Phylomania 2015

University of Tasmania, Hobart
November 11-13

Welcome

Welcome to Phylomania 2015 – hosted by the University of Tasmania's Theoretical Phylogenetics Group

Phylogenetics is concerned with what is probably the most important problem in biology: reconstructing the evolutionary history of present-day organisms from molecular data, such as DNA, or morphological characters. Hidden from view, in phylogenetics software packages used by biologists, are algorithms implementing stochastic and combinatorial methods on binary trees, as well as more general network structures. The mathematics involved represent a unique confluence of probability theory, discrete mathematics, stochastic methods, and statistical inference, as well as algebraic methods such as group theory. There are many important theoretical and practical problems that arise, such as statistical identifiability of models, consistency and convergence of methods. These problems can only be solved using a multi-disciplinary approach. *Phylomania* brings together internationally recognised experts, with the aim of discussing the pressing research problems in phylogenetics.

Phylomania is held within the School of Physical Sciences at the University of Tasmania, Hobart. The meeting brings together mathematicians, statisticians, biologists and computer scientists to discuss current research in phylogenetics. The conference is both informal and leading-edge: some of the very latest ideas, applications, and unsolved challenges, to do with phylogenetic inference and evolutionary biology are voiced here.

Important Information

Location: If you've attended before, you'll already know where the conference is held: Maths and Physics Building (on Google Maps), Clark Road, Sandy Bay Campus.

The lectures will all be held in Physics Lecture Theatre 2, also known as Room 313. Morning and afternoon teas and lunches will all be in Room 333 in the same building, close by.

If you have a poster to show, you may mount it in Room 333.

Wireless Internet Access: If you have `eduroam` set up on your laptop or other mobile computer you can log in to UTAS wireless network with that.

We also have a wireless account set up for the conference. The network name is `UConference`, and the password to access it is `Nov-Conf-2015`.

There are desktop computers available if you need one, with the *username* of `utas008` and the *password* of `Phylomania_2015`.

If you have any problems with IT while you're here, you can call the UTAS Service Desk on 03 6226 1818.

Excursion: We usually go on an excursion on the last afternoon of the conference. There is lots of great walking around Hobart, and we usually take a moderate, low-intensity walk around, before adjourning to a local hostelry for a reward. Make sure you have sensible shoes for this (sneakers are ok).

If you're planning to come along, bring a sun hat and sun screen!

The UV in Tasmania is very strong and you can easily get bad sunburn if you're not well prepared.

Discussions: At the end of each day we often go to a local pub and this year (again) we have chosen Preachers Bar, which is about a 30 minute walk from campus (see map on the back page).

Dinner: We have booked tables for 7pm on Thursday the 12th of November, at Tavern 42 Degrees South T42 is on Elizabeth Street Pier, a few minutes walk from Preachers Bar (see map on back page). The cost of the meal is *not* included in your conference registration, but prices are reasonable, and it's a great spot.

Programme

Day One: Wednesday, November 11th

Session 1: Genomes and Hybridization

08:30 – 09:00 Registration, Tea and Coffee in Room 333

09:00 – 09:10 **John Dickey**, Head of School of Physical Sciences, University of Tasmania

Introduction and Welcome

09:10 – 09:50 **Andrew Francis**, University of Western Sydney

Bacterial phylogeny in the Cayley graph

09:50 – 10:10 **Chris Burridge**, University of Tasmania

Hybridisation in Nothofagus?

