes but do not create new species. Reconciliations map the gene trees into the species tree, explaining the discrepancies by events including gene duplications, losses, and horizontal transfers. Of particular interest is segmental duplications, where several genes may duplicate in a single event. We therefore seek a most parsimonious reconciliation of a set of gene trees into a species tree accounting for segmental duplications, a known NP-hard problem. In this talk, we describe a probabilistic approach to solve this problem: we first assign a Boltzmann probability distribution on the reconciliation space, then use simulated annealing to recover the optimal reconciliation. This approach holds great promise for the future resolution of theoretically difficult reconciliation models.

The Phylogenetic Utility of Shared k-mers (PHUSK)

Michael Charleston, University of Tasmania
michael.charleston@utas.edu.au

Algorithms

I present a fast method to determine whether a given set of unaligned sequences appears to be phylogenetically informative, *Phusk* for Phylogenetic utility of shared k-mers.

The method runs in linear time and can be used to detect broad signals of phylogenetic usefulness of a data set, by comparing the distribution of k-mers across sets of sequences with those expected under simple models of evolution.

Bayesian phylokinetics: ancestral protein reconstruction and functional characterisation in a Bayesian framework

Jordan Douglas (he/him), Department of Physics, University of Auckland
jordan.douglas@auckland.ac.nz

Deep Time

Joint work with Remco Bouckaert, Charles Carter Jr, Peter Wills

Ancestral sequence reconstruction is the task of inferring ancestral nucleic acid or protein sequences using phylogenetic methods. This approach is often followed up by experimental characterisation to infer the biological function of the ancestral gene. The latter task is time consuming and resource intensive and therefore experimental biologists typically do not wish to test more than one or two ancestral constructs (e.g. the maximum likelihood estimate), even knowing that this point-estimate is most certainly wrong.
In this talk we will discuss some of the challenges in inferring biologically realistic protein sequences that fold into stable protein structures. These challenges include the choice of site model and the handling of insertions and deletions.
Lastly, we will explore the notion of ancestral enzyme characterisation in a Bayesian framework. An idealised Bayesian approach here would involve the functional characterisation of a large posterior sample of ancestral genes (rather than a point estimate) and thus offer greater robustness and statistical rigor to the overall experiment. In order to avoid an intractable cost in performing these experiments, we consider the possibility of an expression library: one which co-expresses, co-purifies, and co-assays an entire population of enzymes. The kinetic activities of these enzymes are characterised as a non-homogenenous mixture of enzymes using a specialised kinetic model. This setup would avoid the need to run many experiments in parallel.
This approach is Bayesian phylokinetics - a Bayesian statistical framework from the computer to the test-tube.

From scRNA-seq count matrix to phylogenetic inference

Christiane Elgert, Center for Integrative Bioinformatics Vienna (CIBIV), Max Perutz Labs, University of Vienna and Medical University of Vienna, Vienna BioCenter, Vienna, Austria
christiane.elgert@univie.ac.at

Models and Algorithms

Joint work with Julia Naas, Arndt von Haeseler: Center for Integrative Bioinformatics Vienna (CIBIV), Max Perutz Labs, University of Vienna and Medical University of Vienna, Vienna BioCenter, Vienna, Austria

In molecular phylogenomics one infers a phylogenetic tree for a group of species using genetic information from a multiple sequence alignment that provides information about the evolutionary relationships between the species. But what information lies in-between the characteristics of different species and their carrier of the genetic information, the genome? Can we use the same assumptions and methods from phylogenomics for other kinds of application and genetic information?
In recent decades, there has been increased interest in reconstructing high-resolution developmental lineages (trajectories) based on single-cell RNA sequencing (scRNA-seq) data. While one approach involves experimental efforts to store information about common ancestors in a cell's transcriptomic information, numerous bioinformatic tools have been developed to construct lineages and trajectories based on a cell's gene expression profile. Such tools are highly dependent on the embedding space and distance metric between pairs of cells and the quality and stability of their results are not accessible due to the missing standardization of different pipeline steps, e.g. preprocessing and embedding.
Therefore, we suggest rethinking lineage reconstruction by not developing new algorithms but, on the contrary, recoding scRNA-seq data to fit already well-established phylogenomic tools. Using different simple discretization procedures, we convert the number of transcripts into a multiple-sequence alignment, that serves as input into any preferred tree inference software, e.g. IQ-Tree. Exemplary reconstructed trees from scRNA-seq data, displaying the relationship between single cells instead of species, already resemble clustering annotation attained in scRNA-seq downstream analysis. Furthermore, we can use the distances of cells on the reconstructed tree as a new metric for subsequent downstream analysis, visualisation and changes in gene expression along the tree.

