MBSIG Workshop

"theoretical phylogenetics focus"

As part of **ANZIAM 2018**
Friday 9 Feb 2018

University of Tasmania, Sandy Bay, Clark Road, Maths and Physics Building, Lecture Th 1

Program

9.00am–9:30am	**Hello**
9:30am–10:30am	**Barbara Holland**, University of Tasmania *Extending the standard phylogenetic model to include convergence*
10:30am–11:00am	**Morning tea**
11:00am–11:40am	**Ben Kaehler**, Australian National University *Making the transition from time-reversible to non-stationary Markov chains in evolutionary analysis*
11:40am–12:00pm	**Julia Shore**, University of Tasmania *Pulling origin of life hypotheses out of aaRS*
12:00pm–1:30pm	**MBSIG AGM** followed by **Lunch**
1:30pm–2:10pm	**Venta Terauds**, University of Tasmania *Maximum likelihood distances for circular genomes*
2:10pm–2:30pm	**Afternoon tea**
2:30pm–3:10pm	**Michael Baake**, Bielefeld University *On some aspects of recombination*
3:10pm–3:50pm	**Andrew Francis**, Western Sydney University *Can we future-proof phylogenetic consensus trees?*
3:50pm–	**Discussion + Pub**

PTO for abstracts…

Abstracts

Michael Baake, Bielefeld University
On some aspects of recombination
(Joint work with Frederic Alberti and Ellen Baake)

The process of recombination, in the limit of large population size, leads to a nonlinear equation that can be solved in closed terms. The underlying structure leads to a Markov chain in continuous time that can used to find the solution, but also to construct a Lyapunov function. Some recent results will be reviewed, and different formulations will be compared.

Andrew Francis, Western Sydney University
Can we future-proof phylogenetic consensus trees?
(Joint work with David Bryant and Mike Steel)

Consensus methods are widely used for combining phylogenetic trees into a single estimate of the evolutionary tree for a group of species. But how robust are these methods to future information? If additional species are added to the original set of trees, will the expanded consensus tree simply be an expansion of the original consensus tree? In this talk I will formalise and answer this question.

Barbara Holland, University of Tasmania
Extending the standard phylogenetic model to include convergence
(Joint work with Jonathan Mitchell and Jeremy Sumner)

Phylogenetics provides a rich application area for continuous time Markov models. The most widely used methods of phylogenetic inference are based on models where evolution along an edge of an evolutionary tree is assumed to follow a Markov process, and where speciation events imply that evolution on child edges is conditionally independent given the state at the common ancestor.

Over the last few years we have been working on a class of models we call 'Convergence-Divergence models'. These allow for traditional speciation events where species diverge from a common ancestor but they also allow species to become more similar again. In this talk I will

1. give an overview of the "standard" phylogenetic model;
2. introduce the convergence-divergence model;
3. discuss potential areas of application: morphological convergence, modelling gene content, introgression;
4. discuss issues around identifiability for 3 and 4 taxon cases (with an intriguing link to the molecular clock in the three-taxon case).

Ben Kaehler, Australian National University
Making the transition from time-reversible to non-stationary Markov chains in evolutionary analysis

Some surprising results from mathematical statistics allow us to draw inference regarding the evolution of DNA. We relate present-day species to one another via their shared ancestors that existed millions of years ago. Unfortunately, ubiquitously adopted time-reversible

Markov models can be made to strongly support contradictory hypotheses. My work in this area has been to deploy non-stationary Markov processes that are demonstrably more consistent with the data and have the potential to tell us much more from the same genetic data. I will outline some of the challenges and exciting possibilities of moving from time-reversible to non-stationary models.

Julia Shore, University of Tasmania
Pulling origin of life hypotheses out of aaRS

Enzymes called amino-acytyl tRNA synthase (aaRS) are responsible for attaching amino acids to tRNA during the process of gene expression. In total there are 20 aaRS, one for each amino acid, and their structures strongly suggest that they can be divided into two classes: class I and class II (each class containing 10 aaRS). This has inspired a hypothesis for the origin of life: initially there were only 2 aaRS the first giving rise to class I and the second giving rise to class II.

I will present the results of a study that tested this theory by comparing empirical amino acid substitution rate matrices such as PAM1 and BLOSUM62 to rate matrices generated by the hypothesis. The results of this study show that hypothesised aaRS-class rate matrices do indeed fit empirical models significantly better than rate matrices generated by (random) alternative hypotheses. Further studies show that the polarity of amino acids is also an important factor in these lines of inquiry and that a rate matrix that takes into account both aaRS class and amino acid polarity fits empirical models better than any other method tried.

Venta Terauds, University of Tasmania
Maximum likelihood distances for circular genomes

Evolution of circular genomes is most commonly modelled via gene rearrangement, with evolutionary distance taken to be the minimum number of rearrangements needed to convert one genome into another. Recent work has suggested that maximum likelihood estimates (MLEs) of time elapsed are a better proxy for true evolutionary distance.

We present results to support this claim. By considering potential symmetries of rearrangement models, along with that of the genomes, and by applying techniques from group representation theory, we significantly reduce the combinatorial complexity of the distance calculations. This allows us to compare the properties of minimum distances to MLEs for genomes with up to eleven regions under several distinct rearrangement models.