10:10 – 10:30 **Barbara Holland**, University of Tasmania

Making inferences about hybridization

Break: coffee, tea and snacks provided in Room 333

Session 2: Likelihood

11:00 – 11:40 **Arndt von Haeseler**, University of Vienna

Terraces, Partial Terraces and Phylogenetic Inference

11:40 – 12:00 **Michael Woodhams**, University of Tasmania

Lie-Markov Models

Lunch, in Room 333

Session 3: The Right Model

13:30 – 13:50 **Matt Phillips**, Queensland University of Technology

Ancient DNA, total evidence and the evolution of giant extinct kangaroos from Tasmania

13:50 – 14:10 **Daisy Shepherd**, The University of Auckland

Detecting Heterogeneity in Phylogenetic Inference: Exploring the Sliding Window Approach

14:10 – 14:30 **Andrew Ritchie**, University of Sydney

Evaluating the impact of the tree prior on molecular dating

Break: coffee, tea and snacks provided in Room 333

Session 4: Genomic Evolution

15:30 – 16:10 **Tristan Stark**, University of Tasmania

Preservation of Gene Duplicates: Analysis of the Subfunctionalization Model

16:10 – 16:30 **Michael Charleston**, University of Tasmania

Aligning Biological Networks

Session 5: Day One Discussion

16:45 — late General Discussion at Preachers Bar, 5 Knopwood St, Hobart TAS 7000

Day Two: Thursday, November 12th

Session 6: Invariants and Other Creatures

08:30 – 09:00 Tea and Coffee in Room 333

09:00 – 09:10 Housekeeping

09:10 – 09:50 **Marta Casanellas**, Universitat Politecnica de Catalunya

New approaches to phylogenetic invariants

09:50 – 10:10 **Bennet McComish**, University of Tasmania

Microsatellite evolution at different timescales

10:10 – 10:30 **Barbara Schoenfeld**, University of Tasmania

A pedigree for wild devils

Break: coffee, tea and snacks provided in Room 333

Session 7: Next Gen Surprise

11:00 – 11:40 **Allen Rodrigo**, Australian National University

Evolutionary analysis of short-read sequences from mixed samples of unlabeled individuals.

11:40 – 12:00 **William Dodt**, Queensland University of Technology

Talk Title To Be Confirmed

Lunch, in Room 333

Session 8: Populations

13:30 – 14:10 **Conrad Burden**, Mathematical Sciences Institute, Australian National University

Genetic drift in growing populations with applications to dating mitochondrial Eve

14:10 – 14:30 **Scott Carver**, University of Tasmania

Relatedness and urban development shape viral transmission in bobcats

14:30 – 14:50 **Yuantong Ding**, Duke University/ANU

Modeling single cell phylogenies in cancer under different evolutionary model

Break: coffee, tea and snacks provided in Room 333

Session 9: Phylogenetic Inference

15:30 – 15:50 **David Duchene**, Australian National University

Assessing absolute model performance in phylogenomics

15:50 – 16:10 **Lars Jermiin**, CSIRO

ModelFinder: A new model-selection method improves phylogenetic estimates

16:10 – 16:30 **Ben Rohrlach**, University of Adelaide

Informative Measures of Goodness of Fit for Models of DNA Substitution

Session 10: Day Two Discussion

17:00 General Discussion at Preachers Bar, 5 Knopwood St (see map on back page)

19:30 Dinner at T42, Elizabeth Street Pier (see map on back page)

Day Three: Friday, November 13th

Session 11: Making Things Complicated

08:30 – 09:00 Tea and Coffee in Room 333

09:00 – 09:10 Housekeeping

09:10 – 09:50 **Mike Steel**, Canterbury University

Two results on trees and networks

09:50 – 10:10 **Patricio Russel**, University of Auckland

Success of Bayesian Inference in the Four-Taxon Case.

10:10 – 10:30 **Stephen Crotty**, University of Adelaide

Modelling heterotachous evolution

Break: coffee, tea and snacks provided in Room 333

Session 12: Wrapping Up

11:00 – 11:20 **Rob Lanfear**, Macquarie University

Automated selection of partitioning schemes for phylogenetics: how to cluster sites

11:20 – 12:00 **Jeremy Sumner**, University of Tasmania

A representation-theoretic approach to circular genome rearrangements

Lunch, in Room 333

Friday Afternoon Excursion

A walk and talk; details to be determined closer to the event.