Antivirals can supercharge SARS-CoV-2 evolution in immunocompromised patients

Nick Fountain-Jones, Biological Sciences, University of Tasmania
nick.fountainjones@utas.edu.au

Models and Selection

Joint work with Robert Vanhaeften[1], Jan Williamson[1], Janelle Maskell[1], I-Ly J Chua[1], Michael Charleston[2] & Louise Cooley[1,3]

[1] *Royal Hobart Hospital, Pathology Department, Hobart Australia;* [2] *School of Natural Sciences, University of Tasmania, Hobart Australia;* [3] *School of Medicine, University of Tasmania, Hobart Australia.*

Continued SARS-CoV-2 infection among immunocompromised individuals likely plays a role in generating genomic diversity and the emergence of novel variants. While antiviral treatments such as Molnupiravir are employed to mitigate severe COVID-19 outcomes, the extended effects of these drugs on viral evolution in patients with chronic infections remain uncertain. This study investigates how antiviral treatments influence viral evolution in immunocompromised patients with prolonged infections. The study included five patients treated and four untreated with Molnupiravir with similar background conditions. Samples were collected in patients up to 44 days post-treatment and were sequenced using tiled amplicon sequencing followed by variant calling. The UShER pipeline and UCSC genome viewer provided insights into the global context of variants. Comparisons between treated and untreated patients were conducted, and mutation profiles were visualized to understand the impact of Molnupiravir on viral evolution.Patients treated with Molnupiravir exhibited a large increase in low-to-mid frequency variants in as little as 10 days after treatment, while no such change was observed in untreated patients. Importantly, some of these variants became fixed in the viral population including non-synonymous mutations in the spike protein. The variants were distributed across the genome and included unique mutations not commonly found in global Omicron genomes. Notably, G-to-A and C-to-T mutations dominated the mutational profile of treated patients, persisting up to 44 days post-treatment. The study revealed that Molnupiravir treatment in immunocompromised patients led to the accumulation of a distinctive pattern of mutations far beyond the recommended five days of treatment. As our treated patients maintained persistent PCR positivity for the duration of monitoring, there is clear potential for transmission and subsequent emergence of novel variants.

torchtree: phylogenetic inference using gradient-based algorithms

Mathieu Fourment, University of Technology Sydney
mathieu.fourment@uts.edu.au

Applications

Variational Bayes (VB) offers an alternative approach for Bayesian inference that better handles large datasets than the widely used MCMC algorithm. A key challenge in VB lies in efficiently computing the gradients of complex phylogenetic models. In this context, I introduce torchtree, a Python-based probabilistic framework designed for phylogenetic inference through variational methods and other gradient-based algorithms like Hamiltonian Monte Carlo. Although this framework use automatic differentiation, I'll show that it can be easily extended using C++ extensions, providing a significant speedup.

A new way to encode phylogenetic network classes

Andrew Francis (he/him/his), Western Sydney University
a.francis@westernsydney.edu.au

Networks

The set of forests of phylogenetic trees is in bijection with the set of partitions of finite sets, but is there such a set-theoretic bijection for phylogenetic networks? In this talk I will describe such a bijection between a large class of networks, called 'labellable', and sets of covers of finite sets that are 'expanding'. This new class of networks contains most well studied classes, and can be structurally characterised in graph-theoretic terms. It turns out that many familiar classes of network can also be characterised by properties of expanding covers. This opens questions of whether we can use covers to study networks in a new way.

Multi-Response Phylogenetic Mixed Models: New Horizons for Conservative Trait Correlation

Ben Halliwell, Biological Sciences, University of Tasmania
benjamin.halliwell@utas.edu.au

Applications

Joint work with Barbara Holland, School of Natural Sciences, University of Tasmania, Sandy Bay, Tasmania, Australia;
Luke Yates, ARC Centre of Excellence for Plant Success in Nature and Agriculture, University of Tasmania, Sandy Bay, Tasmania, Australia;
Mark Westoby, School of Natural Sciences, Macquarie University, Sydney, New South Wales, Australia

Modern phylogenies and species trait databases permit many open questions in comparative biology to be addressed with the right statistical tools. Multi-response (MR) phylogenetic mixed models (PMM) offer great flexibility, including support for multi-level meta-analyses and non-gaussian response variables, but are rarely employed, reflecting a technical literature that creates barriers to usage for many biologists. We present an accessible review of MR-PMM integrating recent methodological developments, discuss the biological motivation and interpretation of model parameters, and highlight an emerging synthesis of approaches made possible by this powerful and underutilised model class.

