WAIHEKE 2017

21st New Zealand Phylogenomics Meeting
Onetangi Beach, Waiheke Island
12-17 February 2017

CENTRE FOR COMPUTATIONAL EVOLUTION

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Organisers: David Welch, Alexei Drummond, Carmella Lee

Contents

Programme of Talks

Sunday 12 February 2017

4:30 – 6:00pm	**Registration, wine & nibbles**	

Monday 13 February 2017

8:30 - 9:00am	**Registration**	
9:00 - 9:15am	Welcome	
9:15 - 9:40am	Beta splitting as a model for transmission trees on contact networks	**Welch**
9:40 - 10:05am	Phylogenetic reconstruction of viral quasispecies dynamics	**Boskova**
10:05 - 10:30am	Phylodynamic model adequacy using posterior predictive methods	**Duchene**
10:30 - 11:00am	**Morning tea**	
11:00 - 11:25am	Evolution of Campylobacter jejuni within a long-term human host: adaptations, antimicrobial resistance and clinical implications	**Bloomfield**
11:25 - 11:50am	BEAST for extended sequencing of measles virus for molecular epidemiology of outbreaks	**Schulz**
11:50 -12:15am	Ancient DNA and the evolution of Hep B virus	**Patterson Ross**
12:15 - 2:00pm	**Lunch**	
2:00 - 2:25pm	Cross-species transmission occurs more frequently than co-divergence among viral families	**Geoghegan**
2:25 - 2:50pm	An RNA polymerase stuttering mechanism in paramyxoviruses	**Douglas**
2:50 - 3:15pm	Exploration of rust fungal evolution on native New Zealand Asteraceae	**Padamsee**
3:15 - 3:45pm	**Afternoon tea**	
3:45 - 4:10pm	Ancestral state reconstruction with parsimony	**Herbst**
4:10 - 4:35pm	Exploring the consequences of lack of closure for codon models	**Holland**
4:35 - 5:00pm	Closed codon models: Just a hopeless dream?	**Shore**
5:00 - 5:25pm	Tip of the iceberg: spectacular and unprecedented genome diversity in obligate endosymbionts of cicadas	**Simon**
5:25 - 5:50pm	Genome editing to the fittest?	**Drummond**

Tuesday 14 February 2017

9:15 - 9:40am	Efficient phylogenetics with diffusions	**Bryant**
9:40 - 10:05am	Models matter: Better models for better inference in biogeography	**Matzke**
10:05 - 10:30am	From fossils to trees: the place of Homo naledi in the hominin phylogeny	**Dembo**
10:30 - 11:00am	**Morning tea**	
11:00 - 11:25am	Signal or Noise? Investigating the impact of highly variable sites on phylogenetic inference	**Klaere**
11:25 - 11:50am	X-cactus trees and cactus tree metrics	**Hayamizu**
11:50 -12:15am	Investigating seasonal New Zealand influenza dynamics using genomic and mobile phone data	**Vaughan**
12:15 - 2:00pm	**Lunch**	
2:00 - 2:25pm	Towards a global language tree	**Bouckaert**
2:25 - 2:50pm	Phylodynamics and the the Galapagos of language evolution	**Gray**
2:50 - 3:15pm	Diversification models for language evolution – adequacy and empirical performance	**Ritchie**
3:15 - 3:45pm	**Afternoon tea**	
3:45 - 4:10pm	Languages, parasites and politics: Addressing phylogenetic non-independence, spatial autocorrelation, and environmental covariation in cross-cultural studies	**Schneemann**
4:10 - 4:35pm	Extending the multispecies coalescent to pinpoint substitution rate shifts	**Ogilvie**
4:35 - 5:00pm	Tree-Based Networks	**St. John**
5:00 - 5:25pm	Reconstructing a dated tree of life using phylogenetic incongruence	**Szöllősi**

Wednesday 15 February 2017

Free day, packed lunch provided

Thursday 16 February 2017

9:15 - 9:40am	Lost in space? Generalizing subtree prune and regraft to phylogenetic networks	**Linz**
9:40 - 10:05am	The subnet prune and regraft operation on classes of phylogenetic networks	**Klawitter**
10:05 - 10:30am	Representing orthology relations beyond trees	**Scholz**
10:30 - 11:00am	**Morning tea**	
11:00 - 11:25am	Phylogenetic networks modelling lateral gene transfer events and its reconstruction from triplets	**Pons**
11:25 - 11:50am	Seeing the trees for the tree-children	**Simpson**
11:50 -12:15am	Using subflattenings to rate splits and find trees	**Charleston**
12:15 - 2:00pm	**Lunch**	
2:00 - 2:25pm	Discussion on “small n, large p“ in phylogenetics and application of James-Stein estimator	**Chernomor**
2:25 - 2:50pm	Modeling site heterogeneity with posterior mean site frequency profiles for phylogenomic inference	**Minh**
2:50 - 3:15pm	Alignment uncertainty and phylogenetic reconstruction	**Floden**
3:15 - 3:45pm	**Afternoon tea**	
3:45 - 4:10pm	Lie-Markov models in the main stream: IQ-TREE implementation and phylogenetic accuracy	**Woodhams**
4:10 - 4:35pm	Scholar Relational Network	**Xie**
4:35 - 5:00pm	Doing somatic phylogenomics with DeNovoGear	**Cartwright**
5:00 - 5:25pm	Modelling PCR stochasticity and its effect on quantitative NGS experiments	**Haeseler**
6:45pm	**Dinner at Casita Miro**	