Abstracts

Genetic drift in growing populations with applications to dating mitochondrial Eve

Conrad Burden, Mathematical Sciences Institute, Australian National University

Most population genetics studies have their origins in the Wright-Fisher model or some closely related model. In such models the total population size is specified externally rather than determined dynamically, and each individual essentially chooses its ancestor at random from the previous generation. An alternate approach which has received little attention is to model the population as the result of a Galton-Watson branching process. In this approach, each individual produces a random number of children, and hence the total population size is generated stochastically.

The most recent common ancestor of the human population following only maternal lines of decent is known as the mitochondrial Eve (mtE). O'Connell [1] has used the Galton-Watson model to find an analytic estimate the time since mtE. O'Connell's analysis assumes certain asymptotic approximations which we believe may not necessarily be appropriate in general. Cyran and Kimmel [2] have carried out extensive numerical simulations of the same model, though only over small numbers of generations, and for these cases claim consistency O'Connell's results.

In this talk I present an analysis of the Galton-Watson model of genetic drift which avoids the limitations of the O'Connell approximation through a combination of analytic analysis of the forward Kolmogorov equation and numerical integration. The approach enables an estimate of the time since mtE and an estimate of the population size during the mtE's lifetime.

[1] N. O'Connell, "The genealogy of branching processes and the age of our most recent common ancestor", Adv. Appl. Prob. 27, 418-442 (1995).

[2] K.A. Cyran and M. Kimmel, "Alternatives to the Wright-Fisher model: The robustness of mitochondrial Eve dating", Theor. Pop. Biol. 78, 165-172 (2010).

Hybridisation in Nothofagus?

Chris Burridge, University of Tasmania
(Joint work with James Worth, Greg Jordan, Anna Brüniche-Olsen, Peta Hill, René Vaillancourt. Nothofagus Hybridisation Consortium, University of Tasmania)

Hybridisation is widely implicated when discordant gene trees are observed during phylogenetic and phylogeographic studies of closely related plant species. However, alternate explanations involving lineage sorting or erroneous reconstruction of gene trees are less frequently investigated. Here we document discordant chloroplast and nuclear gene trees for two closely related plant species of the southern temperate rainforest tree genus Nothofagus, where a nuclear gene tree appears 'sorted', while this process appears incomplete for the chloroplast lineage. This discrepancy is difficult to explain as the chloroplast lineage should sort faster than nuclear lineages, but cannot be readily explained by hybridisation either, given the allopatric nature of the present species distribution. Hypotheses involving 'ghost hybrids' have been entertained, but limitations in the coalescent simulation process will also be discussed.

Relatedness and urban development shape viral transmission in bobcats

Scott Carver, University of Tasmania

Phylodynamic approaches, using viral and host phylogenies, can provide important insights into how pathogens spread. By extension landscape factors (such as urbanization and other ecological variables) can strongly shape phylodynamic relationships in wildlife populations. We encompass landscape factors into phylodynamics (termed 'ecophylodynamics'), using Feline immunodeficiency virus (FIV) and bobcats (*Lynx rufus*) in a heavily fragmented landscape in southern California. This bobcat population provides a model system to apply an ecophylodynamic approach, as FIV is fast evolving and doesn't directly lead to host mortality. Using the combination of landscape, host demographic and host genetic information we show landscape variables coupled with bobcat relatedness were important for FIV transmission. Bobcat relatedness in particular is positively correlated with increased human development (i.e., bobcats are more likely to be related the closer they are to anthropogenic disturbance and this is

important for FIV transmission). An ecophylodynamic approach provides a novel way untangle the interplay between host demographic, host genetic, and landscape factors guiding viral transmission.

New approaches to phylogenetic invariants

Marta Casanellas, Universitat Politecnica de Catalunya
(Joint work with Jesús Fernandez-Sánchez, Universitat Politecnica de Catalunya)

We shall discuss the advantages and drawbacks of phylogenetic invariants and their use in phylogenetic reconstruction, model selection, and identifiability issues. We shall present results on real and simulated data of new methods of phylogenetic reconstruction based on geometric tools related to phylogenetic invariants.

Aligning Biological Networks

Michael Charleston, University of Tasmania
(Joint work with Alexandru Radu and Martin McGrane, University of Sydney)

Biological networks are a rich source of information, including their potential to understand evolutionary relationships. However in order to compare networks we need to find their similarities and differences at a structural level, which is a computationally hard problem.

This talk discusses the underlying theoretical problem of aligning networks, describes recent developments in heuristic methods for pairwise and multiple alignment of networks, and shows beginnings of methods with which we can estimate ancestral networks from those at the leaves of a (very simple) tree.

Modelling heterotachous evolution

Stephen Crotty, University of Adelaide

In this talk I will report on progress in

1. defining heterotachous evolutionary models
2. working out when (and when not) we can recover their parameters
3. making it all work for more than toy examples (not that we don't love our 4-taxon simulations as much as the next theoretical phylogeneticist)

Modeling single cell phylogenies in cancer under different evolutionary models

Yuantong Ding, Duke University/ANU

Cancer as an evolutionary process was first described by Nowell in 1976, and since then researchers have identified clonal expansions and genetic heterogeneity within many different types of neoplasms, which have profound clinical implications for neoplastic progression, cancer prevention and cancer therapy. With the technical development of single cell sequencing, it is now feasible to gain more precision and greater insight into the evolutionary process within tumors on the single-cell level. We collected data from the currently available single-cell tumor sequencing studies and found their phylogenies all show an interesting ladder-like pattern. In this study, we used simulations to explore the possible underlying evolutionary models. We used both a homogeneous Wright-Fisher and a spatial Moran model. Our results shows that one possible reason for these latter-like tree may lie in the sampling method used to generate. Besides, an evolutionary model with mildly deleterious passenger mutations can also predict these ladder-like phylogenies in a well-mixed population under a certain parameter regime.

Resolving kangaroo phylogeny with transposable elements

William Dodt, Queensland University of Technology
(Joint work with Matthew Phillips[1], Susanne Gallus[2], Maria Nilsson-Janke[2]; [1]Queensland University of Technology, [2]Senckenberg Biodiversity and Climate Research Centre)

The genus *Macropus* contains many of the most recognizable kangaroo species, which include the largest living marsupials. Despite being a well-studied group, the phylogenetic relationships within this genus remain poorly resolved. With the development of next generation sequencing, it has become possible to investigate phylogenetic relationships using genome level characters. I will discuss the use of retrotransposons as phylogenetic markers, with a focus on kangaroo evolution. A particular class of retrotransposon – an endogenous retrovirus – has been prolific during the evolution of kangaroos. We have utilized presence/absence information of retrotransposons to shed light on the phylogenetic relationships among members of the genus *Macropus*, and close relatives, and propose that incomplete lineage sorting has been prevalent during kangaroo evolution.

Assessing absolute model performance in phylogenomics

David Duchene, Australian National University
(Joint work with Sebastian Duchene, Simon Y.W. Ho)

Statistical phylogenetic methods employ models of the evolutionary process. Approaches for assessing model adequacy are able to falsify evolutionary models and can even provide an indication of which model assumptions are being violated by the data. If the model assumptions are violated by the data, or are otherwise inadequate, they are likely to be poor descriptions of biological reality and lead to misleading phylogenetic estimates. Using simulation and empirical data, I explore whether testing substitution model adequacy in genome-scale data opens an opportunity to select genes and regions that are more accurately described by available models, thereby improving our confidence in the resulting estimates.

Bacterial phylogeny in the Cayley graph

Andrew Francis, University of Western Sydney

Modelling bacterial genome rearrangement operations as group actions on the space of all possible genomes provides a one-to-one correspondence between genome space and the group that acts. This means that a subset of genomes defines a set of points on the Cayley graph of the group, and a phylogeny on those genomes is represented by a Steiner tree on those points. In this talk I will describe this viewpoint and several related results. First, I will discuss a more nuanced view of the "minimal distance" between genomes, and second, I will describe some algorithmic results relating to the median problem for three genomes on the Cayley graph.

Making inferences about hybridization

Barbara Holland, University of Tasmania
(Joint work with Jonathan Mitchell, Jeremy Sumner, and Michael Woodhams)

> "*Mathematically, we can view stochastic uncertainty as being conditional on an assumed model. Mathematics within the model can be precise and potentially within the control of the statistician. However, the choice of model itself carries an inductive uncertainty, which may be less precise and potentially beyond the control of the statistician.*"
>
> - Pawitan (2001)

Statistical models have two main purposes, prediction and explanation. In phylogenetics we are very much focused on the latter purpose - we want to know what the tree was. Indeed, most models we use in phylogenetics implicitly assume that evolution has occurred along a tree. This begs the question, what should we do if what we are interested in is hybridization?

Various authors have proposed likelihood-based methods for phylogenetic networks, for instance calculating the maximum likelihood score by taking a mixture model over the set of trees that are implied by a directed acyclic

graph (DAG). Such DAGs can be constructed by beginning with a good tree and then progressively adding edges until there is no sufficient improvement in fit by continuing to add edges. The trouble with using DAGs to define mixture models is that this approach doesn't actually capture the biological processes of interest within the model. The sort of things we'd like the data to tell us are what is the relative rate of hybridization to mutation or speciation?

To address the Pawitan quote from above, in the context of network evolution we need to make careful decisions about what biological processes should be included within the model such that inferences about reticulate (non-treelike) processes of evolution can be brought within the realm of stochastic uncertainty rather than being left as a source of inductive uncertainty.

This talk will be a hybrid — or perhaps *cocktail* — of the following topics:

1 shot philosophical meanderings along the lines above

1 shot recent results on convergence networks (J. Mitchell, J. Sumner)

1 shot recent results on inferring hybridization on the basis of summary statistics (M. Woodhams)

ModelFinder: A new model-selection method improves phylogenetic estimates

Lars Jermiin, CSIRO
(Joint work with Subha Kalyaansmoorthy[1], Bui Q Minh[2], Thomas KF Wong[1], Vivek Jayaswal[3], Arndt von Haeseler[2,4]; [1]CSIRO Land & Water, Canberra, ACT 2601, Australia. [2]Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria. [3]School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia, [4]Bioinformatics and Computational Biology, Faculty of Computational Science, University of Vienna, Vienna, Austria)

Model-based molecular phylogenetic analysis plays an important role in the comparison of genomic data, allowing us to annotate genomes more accurately, to develop a deeper understanding of the evolution of species and their genes, and to identify the origin and spread of pathogens and agricultural pests. Model selection is a critical step in all such analyses. We present a new model-selection method that incorporates a model of rate-heterogeneity across sites (RHAS) not previously considered. The method also allows the tree topology to be a variable during the model search, and show that it can yield accurate estimates of RHAS, where other model-selection methods fail (e.g., when the RHAS involves bimodal distribution of rates).

Automated selection of partitioning schemes for phylogenetics: how to cluster sites

Rob Lanfear, Macquarie University
(Joint work with Paul Frandsen, Smithsonian Institution)

In this talk, I will present our most recent efforts to automatically select partitioning schemes for the phylogenetic analysis of genome scale datasets. This involves new work on methods to rapidly determine groups of sites that have evolved in similar ways, the discovery of some troubling biases in model selection, methods to overcome these biases, and some unanswered questions that we're still working on.

Microsatellite evolution at different timescales

Bennet McComish, University of Tasmania
(Joint work with Michael A. Charleston, Matthew Parks, Barbara R. Holland, David M. Lambert)

We have analysed 63 published genomes (including 48 birds), and 26 modern and 22 ancient Adélie penguin read sets that have previously not been published. We identified a total of approximately 30 million microsatellites, including 177,974 in Adélie.

Using the ancient samples we tracked homologous microsatellite loci through evolutionary time and determined that there was no discernible drift in their length. Slippage is apparently at a dynamic equilibrium; the length is not changing over time on average, although there is length polymorphism among the individual samples. Consistent with results in human genomics, we found a strong over-representation of trimer repeats in coding regions across 48 avian species. We also discovered that a surprising number of microsatellites persist over very long time scales.

By being able to map so many microsatellites to the phylogeny we could show that more ancient microsatellites had quite different characteristics, both in terms of their motif length and where they are found. Finally we have also shown that individual loci appear to have lost and regained the ability to behave as microsatellites over time.

Ancient DNA, total evidence and the evolution of giant extinct kangaroos from Tasmania

Matt Phillips, Queensland University of Technology
(Joint work with Manuela Cascini, Queensland University of Technology)

Short-faced kangaroos (Sthenurinae) and giant wallabies (Protemnodon) were major components of Australian and New Guinean mammal faunas until the Late Pleistocene megafaunal extinction. Tracing their evolution from fossil records has been controversial, and until recently, DNA from their >40,000 year old remains has been inaccessible. Sthenurine affinities among modern taxa fall in a long debated trichotomy with the dominant living macropod clade, Macropodinae and the monotypic banded hare wallaby, *Lagostrophus fasciatus*. In turn, Protemnodon affinities have been widely distributed within Macropodinae. Llamas *et al*. (2015) recovered the first DNA fragments from these extinct macropodids, but gave precedence to "total evidence" analyses that were strongly informed by morphological characters. We show more generally that the morphological data mislead rather than enhance molecular phylogenetic signal among macropods. We focus instead on reducing base composition bias and phylogenetic signal erosion among the DNA data. Our results confirm a close relationship between Protemnodon and the iconic kangaroo genus, *Macropus*, and favour Lagostrophus being the closest living link to the once diverse sthenurine radiation.

Evaluating the impact of the tree prior on molecular dating

Andrew Ritchie, University of Sydney
(Joint work with Nathan Lo, Simon Y.W. Ho)

Sampling schemes incorporating multiple individuals per species are a common feature of phylogenetic studies. In particular, such mixed datasets are essential for ensuring sufficient coverage of species in groups with unknown or disputed species boundaries. However, these datasets can cause difficulties for molecular dating. Bayesian phylogenetic methods for estimating evolutionary timescales require the specification of a prior probability on the tree and divergence times. Current software packages typically offer time-tree priors that are based on either a speciation or a coalescent process. In the absence of tree priors designed specifically for mixed datasets, researchers must generally choose from among the aforementioned prior classes. It is at present unclear what impact this choice might have on dating results.

Here, we compare the results of Bayesian divergence time estimation on three empirical mixed datasets assuming different time-tree priors. We assess the fit of the most commonly-used tree models to real data and conduct simulations to investigate the effects of different sampling strategies. We make recommendations for choosing appropriate tree priors in molecular clock analyses and discuss implications for the sensitivity of molecular dating methods to the choice of tree prior.

Evolutionary analysis of short-read sequences from mixed samples of unlabeled individuals.

Allen Rodrigo, Australian National University

Next generation sequencing (NGS) provides researchers with several thousand short-read sequences, at a cost that is orders of magnitude cheaper than older Sanger sequencing technologies. When these technologies are applied to multi-individual samples of longer fragments of amplified DNA or RNA from genetically variable populations, it is challenging to estimate the original haplotypes in the samples. This has a flow on effect if the aim of such studies is the estimation of evolutionary parameters. In this talk, I describe some initial steps that we have taken to develop a new Bayesian Markov chain Monte Carlo procedure that allows the use of single-sample multi-individual NGS sequence data in tree-based evolutionary analyses. If this method works, researchers will benefit from the flexibility of tree-based inference, while correctly accounting for the uncertainty associated with the absence of full-length sequences.

Informative Measures of Goodness of Fit for Models of DNA Substitution

Ben Rohrlach, University of Adelaide

Models of DNA substitution describe they way in which DNA evolves over time, and can contain many parameters. Given data (an alignment), and a model of DNA substitution, a best fitting tree can be found. Further, methods exist for finding which DNA substitution models fits the data "best", however, researchers have no simple tool for finding if the "best" substitution model fits the data adequately.

Site patterns form the basis for likelihood calculations for phylogenetic analyses, and hence we can calculate the number of times we expect to observe any site pattern in our sample. In the MISFITS method, they compare the number of times a specific site pattern is observed with the expected number of observations. With these comparisons we begin to form a concept of residuals for phylogenetic models.

However, these methods are not without their limitations. The number of unique site patterns becomes very large for even very small data sets, and so finding the site patterns that we did not observe, even though we expected to, becomes difficult. Similarly, the majority of the possible site patterns are extremely unlikely, and due to the finite length of DNA sequences, comparisons are dominated by the vast number of zero counts for site patterns. We aim to end with a discussion about possible solutions for overcoming these limitations.

Success of Bayesian Inference in the Four-Taxon Case.

Patricio Russel, University of Auckland

Methods of model selection remains an active field in phylogenetic inference. Several simulation-based studies have assessed their performance, primarily using Maximum Likelihood or Maximum Parsimony methods. Here, we investigated the properties of Bayesian phylogenetic tree estimation for the case of four taxa. Simulated data were generated for a wide variety of branch lengths combinations relating to the Felsenstein zone. The different models were compared using Bayes factors. Further, the expected proportion of successes was estimated for each branch length combination. Of particular interest was the impact of rate categories applied to a continuous gamma rates across sites model.

The Devils' pedigree

Barbara Schönfeld, University of Tasmania

The Tasmanian Devil (*Sarcophilus harrisii*) is the largest surviving carnivorous marsupial and is endemic to Tasmania. The species has recently seen a dramatic decline due to Devil Facial Tumour Disease (DFTD). DFTD was first described in 1996 and is one of only two known transmissible cancers. The disease has a mortality rate of 100% within less than a year from infection and extinction of the species in the wild long seemed a distinct possibility. However, populations have been closely monitored and sampled over the last 16 years and adaptive responses to the extreme selective pressures posed by DFTD were observed.

In this talk I present my efforts to construct the most extensive pedigree of a wild population to date. Using Rapture SNP data from more than 3500 Tasmanian Devils, representing 50-100% of their respective populations, this pedigree will allow us to investigate reproductive success correlated to specific phenotypic traits and rapid adaptive changes in a population under extreme selection pressure. Moreover, this dataset offers a unique opportunity to address a range of other evolutionary and genomic questions.

Detecting Heterogeneity in Phylogenetic Inference: Exploring the Sliding Window Approach

Daisy Shepherd, The University of Auckland

Heterogeneity presents many challenges to the modelling process within phylogenetic analyses. Currently there are a number of effective techniques aimed at improving the detection and modelling abilities of heterogeneity. One relatively unused method concerns the sliding window (SW) approach.

We implemented the sliding window approach as a means to assessing conflicting hypotheses and compare competing models. The method performed phylogenetic analysis on a small window of the sequence, before iteratively sliding along and repeating the analysis within each individual new window of sites. The process repeated until all sites had been involved within at least one inference.

Our exploratory analysis demonstrated a number of strengths and advantages associated with this approach, some specific to the method's ability in detecting interesting patterns within the data. Likelihood ratio tests indicated the SW approach outperformed the typical complete alignment inference in pinpointing the presence of rate heterogeneity. A further application to real alignment data highlighted the versatility of the SW approach, when assessing specific problems of interest.

Whilst the sliding window approach was more computationally intensive, the ability to profile and apply multiple inferences across an alignment allowed more insightful and detailed heterogeneity detection within the analysis. Results indicated the approach is undoubtedly a useful tool in detecting general model violations, with the potential to be developed into a powerful inference technique.

Preservation of Gene Duplicates: Analysis of the Subfunctionalization Model

Tristan Stark, University of Tasmania
(Joint work with David Liberles, University of Wyoming, Barbara Holland, Małgorzata O'Reilly)

After duplication, the fate of a duplicate pair of genes is of great interest in genomics. Gene duplication and subsequent loss is believed to contribute to genome diversification. There has been an increasing focus on developing models to describe these processes. Under the subfunctionalization model functions performed by the original gene can eventually be distributed between the two copies, preserving both copies by selective pressure. Alternatively one copy can be lost entirely with the other performing the full set of functions of the original gene.

We construct a mathematical model for the subfunctionalization process, and analyse its performance. This includes the derivation of exact rates of gene loss, and comparison of these rates to existing phenomenological approximations. Our results indicate that the importance of subfunctionalization may have been underestimated by previous analysis, with the qualitative features of the model in close agreement to empirical data.

Two results on trees and networks

Mike Steel, Canterbury University

A recent paper (Semple and Bordewich, 2015) established an elegant result for the smallest number of r-state characters required to 'capture' a phylogenetic tree. In the first part of my talk, I derive a related result by applying the probabilistic method. In the second part of the talk, I describe some recent work with Andrew Francis, motivated by a basic question: When is a phylogenetic network merely a tree with arcs between its branches?

A representation-theoretic approach to circular genome rearrangements

Jeremy Sumner, University of Tasmania
(Joint work with Andrew Francis, University of Western Sydney, and Peter Jarvis, University of Tasmania)

A simple model of circular genome rearrangements is obtained by taking a restricted set of gene permutations S and considering the distance between two genomes to be related to the number of permutations required to convert one genome into the other. Stochastically, we consider genome evolution as a Poisson process where each permutation in the set S is applied with uniform rate, with the distance between two genomes defined as the maximum likelihood time passed. Algebraically, the set of permutations S generates a group G and we need to compute the number of ways of expressing a given element of G as a word of length k in "letters" S. Unfortunately, this algebraic task is combinatorially intensive (factorial in the number of genes/regions) but must be solved in order for the stochastic analysis to proceed. I will present some recent ponderings taking a representation theoretic approach which converts the combinatorial problem into linear algebra and the computation of certain eigenvalues. This improves the complexity of the problem a little (by a square root), but is by no means a silver bullet (square root of factorial complexity is still terrible, really terrible!)

Terraces, Partial Terraces and Phylogenetic Inference

Arndt von Haeseler, University of Vienna
(Joint work with Olga Chernomor and Bui Quang Minh)

We discuss the concept of phylogenetic terraces to improve the computational efficiency of phylogenetic inference programs. To this end, we provide the rules to detect terraces during tree search. More precisely we study the induced partition trees (i.e. gene trees that live inside species trees) and how topological rearrangements on species tree changes the associated partition trees. We characterise the changes for Nearest Neighbour Interchange (NNI), Subtree Pruning and Regrafting and Tree Bisection and Reconnection operations. We further generalize the concept of terraces to partial terraces and study their occurrence for real alignments using NNI neighbourhoods.

Secondly, we provide a phylogenetic terrace aware data structure (PTA) for the efficient analysis of concatenated multiple alignments. Using PTA and the rules developed to detect (partial) terraces in the presence of missing data one saves computational time by avoiding unnecessary recompilations.

Interactions between site rate model and DNA model selection

Michael Woodhams, University of Tasmania

In choosing a model (e.g., using Modeltest) as part of a phylogenetic analysis, we need to both select a DNA model, and also a rates-across-sites model. One might optimize one, then the other, and then stop - as the old Modeltest program did. Is this safe? I investigate using both Lie Markov and Time Reversible models.

Google Maps Clark Rd, Sandy Bay TAS 7005 to Tavern 42 Degrees South Walk 2.8 km, 35 min

Map data ©2015 Google 200 m

via Grosvenor St and Sandy Bay Rd/B68 35 min
2.8 km