The accuracy of species tree inference under gene tree dependence (Student presentation)

Wanting He (she/her/hers), School of Mathematics and Statistic, University of Melbourne
wahe1@student.unimelb.edu.au

Algorithms

Joint work with Yao-ban Chan, Celine Scornavacca

When inferring the evolutionary history of species and the genes they contain, the phylogenetic trees of the genes can be different to that of the species and to each other, due to a variety of causes such as incomplete lineage sorting. We often wish to infer the species tree from an input set of gene trees; methods to do this are known as summary methods, of which ASTRAL is the currently among the most popular. ASTRAL has been shown to be practically accurate through extensive simulations. However, these simulations always assume that the input gene trees are independent of each other. This is known to be unrealistic, as genes that are close to each other on the chromosome (or are related by function) have dependent phylogenies, due to the absence of unlimited recombination between the genes.
We develop a model for generating dependent gene trees within a species tree based on the coalescent with recombination. We then use these trees as input to ASTRAL to reassess its accuracy for dependent gene trees. Our results show that ASTRAL performs worse with higher dependence, both when gene trees are known and estimated from sequences. Indeed, the effect of dependence between gene trees is comparable to (if not larger than) the effect of gene tree estimation error. We also estimate a realistic recombination rate by re-analysing a 37-taxon mammalian data set. Under this recombination rate, the estimated accuracy of ASTRAL is about half of the accuracy previously estimated with independent gene trees, and the effective sample size for this dataset is about one-third of the actual sample size. This shows that the accuracy of ASTRAL has been significantly overestimated by previous studies.

Investigating the performance of information criteria in choosing between partition and mixture models in phylogenetics.

Barbara Holland, University of Tasmania
barbara.holland@utas.edu.au

Models and Selection

Joint work with Qin Liu (UTAS) and Stephen Crotty (U Adelaide)

I will report on some recent results of Qin's following up on a question that arose out of work with Stephen Crotty on trying to decide if we should trust IC in picking between partition models and mixture models (or at all, fact).

A novel algebraic approach to time-reversible evolutionary models

Roser Homs Pons (she/her/hers), Centre de Recerca Matemàtica
rhoms@crm.cat

Algebraic Methods

Joint work with Marta Casanellas (Universitat Politècnica de Catalunya), Angélica Torres (Max Planck Institute for Mathematics in the Sciences)

In the last years algebraic tools have been proven to be useful in phylogenetic reconstruction and model selection by means of the study of phylogenetic invariants. However, up to now, the models studied from an algebraic viewpoint are either too general or too restrictive (as group-based models with a uniform stationary distribution) to be used in practice. In this talk we provide a new framework to work with time-reversible models, which are the most widely used by biologists. In our approach we consider algebraic time-reversible models on phylogenetic trees (as defined by Allman and Rhodes) and introduce a new inner product to make all transition matrices of the process diagonalizable through the same orthogonal eigenbasis. This framework generalizes the Fourier transform widely used to work with group-based models and recovers some of the well known results. As illustration, we exploit the combination of our technique with algebraic geometry tools to provide relevant phylogenetic invariants for trees evolving under the Tamura-Nei model of nucleotide substitution.

Trees (phylogenetic) and leaves (green not graph)

Greg Jordan, Biological Sciences, University of Tasmania
greg.jordan@utas.edu.au

Macroevolution

Joint work with Katya Bandow, Ben Halliwell, and Barbara Holland (UTAS)

We discuss the application of Bayesian phylogenetic mixed models to investigate the relationships between leaf size and climate in conifers. These models provide a substantial advance on older evolutionary comparative methods because they allow variances and covariances to be partitioned between phylogenetic effects and phylogenetically independent effects. Conifers are an important group of plants that are often considered to be very conservative in their evolution. Our analyses show that both leaf size and climatic traits show very high phylogenetic signals, which supports the idea of such evolutionary conservatism in both leaf traits and niches. More importantly, there are strong correlations between leaf size and climate at the phylogenetic level but not at the residual (phylogenetically independent) level. These relationships are stronger within individual subclades. These results support the importance of occasional, functionally significant innovations followed by long time evolutionary stability. Thus, clades maintain both their leaf traits and their climatic preferences, but survive long term climate change by moving to track their favoured climates. These results also provide an evolutionary basis for using leaf size of fossils as a proxy for past climates. Ben Halliwell and Barbara Holland also made important contributions to this study.

How far into the past can we see? (Student presentation)

Yapeng Lang (Him), Australian national university, research school of biology, evolution and ecology
u7181074@anu.edu.au

Deep Time

The current phylogenetic reconstruction protocol lacks consideration of whether an inference is at its limit. The limit designates the point in evolutionary time beyond which no reliable relationship can be inferred. Existing methods to detect site-saturation indirectly measure the proximity to this limit of inference. However, theories established on the identifiability of Maximum Likelihood estimation in phylogenetics clearly prove the limit is associated with the properties of probability matrices which model the evolutionary process. It indicates that methods for describing an alignment characteristics cannot detect the innate limit of the phylogenetic models. Motivated by this gap, this thesis implements the theories of identifiability and develops a de novo method for evaluating the estimated models in this field. The comparison between the results of the identifiability evaluation and site-saturation test illustrates the discrepancy between the two approaches to find the limit. It reinforces the hypothesis that a failure in phylogenetic reconstruction based on a saturated alignment is not necessarily relevant to the underlying identifiability of the model it used. The method of diagnosing identifiability facilitates further research on conflicting phylogenies, which may detect what is an artefact raised from being beyond the limit of inference.

Bayesian Assessment of Marsupial Phylogenomics and Calibrations Consistency. (Student presentation)

cinthy lorena jimenez silva (she/her), The University of Auckland
cjim882@aucklanduni.ac.nz

Eclectica

Joint work with Jordan Douglas[1,3]*, Alexei J. Drummond*[1,2]*, Remco Bouckaert*[1,4]
[1] *Centre for Computational Evolution,* [2] *School of Biological Sciences,* [3] *Department of Physics,* [4] *School of Computer Science, University of Auckland, Auckland, New Zealand.*

Phylogenetic studies into marsupials have historically encountered challenges related to low resolution and inaccurate dating, primarily due to using distinct genetic markers in separate datasets. Key challenges in enhancing dating accuracy include variations in DNA substitution rates and the reliability of fossil records for calibration. Various analyses have estimated the age of the marsupial clade to be between 78.1 and 83.9 million years ago (Ma), supported by 32 hard constraints from fossil records and prior phylogenetic studies. Previous Bayesian analyses suggest an age range of 59 to 80 Ma. Recent methodological advancements have provided robust approaches to estimating relative divergence times in the presence of nucleotide substitution rate heterogeneity across lineages. However, limited attention has been given to the precision of fossil calibration points for translating relative divergence times into time intervals. Given the disparities among different studies' hypotheses and the potential bias stemming from multiple calibration points, we conducted a Phylogenomics analysis utilizing hundreds of loci under the StarBEAST3 method to enhance our understanding of marsupial phylogeny. We introduce a novel Bayesian cross-validation approach to identify inconsistent fossils when multiple calibration points are available for a clade and apply this method to a Marsupial molecular phylogeny. We identified two misleading calibrations using the foss-validation BEAST-2 package, which contains tools to identify and eliminate inconsistent calibration nodes. Employing consistent calibrations, our findings revealed that despite their ancient lineage, the two most species-rich marsupial clades underwent substantial diversification near the Cretaceous-Paleogene boundary. Today, many of these genuinely ancient marsupial lineages are represented by only a few, often endangered species, underscoring their critical importance for conservation efforts.

Can phylogenetics map traits to genes? Developing PhyloG2P Methods (Student presentation)

Arlie Macdonald (They/Them), Mathematics, University of Tasmania; ARC CoE Plant Success
Arlie.Macdonald@utas.edu.au

Applications

The physical forms of organisms are incredibly complex, and the details of how this complexity arises from underlying genetic information are still poorly understood. Successfully mapping the physical traits of an organism to its genetic information can greatly improve our understanding of the organism and how it interacts with its environment. However, traditional methods of trait-to-gene mapping require detailed genetic information and substantial experimental work. My PhD project will help address these issues by developing a novel trait-to-gene mapping method involving previously un-utilised phylogenetic information. This method takes advantage of the phylogenetic patterns observed in the parallel evolution of a trait (where a trait arises multiple times independently) to identify genes associated with such traits. This removes many of the informational and experimental requirements associated with traditional methods, as well as facilitating the mapping of traits that do not vary within a single species.

I will begin this talk by outlining the biological background and motivation for the problem, including the types of phylogenetic signals we may expect to see in cases of parallel evolution. I will then go over the process I have currently implemented to identify sections of DNA associated with those phylogenetic signals, which includes the use of Hidden Markov Models. I will conclude by providing some of the technical details of my work and discussing some of the additional complexity in the problem that we will work on in future.

Spectral Cluster Supertree: A fast and accurate supertree method for rooted trees (Student presentation)
Robert McArthur (he/him/his), Research School of Biology, Australian National University
robert.mcarthur@anu.edu.au
Models and Algorithms

Joint work with Gavin Huttley (ANU), Ahad N. Zehmakan (ANU), Michael Charleston (UTAS)

Spectral Cluster Supertree is a new, scalable, and parallelisable supertree method for merging rooted phylogenetic trees. Spectral Cluster Supertree is a method derived from Min-Cut Supertree (Semple & Steel, 2000), reliant on spectral clustering as a means of partitioning an internal graph. We also extend Spectral Cluster Supertree to make use of branch lengths within the input trees. The new method is capable of resolving hundreds of source trees over upper thousands of taxa in a magnitude of minutes. We present early comparisons to state-of-the-art methods over new and existing datasets.

Convergence-Divergence Models: Alternatives to Phylogenetic Networks
Jonathan Mitchell (he/him), University of Tasmania
Jonathan.Mitchell@utas.edu.au
Models and Selection

Joint work with Barbara Holland (University of Tasmania), Yongfu Tao (The University of Queensland), David Jordan (The University of Queensland), Emma Mace (The University of Queensland), Jeremy Sumner (University of Tasmania)

Phylogenetic trees are simple models of evolutionary processes. They describe the conditionally independent divergent evolution of species from common ancestors. Trees commonly do not have enough freedom to adequately model all evolutionary processes. For example, introgressive hybridization, where genes are exchanged repeatedly over time from one diverged population to another. Phylogenetic networks model evolution not fully described by a tree. However, phylogenetic networks assume that ancestral populations merge instantaneously to form "hybrid" descendant populations. Our *convergence-divergence models* contrast with phylogenetic networks. They permit continuous gene flow over potentially long time frames. Inference is statistically consistent and we demonstrate good performance of our convergence-divergence models on simulated datasets. We apply our algorithms to *Sorghum*, an important cereal grain consumed worldwide, with many instances of gene flow between populations. From a *Sorghum* gene family presence/absence dataset, we infer a convergence-divergence model with many instances of gene flow concordant with the established literature.

Proximity measures for phylogenetic networks
Yukihiro Murakami, Department of Applied Mathematics; Delft University of Technology
y.murakami@tudelft.nl
Networks

Joint work with Leo van Iersel, Mark Jones, Esther Julien (all from Delft University of Technology)

Phylogenetic networks are used to represent the evolutionary history of species. Recently, the new class of orchard networks was introduced, which were later shown to be interpretable as trees with additional horizontal arcs. This makes the network class ideal for capturing evolutionary histories that involve horizontal gene transfers. Here, we study the minimum number of additional leaves needed to make a network orchard. We demonstrate that computing this proximity measure for a given network is NP-hard and describe a tight upper bound. We also give an equivalent measure based on vertex labellings to construct a mixed integer linear programming formulation. Our experimental results, which include both real-world and synthetic data, illustrate the efficiency of our implementation.

Simulation of conservation management strategies: Trade-offs between genetic diversity and extinction probability (Student presentation)

Habtu Kiros Nigus, Mathematics, University of Tasmania; ARC CoE Plant Success
habtukiros.nigus@utas.edu.au

Eclectica

In efforts to prevent endangered species extinction, genetic rescue has been applied to rescue small, isolated populations facing extinction risk due to genetic drift and inbreeding. It's implemented with the hope of increasing genetic variation and reducing inbreeding depression of the small population. However, the long-term consequences of genetic rescue are not well studied. Nonetheless, it is usually presumed that introduced individuals increased the genetic diversity of the residents and promoted evolutionary potential, thereby enhancing disease resistance, reproductive success, and survival rates. Furthermore, the persistence of species despite their low genetic variation prompted a question on the role of genetic rescue in the survival of species.
In the Phylomania meeting, I will talk about the efficacy of one-off and ongoing supplementation management strategies for genetic rescue. We compared them based on the number of times they experienced extinction, the lower extinction rate, and the better strategy. We performed our simulations in SLiM, the powerful and leading-forward time simulator. We introduced beneficial alleles that we assume arise spontaneously from the wild population that help grow fast (breed early), increase the likelihood of not contracting the disease, and live longer even after contracting the disease. The beneficial mutation could increase in frequency and sweep, which led wild populations to lose their diversity but increase the population size, that is, the benefit of the beneficial mutation outweighs the downside of reduced diversity.

Markovian Binary Trees: Likelihood of Reconstructed Species Trees

Małgorzata O'Reilly, University of Tasmania
malgorzata.oreilly@utas.edu.au

Models and Algorithms

Joint work with Albert C. Soewongsono[1], Barbara R. Holland[2].*
[1] *Washington University;* [2] *University of Tasmania*

We consider a Markovian Binary Tree (MBT) model and describe its application to the evolution of species. We detail how the MuSSE and the BiSSE models are special cases of the MBTs. We derive the likelihood of observing a reconstructed tree under an assumption that some information might be missing due to lack of data. We state the results in a convenient matrix form, provide their physical interpretations, and illustrate the theory through numerical examples.
This work forms a section in the paper:
Matrix-analytic methods for the evolution of species trees, gene trees, and their reconciliation. Albert C. Soewongsono, Jiahao Diao, Tristan Stark, Amanda E. Wilson, David A. Liberles, Barbara R. Holland, Malgorzata M. O'Reilly. Submitted. https://doi.org/10.48550/arXiv.2309.06447

New phylogenetic models of amino acid sequence evolution for probing the relationships between prokaryotes and eukaryotes.

Andrew Roger, Dept. of Biochemistry and Molecular Biology, Dalhousie University
andrew.roger@dal.ca

Deep Time

Joint work with Edward Susko, Hector Baños, Charley McCarthy, Kelsey Williamson, Laura Eme, Alastair Simpson, Matt Brown, Minh Bui

The origin of eukaryotic cells from prokaryotic ancestors remains one of the most enigmatic major transitions in the evolution of life on Earth. It is widely accepted that the nucleocytoplasmic lineage of eukaryotes is related to asgard Archaea and mitochondria evolved from endosymbionts related to Alphaproteobacteria. However, the precise phylogenetic positions of these two ancestral lineages, the nature of additional genetic contributors, the position of the root of the eukaryotic tree and the relative timing of events that occurred in eukaryogenesis remain unclear. This lack of clarity stems, in part, from artefacts induced by the inadequacy of standard phylogenetic models of amino acid sequence evolution (e.g. LG+Gamma) to capture the dynamics of sequence change on the billion-year timescale. Here we introduce several new mixture models for deep-time phylogenetic estimation that allow for: i) heterogeneity in the amino acid substitution process over sites customized to your data, ii) functional shifts in molecules over branches (splits), and iii) simultaneous site- and branch-heterogeneity in amino acid frequencies. We show how these methods can be applied to robustly determine the root of the eukaryote tree and other eukaryogenesis-related phylogenetic problems.

A Diffusion-Based Approach for Simulating Forward-in-Time State-Dependent Speciation and Extinction Dynamics

Albert Christian Soewongsono (he/him/his), Department of Biology, Washington University in St. Louis
soewongsono@wustl.edu

Models and Algorithms

Joint work with Michael J. Landis, Department of Biology, Washington University in St. Louis

We establish a general framework using diffusion processes to simulate forward-in-time state frequencies given state-dependent diversification models. We apply the framework to the geographic-state speciation-extinction (GeoSSE) model. We show that the species range state dynamics simulated under tree-based and diffusion-based processes are comparable. We derive theoretical results for inferring rate parameters that generate observed stationary state frequencies under our framework. We design an iterative method to detect changes in rate parameters leading to shifts in stationary state frequencies across multiple time periods. Finally, we discuss how this framework can be utilized to gain a better understanding between pattern and data generating process from an evolutionary model.

Keywords: evolutionary processes, diffusion processes, state frequencies, GeoSSE.

Rearrangement distances on circular genomes (Student presentation)

Joshua Stevenson (he/him/his), University of Tasmania
joshua.stevenson@utas.edu.au

Algebraic Methods

In addition to point mutations, genomes may also undergo large-scale rearrangement events, in which sections of the genome are broken apart and reattached in a different order or orientation. By considering the possible combinations of rearrangements that could convert one genome into another, we can identify how closely related the organisms are.

I will give an overview of genome rearrangement modelling as an applied combinatorial problem, and then present an algebraic framework for modelling genomes, along with some distance estimates which we can compute under this framework.

Intertwiners for phylogenetic branching operations

Jeremy Sumner (he/they), University of Tasmania
jsumner@utas.edu.au

Algebraic Methods

I will present recent musings on interwiners for phylogenetic branching operations. Being exceptional in the context of algebraic representation theory, we have made recent progress on understanding the mathematical origin of these interwiners in the context of Markov chain model on phylogenetic tree. I will also discuss a conjecture relevant to applications to modelling convergent evolution.

Optimal Phylogenetic Reconstruction of Insertion and Deletion Events (Student presentation)

Sanjana Tule (she/her/hers), School of chemistry and molecular biosciences, University of Queensland
s.tule@uqconnect.edu.au

Algorithms

Motivation: Ancestral sequence reconstruction (ASR) provides insight into molecules' evolutionary history and aids in discovering novel enzymes with properties not found in existing enzymes. While substitutions are routinely inferred using established substitution models, indels are frequently overlooked or eliminated in the ASR process. However, indels embody a crucial evolutionary signal and are integral to the design of enzymes with desired functionalities. In this study, we advance the hypothesis that optimal indel events within a phylogenetic tree can be ascertained by framing evolutionary indel events as a global optimization challenge, while adhering to the structural constraints imposed by the phylogenetic tree. To uncover the most optimal indel events for a given set of extant sequences, we cast the problem of determining likely indel events within a phylogenetic tree as a Mixed-Integer Programming (MIP) optimization problem. Our MIP-based indel solution offers three key benefits: 1) identification of the optimal minimum indel events, 2) preservation of cohesiveness across the phylogenetic tree, and 3) conforming with the indel patterns observed in the extant sequences. Furthermore, our MIP formulation can also find alternate optimal indel histories if available.
Results: To rigorously assess the efficacy of our proposed MIP indel inference model, we conducted extensive evaluations on a dataset comprising 13 phylogenetic trees associated with protein families ranging from 500 to 2000 extant sequences. Additionally, we tested our approach on 60 synthetic trees spanning various levels of complexity. Comparative analyses were performed against existing strategies, including BEP (bi-directional edge maximum parsimony), PIP (position-specific maximum parsimony), and SICP (simple indel-coding maximum parsimony). Remarkably, in almost all cases, our MIP approach proved to be the best one by reporting the fewest indel events, while ensuring phylogenetic coherence, and offering an improved justification for the indel patterns of the extant sequences.This research presents a comprehensive framework for enhancing ASR through optimal indel event inference, improving our understanding of molecular evolution, and offering insights for the design of enzymes with tailored functionalities.

The weighted total cophenetic index: A novel balance index for phylogenetic networks

Kristina Wicke, New Jersey Institute of Technology
kristina.wicke@njit.edu

Networks

Joint work with Linda Knüver (University of Greifswald), Mareike Fischer (University of Greifswald), and Marc Hellmuth (Stockholm University)

Phylogenetic networks play an important role in evolutionary biology as they can be used to depict complex evolutionary scenarios involving reticulate events such as lateral gene transfer and hybridization. Recent research has provided significant progress concerning the reconstruction of such networks from data as well as insights into their graph theoretical properties. However, methods and tools to quantify structural properties of networks or differences between them are still very limited. For example, for phylogenetic trees, it is common to use balance indices to draw conclusions about the underlying evolutionary process, and more than twenty such indices have been proposed in the literature. For networks, on the other hand, balance indices are to-date still scarce.
In this talk, we introduce the weighted total cophenetic index, a generalization of the well-established total cophenetic index for trees, as a balance index for networks and discuss its properties. In particular, we investigate its maxima and minima as well as the structure of networks that achieve these values within the space of so-called level-1 networks.

Phylo-transcriptomics: a new method for analysis of gene-expression data across multiple related species.

Luke Yates (he/him/his), University of Tasmania
luke.yates@utas.edu.au

Applications

Joint work with Barbara Holland, Michael Charleston, & Ben Halliwell, University of Tasmania

The statistical methods of phylogenetic comparative analysis provide a means to compare measurable traits across many different species while taking into account the phylogenetic relatedness of each species pair. Transcriptomics involves the measurement and analysis of gene expression data, typically collected under experimental conditions for the purpose of determining which genes are involved in the response of a given plant or animal. Here we unify these two approaches for the analysis of phylogenetically structured transcriptomics data. This novel methodology allows us to discover the degree to which genetic mechanisms for species traits are driven by phylogenetic relatedness versus adaptation to environmental conditions. This talk discusses the statistical and computational foundations of the method and outlines plans for an exciting new experiment to test for replicated evolution of drought resistance across ten pairs of eucalyptus species.

Tropical Neural Networks for classifying phylogenetic trees

Ruriko Yoshida, Naval Postgraduate School
ryoshida@nps.edu

Algebraic Methods

Deep neural networks show great success when input vectors are in an Euclidean space. However, those classical neural networks show a poor performance when inputs are phylogenetic trees, which can be written as vectors in the tropical projective torus. Here we propose tropical embedding to transform a vector in the tropical projective torus to a vector in the Euclidean space via the tropical metric. We introduce a tropical neural network where the first layer is a tropical embedding layer and the following layers are the same as the classical ones. We prove that this neural network with the tropical metric is a universal approximator and we derive a backpropagation rule for deep neural networks with the tropical metric. Then we provide TensorFlow 2 codes for implementing a tropical neural network in the same fashion as the classical one, where the weights initialization problem is considered according to the extreme value statistics. Finally we apply our method to empirical data including sequences of hemagglutinin for influenza from New York.

List of attendees

Appaw, Raima . . . . . raima.appaw@utas.edu.au
Banos, Hector . . . . . hector.banos@csusb.edu
Blomberg, Simone . . . . . s.blomberg1@uq.edu.au
Butryn, Alexander . . . . . a.butryn2@newcastle.ac.uk
Casanellas, Marta . . . . . marta.casanellas@upc.edu
Chan, Yao-ban . . . . . yaoban@unimelb.edu.au
Charleston, Michael . . . . . michael.charleston@utas.edu.au
Ching Lau, Ching . . . . . ChingChing.Lau@anu.edu.au
Diao, Jiahao . . . . . jiahao.diao@unimelb.edu.au
Douglas, Jordan . . . . . jordan.douglas@auckland.ac.nz
Elgert, Christiane . . . . . christiane.elgert@univie.ac.at
Fountain-Jones, Nick . . . . . nick.fountainjones@utas.edu.au
Fourmant, Mathieu . . . . . mathieu.fourment@uts.edu.au
Francis, Andrew . . . . . a.francis@westernsydney.edu.au
Halliwell, Ben . . . . . benjamin.halliwell@utas.edu.au
He, Wanting . . . . . wahe1@student.unimelb.edu.au
Holland, Barbara . . . . . barbara.holland@utas.edu.au
Homs Pons, Roser . . . . . rhoms@crm.cat
Jarvis, Peter . . . . . peter.jarvis@utas.edu.au
Jordan, Greg . . . . . greg.jordan@utas.edu.au
Lang, Yapeng . . . . . u7181074@anu.edu.au
Lorena Jimenez Silva, Cinthy . . . . . cjim882@aucklanduni.ac.nz
Macdonald, Arlie . . . . . Arlie.Macdonald@utas.edu.au
McArthur, Robert . . . . . robert.mcarthur@anu.edu.au
McComish, Bennet . . . . . bennet.mccomish@utas.edu.au
Mitchell, Jonathan . . . . . Jonathan.Mitchell@utas.edu.au
Murakami, Yukihiro . . . . . y.murakami@tudelft.nl
Navarro, Walter . . . . . walter.navarro@csusb.edu
Nigus, Habtu . . . . . habtukiros.nigus@utas.edu.au
O'Reilly, Małgorzata . . . . . malgorzata.oreilly@utas.edu.au
Roger, Andrew . . . . . andrew.roger@dal.ca
Serra Silva, Ana . . . . . ana.silva.17@ucl.ac.uk
Soewongsono, Albert . . . . . soewongsono@wustl.edu
Stevenson, Joshua . . . . . joshua.stevenson@utas.edu.au
Sumner, Jeremy . . . . . jsumner@utas.edu.au
Swinsburg, Kiah . . . . . kiah.swinsburg@utas.edu.au
Tule, Sanjana . . . . . s.tule@uqconnect.edu.au
Wicke, Kristina . . . . . kristina.wicke@njit.edu
Yates, Luke . . . . . luke.yates@utas.edu.au
Yoshida, Ruriko . . . . . ryoshida@nps.edu

List of talks