Friday 17 February 2017: Mini Symposium

10:00 - 11:00am	Mind Body Quantum Mechanics	**Kauffman**
11:00 - 11:25am	Chasing the tail: mathematical aspects of collectively autocatalytic sets	**Steel**
11:25 - 11:50pm	Adaptive self-preservation at the origin of life	**Egbert**
11:50 - 12:15pm	The end of the RNA World	**Wills**
12:15pm	**Lunch and finish**	

Abstracts

listed alphabetically by last name of presenter (presenter underlined where multiple authors)

Evolution of Campylobacter jejuni within a long-term human host: adaptations, antimicrobial resistance and clinical implications

Samuel Bloomfield[1], Jackie Benschop[1], Anne Midwinter[1], Patrick Biggs[1], David Hayman[1], Jonathan Marshall[1], Philip Carter[2] and Nigel French[1]
[1] mEpiLab, Palmerston North, New Zealand
[2] ESR, Keneperu, New Zealand

Campylobacteriosis is inflammation of the gastrointestinal tract as a result of Campylobacter. Most campylobacteriosis cases are acute and self-limiting, with symptoms and Campylobacter excretion ceasing after a few weeks. However, we recently identified an individual who has been excreting Campylobacter for ten years. Campylobacter isolates spanning this time period were collected from the patient, their genomes sequenced, their antimicrobial susceptibility patterns measured and their motility assessed. Phylogenetic analyses estimated that the isolates shared a date of common ancestor around the year 2000, coinciding with the onset of symptoms for the patient. Analysis also identified selection for changes in motility. Antimicrobial susceptibility testing suggested that the Campylobacter developed resistance to several antibiotics coinciding with periods of antibiotic therapy. We concluded that the patient was consistently colonised with Campylobacter that adapted to the internal environment of the patient. Our results demonstrate how phylogenetic techniques can be used to give insight into a patient's condition and the effect of antimicrobial treatment on long-term Campylobacter colonisation.

Phylogenetic reconstruction of viral quasispecies dynamics

Veronika Boskova
ETH Zurich, Switzerland

Especially in fast evolving and reproducing populations such as RNA viruses, the population of sequences present in one host at a time is often very diverse but also very repetitive. Deep-sequencing approaches allow for quantification of sequences and their frequency.

The amount of sequences from sequencing efforts represents a computational overload for current phylogenetic and phylodynamic model implementations in a full Bayesian framework. Heuristic approaches aim at reducing the computational burden by applying the inference models only to a subset of the sequences. One can only use the unique sequences, i.e. ignoring frequencies of the different sequences and instead assuming each one occurs only once. Alternatively, only a random subsample of the full dataset can be used to reconstruct the phylogeny and the corresponding population dynamics.

We investigated these heuristics in terms of how much loss of information on dynamic properties of the process occurs. We found that in both cases the computational time is drastically reduced, however, the parameter estimates are less exact, and/or less precise.
Based on the identified drawbacks of the heuristics, we propose a new tool for efficient reconstruction of viral dynamics from an alignment of unique sequences and their frequencies and implement it in BEASTv2.

Towards a global language tree
Remco Bouckaert
Centre for Computational Evolution and Department of Computer Science, The University of Auckland, New Zealand

As the tree of life is a phylogeny of all species, the global language tree is a phylogeny of all (circa 7500) languages. We aim to construct a global language tree using a varied set of sources such as structural language features, lexical data, geography, time constraints and topology constraints, and combine all this data in a Bayesian framework using BEAST. There are computational hurdles, data acquisition problems and modelling challenges. In this talk, I will give an overview of our attempts to tackle these issues.

Efficient phylogenetics with diffusions
David Bryant
University of Otago, New Zealand

Diffusion processes for gene frequency changes are some of the oldest mathematical models in population genetics. They have not, however, been widely used for `full-likelihood'-based inference, in part for the computational difficulties of computing `full-likelihoods'. I'll talk about recent progress in this area, including a new spectral-based method developed by me, Gordon Hiscott and Colin Fox.

Doing somatic phylogenomics with DeNovoGear
Reed A. Cartwright
School of Life Sciences, Human and Comparative Genomics Laboratory, The Biodesign Institute, Arizona State University, USA

DeNovoGear is a suite of methods to identify de novo mutations from next-generation sequencing data. DeNovoGear supports the joint analysis of multiple individuals and multiple samples from an individual. For instance, it can detect de novo mutations in a large pedigree where each individual has been sequenced multiple times. DeNovoGear uses hidden-data methods and the pruning/peeling algorithm to calculate the probability of de novo mutation based on sequencing data.

Here we report our success in adapting DeNovoGear to estimate whole-genome phylogenies from NGS data. We analyze a dataset consisting of 24 whole-genome samples taken from a single Eucalyptus melliodora tree. Our estimated somatic phylogeny is consistent with the physical tree topology (a positive control) and allows us to infer the location and patterns of somatic mutation in this individual. In the future our methods will be extended to study tissue relationships in mammals.

Using subflattenings to rate splits and find trees
Michael A. Charleston, Jeremy G. Sumner
University of Tasmania, Australia

A novel method of dimensional reduction for phylogenetic tree models was recently developed by Sumner,[1]. In that method, the patterns at each site in a multiple sequence alignment are counted, and compact sub-matrices of signed counts under a Hadamard (or similar) transformation are constructed. The choice of which transformation to apply is a matter for future investigation: at present we only require that it be orthogonal.

This extends the “flattening” concept of Allman, Kubatko & Rhodes, [2] such that these sub-matrices can be constructed for each possible bipartition, or split, of the taxa in the alignment. We refer to the sub-matrices as sub-flattenings. They are significantly smaller — of size that is quadratic vs exponential in the number of taxa involved. They also have analogous properties to the flattenings, in particular that, in the absence of stochastic (sampling) error, the rank of such matrices is minimal for splits that correspond to the underlying tree, whereas for splits that are not in the underlying tree, the rank is generically higher.

The mathematical elegance of the sub-flattenings is appealing of itself, but of great practical interest is also how the rank estimate varies in the presence of sampling error, caused by having finite sequences. Since stochastic error generally makes the sub-flattening matrices have full rank, we use singular value decomposition (after Erikson[3]) to obtain the singular values of each sub-flattening matrix, and then measure the sum of squares of all the excess singular values — that is, avoiding the first k terms. This error is smallest for sub-flattenings of minimal generic rank, so we attempt to use it as a proxy for the expected rank.

We have some preliminary results from a user-friendly C++ program that we have developed to investigate this idea further, showing that the rank error terms do indeed segregate well for splits that are part of the underlying tree. The speed of the implementation even at this early stage is promising, and we look forward to developing it further.

1. Sumner, J. Dimensional Reduction for Phylogenetic Tree Models *arXiv preprint*, 2016
2. Allman, E. S.; Kubatko, L. S. & Rhodes, J. A. Split scores: a tool to quantify phylogenetic signal in genome-scale data. *Systematic Biology*, 2016
3. Eriksson, N. Tree construction using singular value decomposition. In Pachter, L. & Sturmfels, B. (Eds.) *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005

Discussion on “small n, large p“ in phylogenetics and application of James-Stein estimator

Olga Chernomor and Arndt von Haeseler
University of Vienna, Austria

Undersampling influences many high-dimensional problems in various areas of research. A small sample size hampers an accurate estimation of particular characteristics of interest. In phylogenetics a multiple sequence alignment is considered to be a sample from multinomial distribution with category probabilities (i.e., site pattern probabilities) defined by the unknown evolutionary past. The alignment length, i.e. sample size, has a big influence on the inference. It is widely acknowledged, that the longer the alignment is the more information is available and the more accurate the result is. The number of observed site patterns in nearly all biological alignments is much smaller than the number of all possible site patterns for a given number of species. Therefore, phylogenetic inference is among the examples of high-dimensional problems with small sample sizes, i.e. "small n, large p". Here, we explore the question whether it is possible to improve the inference for short alignments. One of the helpful tools that are generally applied in "small n, large p" regimes is the James-Stein estimator. We will present the preliminary results from the application of such an estimator for site pattern probabilities in the context of phylogenetic inference.

From fossils to trees: the place of Homo naledi in the hominin phylogeny

Mana Dembo, Arne Mooers, and Mark Collard
Simon Fraser University, British Columbia, Canada

A new fossil hominin species, *Homo naledi*, was discovered in the Dinaledi chamber of the Rising Star cave system in South Africa. While the fossils at Rising Star represent the largest collection of hominin fossils ever recovered from a single site, we do not know where *Homo naledi* fits in the hominin evolutionary tree or how old it is. We used a large supermatrix of morphological characters and dated Bayesian phylogenetic techniques to carry out three analyses. First, we performed a dated Bayesian analysis to generate estimates of the evolutionary relationships of fossil hominins including *H. naledi*. Then we employed Bayes factor tests to compare the strength of support for hypotheses about the relationships of *H. naledi* suggested by the best-estimate trees. Lastly, we carried out a resampling analysis to assess the accuracy of the age estimate for *H. naledi* yielded by the dated Bayesian analysis. Our findings have a number of implications. Most notably, they support the assignment of the new specimens to *Homo*, cast doubt on the claim that *H. naledi* is simply a variant of *H. erectus*, and suggest *H. naledi* is younger than has been previously proposed.

An RNA polymerase stuttering mechanism in paramyxoviruses

Jordan Douglas
Centre for Computational Evolution and Department of Computer Science, University of Auckland, New Zealand

The paramyxoviruses are a subfamily of single-stranded RNA viruses with a 15-19kb genome. Notable paramyxoviruses include Mumps virus and Measles virus. The paramyxoviral P gene encodes 3 proteins with different functions - one protein from each reading frame. These 3 proteins coexist inside the same virus infected cell. Transcription of the P gene has a high insertion rate at the repetitive ‘edit site’ which results in uncoded guanine bases being inserted into the transcript. The number of guanines inserted determines the reading frame of the protein product.

It is believed that an RNA polymerase stuttering mechanism causes the stochastic number of inserts. The probability distribution of observing each insert size varies between viral species. Some P genes are unedited the majority of the time, while others have predominantly 1 or predominantly 2 bases added. This stuttering process also causes the expansion or contraction of microsatellites and other highly repetitive sequences in the human genome. I will discuss efforts I have made on the physical modeling of this stuttering process.

Genome editing to the fittest?

Alexei Drummond
Centre for Computational Evolution and Department of Computer Science, University of Auckland, New Zealand

No abstract.

Phylodynamic model adequacy using posterior predictive methods
Sebastian Duchene
Centre for Systems Genomics, University of Melbourne, Melbourne, Australia

Microbial pathogens, such as viruses and some bacteria, evolve sufficiently rapidly as to accumulate genetic change over the course of epidemiological processes. Phylodynamic methods take advantage of genomic information to infer how these pathogens spread through a host population. For example, genomic data from the recent Ebola virus outbreak in West Africa made it possible to infer the origin and basic reproduction number (R0) of the virus. These inferences, however, are contingent on whether the model provides an adequate description of the data. To this end, a Bayesian approach to assess model adequacy consists in sampling from the posterior to simulate data, which are known as posterior predictive simulations. If the model is adequate, the posterior predictive simulations should be very similar to the empirical data. However, the efficiency of this comparison is contingent on a test statistic that summarises some aspect of the data. I will present a simulation study of the performance of test statistics that summarise different aspects of phylogenetic trees. These results provide guidelines to assess the adequacy of some commonly used phylodynamic models, such as the Coalescent and Birth-Death Skyline. Finally, I illustrate the power of these methods in a range of bacteria and virus outbreaks.

Adaptive self-preservation at the origin of life
Matthew Egbert
Centre for Computational Evolution and Dept of Computer Science, University of Auckland

Self-preservation is generally considered to be a characteristic of living systems, but a number of non-biological systems such as motile oil-droplets, reaction-diffusion spots and hurricanes also demonstrate surprisingly adaptive and life-like forms of self-preservation. What role might this kind of behaviour have played in terms of facilitating the origins and/or early evolution of life?
In this talk, I will review my theoretical work in this area and describe a new experimental project where we are investigating the adaptive self-preservation demonstrated in abiotic dissipative structures.

Alignment uncertainty and phylogenetic reconstruction
Evan Floden, Maria Chatzou, Cedric Notredame
Centre for Genomic Regulation (CRG), Barcelona, Spain.

A multiple sequence alignment (MSA) is a set of statements implying evolutionary events within a set of homologous sequences. Whilst recent advances in the field have focused on improving the accuracy of MSA methods, particularly on large datasets, other developments have been made in quantifying the uncertainties within alignments. The existence of such uncertainty has major implications on the predictive capacities of MSAs. Quantifying and effectively accounting for the effect of this uncertainty is therefore a pressing issue. We approach this problem by exploring variability among alternative MSAs. In doing so we move away from the 'one correct alignment' paradigm and open the door on a world of alternative alignments in which the information from multiple, previously competing alignments can be harnessed. Using hundreds of alternative alignments generated from different guide-trees, dynamic programming tie-breaks and gap-penalties we introduce a new approach for phylogenetic tree reconstruction as well as a new support value, based on the bootstrap support, which aids in discriminating between the branches in a given tree.

Cross-species transmission occurs more frequently than co-divergence among viral families

Jemma Geoghegan
University of Sydney, Australia

The cross-species transmission of viruses from one host species to another is responsible for the majority of emerging infections. However, it is unclear whether some virus families have a greater propensity to jump host species than others. If related viruses have an evolutionary history of co-divergence with their hosts there should be evidence of topological similarities between the virus and host phylogenetic trees, whereas host jumping generates incongruent tree topologies. By analysing co-phylogenetic processes in 19 virus families and their eukaryotic hosts we provide the first quantitative and comparative measure of the relative frequency of virus-host co-divergence versus cross-species transmission among virus families. Notably, our analysis reveals that cross-species transmission is a near universal feature of the viruses analysed here, with virus-host co-divergence only apparent in a small number of families and always involving a subset of viruses. Despite the overall high topological incongruence among virus and host phylogenies, virus families that possess double-stranded DNA genomes exhibited more frequent co-divergence than the other virus families. At the other extreme, the virus and host trees for all the RNA viruses displayed high levels of topological incongruence. Overall, we show that cross-species transmission plays a major role in virus evolution.

Phylodynamics and the the Galapagos of language evolution

Russell Gray
Centre for Computational Evolution and School of Psychology, University of Auckland, NZ
Max Planck Institute for the Science of Human History, Jena, Germany
No abstract.

Modelling PCR stochasticity and its effect on quantitative NGS experiments.

Florian G. Pflug and Arndt von Haeseler
University of Vienna, Austria

Many protocols in modern-day biology use next-generation sequencing (NGS) as a quantitative method, i.e. to measure the abundance of particular DNA molecules. Then, any molecule that remains unsequenced causes a measurement error, and if this affects molecules non-uniformly, results are systematically biased. A major source of such biases is the Polymerase Chain Reaction (PCR), used to amplify DNA prior to sequencing. If it can be adequately modelled, its biases can be predicted and corrected for. Different models of PCR haven been proposed, but none have yet found their way into standard analysis pipelines, owing to a lack of parameter estimates for specific conditions. We thus focus on describing a model whose parameters can be estimated from actual experimental data, while still capturing the main source of biases. We show that this is achieved by viewing PCR as a branching process which, during each cycle, duplicates each DNA molecule with a certain probability, called the reactions efficiency. We combine this model with a simple model of the sampling behaviour of NGS and apply it to published RNA-Seq data. We demonstrate that the reaction efficiency can be estimated from the data, and that the data matches the models predictions well. In particular, we find that the model explains the main observed stochastic effects. Finally, we explore how well we can correct for unobserved molecules, and how much this improves the accuracy of the measured gene transcript abundances.

Ancestral state reconstruction with parsimony

Lina Herbst and Mareike Fischer
Ernst-Moritz-Arndt University, Greifswald, Germany

Evolutionary relationships of different species can be represented by phylogenetic trees. There exist different methods for reconstructing phylogenetic trees and ancestral (DNA) sequences. In my talk, I will focus on the latter.

The reconstruction accuracy (i.e. the probability of reconstructing the correct ancestral sequence) of each method is an important aspect. Considering Maximum Parsimony (one of the most frequently used methods for ancestral state reconstruction) an interesting question is to find the minimum number of present day species which have to be assigned one specific state, e.g. *A*, such that the last common ancestor is also assigned *A*. This minimal number depends on the tree topology as well as on the tree height. In my talk, I will consider a specific topology and generalize a result presented by Charleston and Steel in 1995.

X-cactus trees and cactus tree metrics

Momoko Hayamizu
The Institute of Statistical Mathematics, JST PRESTO, Japan

Research on phylogenetic networks is biologically importantas the process of evolution is too complex to be precisely representedby a phylogenetic tree (X-tree). In this talk, I will introduceX-cactus trees and cactus tree metrics to extend the notions of X-trees and tree metrics. The results include a generalisation of the fundamental theorem of phylogenetics (i.e., the four-point condition and uniqueness of X-tree representation) and a connection to split decomposition theory (i.e., total decomposability).

Exploring the consequences of lack of closure for codon models.

Michael Woodhams, Barbara Holland, David Liberles, Michael Charleston, Jeremy Sumner
University of Tasmania, Australia

Codon models of evolution are commonly used to try and identify selection. Selection, by its nature, is a heterogeneous process, i.e. it acts on some branches of the evolutionary tree and not others. Our previous work (Is the GTR model bad for phylogenetics?) has shown that when evolution occurs under a heterogeneous process it is important to consider the closure properties of models, as non-closed models will give biased estimates of evolutionary distance. It is relatively easy to show that codon models which account for the genetic code are not closed, this is a consequence of the fact that they are not linear (meaning that the sum of two codon rate matrices is not a codon rate matrix). This raises the concern that a single codon model fit to a heterogeneous process might be biased into either over or underestimating the effect of selection and likewise over or underestimating branch lengths. In this talk we will demonstrate the consequences of lack of closure for estimation of both ω (the selection parameter) and branch lengths.

The standard Muse-Gaut model is constructed from an underlying F81 model that acts independently at each codon position, and is then influenced by the genetic code via a parameter ω which modifies the rate of transitions between codons that code for different amino-acids. We consider using different DNA models within a similar structure. However, a DNA model being closed does not imply that the codon model constructed from it is closed (the genetic code prevents this). We were curious to see if using closed (Lie Markov) DNA

models helps to reduce the bias that arises from lack of closure in the codon model. The short answer is that it doesn't.

Mind Body Quantum Mechanics
Stuart Kauffman

Neurobiologists believe the mind brain system is and must be classical physics. For many, at some complexity, consciousness arises. This could be correct but faces what I will call the Stalemate: Such a mind can at most witness the world but, due to the causal closure of classical physics, cannot act upon that world. Such a consciousness must be merely epiphenomenal. Quantum biology is exploding, showing that quantum effects can and do arise at body temperature. Quantum mechanics allows a partially quantum mind to have ACAUSAL consequences for the "meat" of the brain. I shall discuss these issues and the newly discovered Poised Realm, hovering reversibly between quantum and "classical" behaviors, as a new basis both for the mind body system and a new class of constructable and evolvable "computers" which are not algorithmic, Trans Turing Systems.

Signal or Noise? Investigating the impact of highly variable sites on phylogenetic inference
Steffen Klaere, Ole Geldschläger, Jale Basten, and Daisy Shepherd
Department of Statistics, The University of Auckland, New Zealand

Tools to select a best model from a set of given models are abundant in phylogenetics. However, as has been pointed out repeatedly since the 1980s, an essential step in a statistical inference, the data-to-model fitness remains under appreciated. There have been few attempts at introducing general tests of fitness with mixed success and little to no implementation. Residual diagnostic tools on the other hand have been studied, primarily in the framework of tree of life inference, where the common ancestor is so old that some sites accumulated so many substitutions that they might support an alternative history. This field has produce multiple indices to assess the noise-level of a site, and thus its variability. The usual process is the to remove noisy sites until the inferred topology remains stable. There seems to be a strong disagreement about which approach is most suitable for the questions asked or how to provide an automated framework to provide a quantitative approach for de-noising.

Here, we will use a mixture of proposed indices to explore their relatedness, visualise the signal level in an alignment and provide a few conclusions and challenges.

The subnet prune and regraft operation on classes of phylogenetic networks
Jonathan Klawitter
Centre for Computational Evolution and Department of Computer Science, The University of Auckland, New Zealand

The subnet prune and regraft operation (SNPR) is a generalisation of the widely-used SPR operation on phylogenetic trees to phylogenetic networks. In this talk we discuss the SNPR operation on different classes of phylogenetic networks.

Lost in space? Generalizing subtree prune and regraft to phylogenetic networks

Simone Linz[1], Magnus Bordewich[2], Charles Semple[3]
[1]Centre for Computational Evolution and Department of Computer Science, The University of Auckland, New Zealand,
[2]Durham University, United Kingdom
[3]University of Canterbury, Christchurch, New Zealand

Over the last fifteen years, phylogenetic networks have become a popular tool in analyzing relationships between species whose past includes reticulation events such as hybridization or horizontal gene transfer. However, the space of phylogenetic networks is significantly larger than that of phylogenetic trees, and how to analyze and search this enlarged space remains a poorly understood problem. In this talk, we introduce a new operation---called subnet prune and regraft---that generalizes the widely-used subtree prune and regraft operation for phylogenetic trees to networks. We show that this new operation induces a metric on the space of all rooted phylogenetic networks on a fixed set of leaves and discuss several questions related to this new metric. This is joint work with Magnus Bordewich and Charles Semple.

Models matter: Better models for better inference in biogeography

Nick Matzke
The Australian National University, Canberra, Australia

Biology and biogeography are in the midst of a data revolution. Often, the scientific questions we wish to answer are limited more by the shortcomings in the models used for statistical inference, than by shortcomings in data. However, these problems can be overcome by creative thinking and programming to develop new, more realistic models, followed by rigorous statistical model comparison. I illustrate using my R package BioGeoBEARS to test new models in biogeography. I show that traditional models are often outperformed (across many clades) by models that allow more sophisticated models of dispersal, and dispersal functions that depend on distance, environmental distance, and/or evolving traits.

Modeling site heterogeneity with posterior mean site frequency profiles for phylogenomic inference

Bui Quang Minh[4], Huai-Chun Wang[1,2,3], Edward Susko[1,3], and Andrew J. Roger[2,3,5]
[1]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada,
[2]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada,
[3]Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University
[4]Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna and Medical University of Vienna, Austria
[5]Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Canada

Proteins have distinct structural and functional constraints at different sites that lead to site-specific preferences for particular amino acid residues as the sequences evolve. Heterogeneity in the amino acid substitution process between sites is not modeled by commonly used empirical amino acid exchange matrices. Such model misspecification can

lead to artefacts in phylogenetic estimation such as long-branch attraction. Although sophisticated site-heterogeneous mixture models have been developed to address this problem in both Bayesian and maximum likelihood (ML) frameworks, their formidable computational time and memory usage severely limits their use in large phylogenomic analyses. Here we propose a posterior mean site frequency (PMSF) model as a rapid and efficient approximation to full empirical profile mixture models for ML analysis. Compared to widely used empirical mixture models with k classes, our implementation of PMSF models in IQ-TREE (http://www.iqtree.org) speeds up the computation by k/3 fold and requires a small fraction of the RAM. Our simulations and empirical data analyses demonstrate that the PMSF models can effectively ameliorate long branch attraction artefacts.

Extending the multispecies coalescent to pinpoint substitution rate shifts
Huw Ogilvie
Australian National University, Canberra, Australia
Centre for Computational Evolution, University of Auckland

Molecular substitution rates vary between lineages and are correlated with a multitude of traits including metabolic rate, body size, and fecundity. Relaxed clock models are used to estimate differences in substitution rates between lineages, and Bayesian methods like BEAST jointly infer tree topology, divergence times and substitution rates from concatenated molecular sequences. However concatenation does not account for coalescent processes such as incomplete lineage sorting, and will confidently estimate substitution rate variation where none exists. We introduce species tree relaxed clocks, which model substitution rate variation within a multispecies coalescent framework. We have implemented three such models (uncorrelated log-normal, uncorrelated exponential, and random local clocks) in StarBEAST2. Preliminary results from a concatenation analysis suggest substantial rate variation between lineages of Australian skinks, and we will be using StarBEAST2 to identify where and when substitution rate shifts have truly occurred.

Exploration of rust fungal evolution on native New Zealand Asteraceae
Mahajabeen Padamsee and Eric McKenzie
Systematics Team, Landcare Research, Auckland, New Zealand

Rust Fungi (Pucciniales, Basidiomycota) are devastating obligate plant pathogens; in New Zealand, nearly two-thirds of all endemic rusts infect alpine or sub-alpine plant species, and many of the rust species are specialised on their hosts. There have been multiple studies that have investigated whether the evolution of rust fungi is tightly coupled with that of their host plants; however, recent work suggests that host jumps may have shaped rust fungal diversity. We set out to explore whether this is also the case with our endemic rust fungi. We focus on the rust fungi on native Asteraceae and use multi-gene phylogenetics to investigate whether co-evolution or host jumps have shaped rust fungal diversity.

Ancient DNA and the evolution of hepatitis B virus
Zoe Patterson Ross
University of Sydney, Australia

Ancient DNA (aDNA) samples expand the investigative time frame for evolutionary analyses, enabling more accurate estimates of evolutionary origins and substitution rates. In this study, hepatitis B virus (HBV) DNA from an Italian 16th century mummy was sequenced with hopes of addressing much-debated questions around HBV evolutionary history and dynamics. HBV

is a ubiquitous virus that has infected over ~2 billion humans, is chronically carried by ~350 million and causes ~1 million deaths annually. Statistical analysis of the cytosine deamination patterns of the next-generation sequencing reads of viral and host mtDNA fragments from our subject revealed patterns indicative of ancient provenance. However, phylogenetic analyses of the assembled whole-genome HBV inferred a close relationship to modern HBV sequences within a data set for which collection date of samples is known. Linear regression and a series of Bayesian MCMC analyses were used to test for temporal structure, as well as substitution rates and divergence times, with no temporal structure detected.

Phylogenetic networks modelling lateral gene transfer events and its reconstruction from triplets

Joan Carles Pons
University of Baleanric Islands, Spain

A variant of phylogenetic networks, called LGT networks, models specifically lateral gene transfer events, which cannot be properly represented by generic phylogenetic networks. This model allows to distinguish between the principal line of evolution of the species under study and the secondary lines determined by the LGTs. Phylogenetic networks have gained attention from the scientific community due to the evidence of evolutionary events that cannot be represented using trees and, especially, their reconstruction. We treat the problem of the reconstruction of LGT networks from substructures induced by three leaves. We first restrict ourselves to a class of LGT networks that are both mathematically treatable and biologically significant, called BAN-LGT networks. Then, we study the decomposition of such networks in subnetworks with three leaves and ask whether or not this decomposition determines the network. The answer to this question is negative, but if we further impose time-consistency (species involved in a later gene transfer must coexist) the answer is affirmative, up to some redundancy in the gene transfers that can never be recovered.

Diversification models for language evolution – adequacy and empirical performance

Andrew Miles Ritchie, Simon Y. W. Ho, and Nathan Lo
School of Life and Environmental Sciences, University of Sydney, Australia

Phylogenetic methods are increasingly being employed to reconstruct the history of human languages using lexical cognacy relations. Such studies have led to a number of exciting applications in recent years, such as testing theories of the origin of Indo-European languages, inferring the dynamics of Austronesian expansion, and shedding new light on the indigenous peopling of Australia. At the same time, these analyses raise the question of how models from evolutionary biology behave when applied to these new kinds of study system. In particular, tree-based models of diversification are ubiquitous in lexical dating, despite widely theorised deviations from these processes in language evolution. Using inappropriate tree priors has been observed to lead to severe inaccuracies in molecular dating results.

In order to assess these risks, we undertake a rigorous evaluation of the suitability of biological diversification models for analysing an array of language datasets. We observe substantial differences between dates inferred under birth-death and coalescent-based priors, with smaller differences between constant-rate and skyline models. We further characterise tree models under a comprehensive model adequacy framework. Our results make it clear that the choice of tree prior in linguistics, as in biology, requires careful consideration of the data and systems in question.

Languages, parasites and politics: Addressing phylogenetic non-independence, spatial autocorrelation, and environmental covariation in cross-cultural studies
Hilde Schneemann, Xia Hua, Marcel Cardillo, Simon Greenhill , Lindell Bromham
Research School of Biology, Australian National University, Canberra, Australia

Evolutionary thinking extends beyond the boundaries of biology, and evolutionary principles are commonly employed to explain human cultural variation. Just as for cross-species analysis, it is essential that spatial and phylogenetic relationships are taken into account when comparing languages, countries or cultures. I will illustrate how cultural phylogenies and methods from biology can be used to test the robustness of correlations between culture and environment. First of all, I will demonstrate how these methods can be employed to study the influence of environment on spatial patterns of language diversity. Secondly, I will present a study which tests the hypothesis that high parasite loads drive the evolution of human cultural traits. I will show that accounting for covariation and statistical non-independence - due to relatedness and spatial proximity - challenges causal relationships between parasite stress and culture. Our re-analysis raises new questions regarding the role of environment in explaining latitudinal patterns in human cultural variation.

Representing orthology relations beyond trees
Guillaume Scholz
University of East Anglia, Norwich, United Kingdom

Reconstructing the evolutionary past of a gene family is an important aspect of many genomic studies. To help with this, simple operations on a set of sequences, called orthology relations, have been introduced. In addition to being interesting from a practical point of view they are also attractive from a theoretical perspective in that a characterization is known for when such a relation is representable by a certain type of phylogenetic tree. For many reasons, it is however too much to hope for that real biological orthology relations satisfy this characterization. Rather than trying to correct the data, we propose representing orthology relations in terms of a structurally very simple phylogenetic network called a level-1 network, which generalizes phylogenetic trees.

To help compute such a network, we introduce a mathematical formalization of orthology relations in terms of the novel concept of a symbolic 3-dissimilarity. This is motivated by the biological concept of a "cluster of orthologous groups", or COG for short. We start with some background results on the topic, and then present our most recent results on this.

BEAST for extended sequencing of measles virus for molecular epidemiology of outbreaks
Helene Schulz[1], Joanne Hiebert[2], and Alberto Severini[1,2]
[1]Dept. of Medical Microbiology, University of Manitoba, Winnipeg, MB, Canada;
[2]National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada

Endemic circulation of measles virus in the Americas has been interrupted since 2002 and all cases result from importation from endemic areas. Ongoing elimination efforts have contributed to the decreasing diversity of measles virus genome, making it impossible to distinguish independent importations from local transmissions when relying on traditional WHO-recommended genotyping targets. Therefore, a higher-resolution genotyping or whole genome sequencing approach is needed.

Our objectives are: 1) calculate the rate of change and determine the relatedness of cases in outbreaks, and 2) analyze the sequences for hypervariable regions for development of high-resolution genotyping test. Based on whole genome sequencing, the 1000nt M/F non-coding region (MF-NCR) accounts for most of the variability seen in outbreak-associated viruses. BEAST revealed a clock rate of 3x10-3 and 7x10-4 changes/nucleotides/year for the MF-NCR and whole genome, respectively. Based on the node ages, the main outbreak is distinguished from other unrelated importations. BEAST analysis of whole genome sequencing and extended genotyping of measles virus will empower epidemiological investigation as global measles elimination is approached, and the diversity of the measles genome decreases.

Closed codon models: Just a hopeless dream?

Julia Shore, Jeremy Sumner, Barbara Holland
University of Tasmania, Australia

Woodhams et al. (2016) have found problems with codon models concerning the misestimation of ω, the ratio of synonymous to non-synonymous substitutions, when averaging along a lineage. Inspired by the successes of Sumner et al. (2012), we thought a possible solution to this problem was to find the Lie closure of a model to use as an alternative to the original. However, the Lie closures of even simple MG-style (Muse Gaut, 1994) models are ridiculously large, and hence are of no practical use. Investigations are being conducted to pinpoint the properties of MG-style models that make them mathematically intractable. We are also developing a theory of approximately closed MG-style models of manageable size.

Sumner, J. G., Fernández-Sánchez, J., & Jarvis, P. D. (2012). Lie Markov models. Journal of Theoretical Biology, 298, 16-31.
Woodhams, Liberles, Holland, Charleston, Sumner (2016) In preparation.
Muse, S. V., & Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Molecular Biology and Evolution, 11(5), 715-724.

Tip of the iceberg: spectacular and unprecedented genome diversity in obligate endosymbionts of cicadas

Chris Simon, Russ Meister, David Marshall, Kathy Hill, