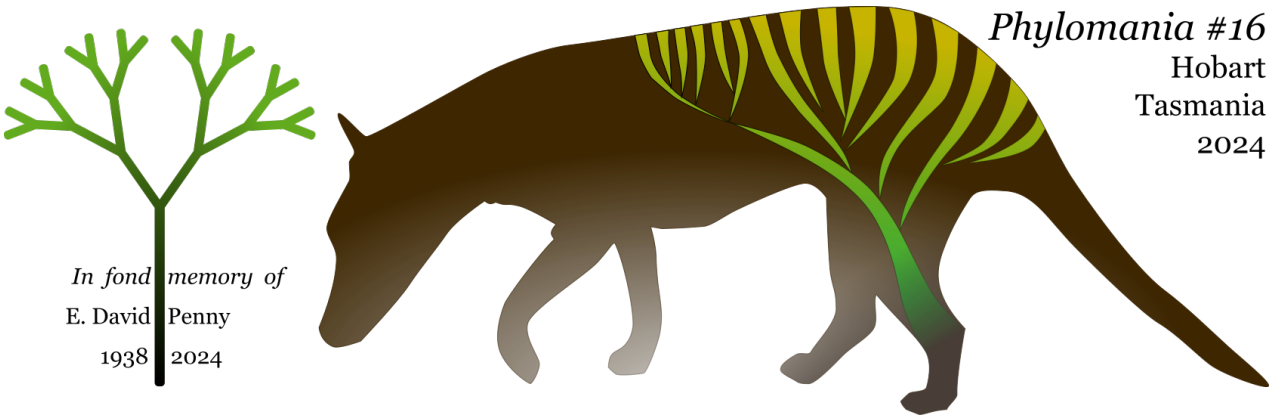
Phylomania 16
Upping the Ante

Hobart
November 27-29
2024



Conference Programme

Session Guide

27 NOV, WED	9:30 – 10:30am	**Registration**
	10:30 – 11am	**Morning Tea**
	11 – 11:20am	**Acknowledgement of Country**
	11:30am – 12pm	**Simone Blomberg** The Wiener sausage as a model for the filling of the phylomorphospace?
	12 – 12:30pm	**Joshua Stevenson** Modelling genome rearrangement events
	12:30 – 2pm	**Lunch**
	2 – 2:30pm	**Nhan Ly-Trong** TreeFormer: A transformer-based tree rearrangement operation for phylogenetic reconstruction
	2:30 – 3pm	**Michael Hendriksen** Counting caterpillar phylogenetic networks
	3 – 3:30pm	**Afternoon Tea**
	3:30 – 4pm	**Jonathan Klawitter** Tractable Tree Distributions
	4 – 4:30pm	**Qin Liu** Robust Phylogenetics
	4:30 – 5pm	**Luke Yates** Prior distributions for correlation matrices in phylogenetic comparative methods
	5 – 6:30pm	**Pub O'clock**
28 NOV, THU	10:30 – 11am	**Morning Tea**
	11 – 11:30am	**Peter Jarvis** Quantum Evolution
	11:30am – 12pm	**Jeremy Sumner** Time-inhomogeneous models and the magic algebraic properties of the equal input model
	12 – 12:30pm	**Kiah Swinsburg** On the Algebra of Species
	12:30 – 2pm	**Lunch**
	2 – 2:30pm	**John Hewson** Two state Markov invariants
	2:30 – 3pm	**Kevin Downard** Structural Phylogenetics with Mass Spectrometry
	3 – 3:30pm	**Afternoon Tea**
	3:30 – 4pm	**Michael Charleston** Sensitivity of phylogenetic estimation to alignment
	4 – 5:30pm	**What Next, Phylogenetics?** Informal Discussion of the next Very Interesting Questions, in memory of David Penny
	5:30 – 6pm	**Prizes and Closing**
	6 – 9:30pm	**Conference Dinner** The Metz Sandy Bay
29 NOV, FRI	10am – 2pm	**Excursion - walk / bus up Mount Nelson** (Packed lunch provided)
	2 – 3:30pm	**Lunch & dispersal**

Abstracts

The Wiener sausage as a model for the filling of the phylomorphospace?

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

Joint work with Cooper Janke

It has long been known that the morphospace is not filled up by species: There are "holes" in the morphospace which are not, and have never been, occupied. The question arises, "Are these regions unoccupied because a) they are open to filling, but evolution hasn't been there (yet) or b) they are "no-go" regions where taxa are absent due to structural or functional constraints. As an attempt to operationalise this problem, we introduce the Wiener sausage: a Brownian motion with a "radius" r. In a phylogenetic stochastic process, we imagine Wiener sausages evolving along a phylogeny, taking up space but also leaving holes. We would like to compare the total amount of the morphospace that has been visited by all taxa in a phylogeny, and compare the size distribution of holes in the Wiener sausage tree to the holes in real phylomorphospaces. We would like any input on the (in)sanity of this project.

Sensitivity of phylogenetic estimation to alignment

Michael Charleston (he/him), University of Tasmania
michael.charleston@utas.edu.au

Molecular phylogenetic estimation relies not only on the right model selection from an alignment, but also on accurate alignment in the first place. When there are insertion/deletion events (indels), then different alignment tools can disagree on what is correct, so there must be downstream effects on model selection and tree estimation. This talk describes a small study in which sequences were simulated under a single standard model (HKY85) using `iqtree2`'s AliSim tool with differing indel rates, aligned using standard tools (`clustal`, `mafft` and `muscle`), and then the phylogeny re-estimated with `iqtree2`. Interestingly, there appears to be a "sweet spot" of positive indel rate at which the branch error between the original and estimated tree is minimised. This suggests that there is definitely something weird going on.

Structural Phylogenetics with Mass Spectrometry

Kevin Downard (he/him/his), Prince of Wales Clinical Research Sciences, UNSW Sydney
kevin.downard@scientia.org.au

For over a decade, we have developed and applied a protein phylogenetic method which utilizes numerical mass datasets acquired in mass spectrometry. This approach avoids the need to generate sequence data or perform a sequence alignment. Sets of peptide masses produced from the digestion of a common protein, expressed across different species, are compared in a pairwise manner and the species are resolved accordingly. Referred to as "phylonumerics'', single point substitutions, which distinguish one protein from the next, are calculated from peptide mass differences during the same tree building step. These are displayed at branch nodes throughout the tree.

Such substitutions or mutations can have profound impacts on a protein's structure at the local and global level. Proline, in particular, is associated with the disruption and promotion of alpha-helices and beta-turns respectively. In this presentation, we demonstrate how the identification of such structural mutations for a diverse set of animal species allow the evolutionary history of such species to be resolved and displayed. A mass tree built for 15 diverse species, easily resolve the birds from mammals, and the ruminant mammals from the remainder of the animals based on structural mutations within the helices of protein myoglobin. A further mass tree study of six more closely related primates, resolves gorilla from the other primates based on a P22S mutation within the B-helix. The remaining five primates are resolved into two groups based on whether they contain a G23S mutation within the same helix.

Overall, mutations that have the greatest impact on a protein's structure, its function, and ultimately the evolution of the species, can be selectively tracked across mass trees providing a new means with which to implement structural phylogenetics.

Counting caterpillar phylogenetic networks

Michael Hendriksen (he/him/his), UNSW, School of Mathematics and Statistics
m.hendriksen91@gmail.com

Joint work with Andrew Francis

Expanding covers provide a way to encode a large class of phylogenetic networks, called labellable networks (Francis & Steel, 2023). This class includes many familiar types of networks, including orchard, normal, tree-child and tree-sibling networks. As expanding covers are a combinatorial structure, it is possible that they can be used as a tool for counting such classes for a fixed number of leaves and reticulations, for which, in many cases, a formula has not yet been found. A more recent paper also introduced a new class of networks, called spinal networks, which are analogous to caterpillar trees for phylogenetic trees and can be fully described using covers (Francis, Marchei, Steel, 2024). In the present talk, we describe a method for counting the intersection of spinal networks with some familiar classes, with the hope that these form a base case from which to attack the more general classes.

Two state Markov invariants

John Hewson, University of Tasmania
john_hewson@icloud.com

Markov invariants associated with tree structures can have utility in terms of measuring the correctness of fit of particular phylogenetic models. The simplest examples are those based on a two state model which give rise to a graded ring structure and associated algebraic invariants. This talk details some aspects of the construction of a generating set for these rings of invariants.

Tractable Tree Distributions

Jonathan Klawitter, University of Auckland
jo.klawitter@gmail.com

The sample from an MCMC algorithm in a Bayesian phylogenetic analysis usually contains at most a few thousand trees and can thus be a poor estimate of the posterior tree distribution for complex instances due to the super-exponential growth of tree space. With the assumption that what happens in one part of the tree is independent of what happens in another part, we can build a better model for a tree distribution. We use a binary graph on clades and clade splits that contains all trees that can build from observed clades or clade splits; each clade has an edge to its clade splits, and each clade split to its child clades. By adding probabilities on choosing a clade split at a clade, we get a Conditional Clade Distribution (CCD), a tractable distribution on phylogenetic trees. For non-trivial instances, CCDs offer better estimates of posterior tree distribution than the sample.
Because of their underlying graph, CCDs are fun to work with and we can develop many interesting algorithms. For example, it is straightforward to sample from a CCD, and with recursive algorithms we can efficiently compute the tree with maximum probability and the phylogenetic entropy of the distribution. In this talk, I will explain CCDs in more detail and showcase a few algorithms and applications we have developed. These include the detection of rogues, unearthing skeletons, estimating ESS, and convergence.

Robust Phylogenetics (Student presentation)
Qin Liu (Qin), School of Natural Sciences, University of Tasmania
qin.liu@utas.edu.au

We present a robust phylogenetic inference method, also called the trimmed loglikelihood method, which effectively identifies fast-evolving, saturated, or erroneous sites in both simulated and empirical multiple sequence alignments. This method avoids circularity by dynamically identifying and removing 'noisy' sites without relying on an initial tree, allowing the specific sites removed to change as tree topology and branch lengths are estimated. Our analyses demonstrated that this method outperforms existing approaches, such as the Slow-Fast method, Tree Independent Generation of Evolutionary Rate approach, and Le Quesne Probability statistics, by removing fewer sites while still inferring optimal phylogenies. Implemented in IQ-TREE 2, the trimmed loglikelihood method is user-friendly with a simple command-line interface. However, challenges remain in addressing heterogeneous evolutionary processes including compositional biases, such as GC bias. Despite these challenges, our approach offers a practical solution for improving phylogenetic inference by effectively identifying down-weighted sites. We recommended that researchers compare trees inferred with varying numbers of down-weighted sites to monitor changes in tree topology to identify a set of candidate tree topologies for further consideration.

TreeFormer: A transformer-based tree rearrangement operation for phylogenetic reconstruction
Nhan Ly-Trong (he/his/him), School of Computing, Australian National University
trong.ly@anu.edu.au

Joint work with Frederick A. Matsen IV - Fred Hutchinson Cancer Research Center; Genome Sciences & Statistics, University of Washington; Howard Hughes Medical Institute, and Bui Quang Minh - School of Computing, Australian National University.

Phylogenomics plays a key role in uncovering the evolutionary histories of species on Earth, ranging from early life forms billions of years ago to, kangaroos, gum trees, and the recent emergence of the SARS-CoV-2 virus causing the COVID-19 pandemic. Popular phylogenetic inference tools, such as IQ-TREE, RAxML, and PHYML, rely on heuristic tree search algorithms to construct phylogenetic trees that maximize the likelihood of observed genetic data. However, tree search is time-consuming and often prone to local optima. To address these issues, we introduce TreeFormer, a new Transformer-based tree rearrangement operation for tree search. Experimental results show that TreeFormer is over 450 times faster than IQ-TREE, while maintaining reasonable accuracy.

Modelling genome rearrangement events (Student presentation)
Joshua Stevenson (he/him/his), University of Tasmania
joshua.stevenson@utas.edu.au

Evolution can occur via a number of mechanisms, and one such mechanism is the occurrence of large-scale genome rearrangement events, in which sections of the genome are broken apart and reattached in a di erent order or orientation. By considering the possible combinations of rearrangements that could convert one genome into another, we can estimate how closely related the organisms are. This problem has often been approached by considering the shortest possible path between genomes under a given set of rearrangements, but modelling genome rearrangement as a Markov process (where each rearrangement is assigned a probability) opens up opportunities for additional distances to be considered. I will give an overview of genome rearrangement modelling, and then present our algebraic framework for modelling genomes, along with some distance estimates that we can compute under this framework.

Time-inhomogeneous models and the magic algebraic properties of the equal input model

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

I will review some work on time-inhomogeneous Markov chains undertaken early 2024 with my collaborator Michael Baake (U Bielefeld). Under relatively mild restrictions on time-dependent Markov generators (rate matrices), we provide a method for finding explicit (integral) solutions for the corresponding probability transition matrices. I will make explicit connections to phylogenetics by discussing our technique in the context of well-known DNA substitution models, in particular the Kimura 3P and Felsenstein '81 models. The talk will also provide a link to our prior work on 'Lie-Markov' models, which occurs naturally since the current setting relies on the theory of Lie groups and algebras via the 'Magnus expansion" for the solution of linear systems of ODEs with time varying coefficients

On the algebra of species: an algebraic approach to modeling speciation (Student presentation)

Kiah Swinsburg (he/him), University of Tasmania
kiah.swinsburg@utas.edu.au

Phylogenetic inference aims to determine evolutionary relationships between species, from the strings of nucleotides that make up their DNA. In this context, nucleotide substitution – which is the driver of many evolutionary adaptations - is modelled by a Markov process. Markov Processes are a class of stochastic process that has a memoryless property, such that the probability of future substitutions depends only on the present state. This memoryless property, remarkably, sets the stage for matrix algebra to play a major role in the analysis of molecular evolution.

Prior work by Sumner *et. al.* has noted the importance of multiplicative closure for phylogenetic Markov models. This motivates the study of so-called Lie-Markov models, in which nucleotide substitution rates vary according to the action of a Lie algebra on a probability vector space.

This talk will provide an overview of how to extend the Lie-Markov approach to phylogenetics, by providing a mathematical framework , in the language of representation theory, for modelling speciation events.

Prior distributions for correlation matrices in phylogenetic comparative methods

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

Joint work with Barbara Holland and Ben Halliwell, University of Tasmania

Estimating correlation matrices for multi-trait phylogenetic comparative analyses is both statistically and computationally challenging. In this talk, we review several parametrisation approaches for covariance matrices, including separation strategies, precision matrices, partial correlations, and Cholesky factorisation. Using these parametrisations, we compare how traditional inverse Wishart and Lewandowski-Kurowicka-Joe (LKJ) priors perform against the sparsity-inducing horseshoe prior for correlation matrix estimation in the challenging but common scenario of many traits observed across relatively few species.

List of attendees

Blomberg, Simone . . . . . s.blomberg1@uq.edu.au
Casanellas, Marta . . . . . marta.casanellas@upc.edu
Charleston, Michael . . . . . michael.charleston@utas.edu.au
Downard, Kevin . . . . . kevin.downard@scientia.org.au
Edwards, Claire . . . . . claire.edwards@utas.edu.au
Grant, Gus . . . . . augustus.grant@utas.edu.au
Hendriksen, Michael . . . . . m.hendriksen91@gmail.com
Hewson, John . . . . . john_hewson@icloud.com
Holland, Barbara . . . . . Barbara.Holland@utas.edu.au
Jaiswara, Ranjana . . . . . ranjana.jaiswara@gmail.com
Jarvis, Peter . . . . . Peter.Jarvis@utas.edu.au
Klawitter, Jonathan . . . . . jo.klawitter@gmail.com
Liu, Qin . . . . . qin.liu@utas.edu.au
Ly-Trong, Nhan . . . . . trong.ly@anu.edu.au
McComish, Bennet . . . . . bennet.mccomish@utas.edu.au
Mitchell, Jonathan . . . . . Jonathan.Mitchell@utas.edu.au
Silva, Anna Serra . . . . . ana.silva.17@ucl.ac.uk
Stevenson, Joshua . . . . . joshua.stevenson@utas.edu.au
Sumner, Jeremy . . . . . jsumner@utas.edu.au
Swinsburg, Kiah . . . . . kiah.swinsburg@utas.edu.au
Tiangco, Narra Maria . . . . . narramaria.tiangco@utas.edu.au
Wilson, Amanda Erin . . . . . aw3746@cumc.columbia.edu
Yates, Luke . . . . . luke.yates@utas.edu.au

List of talks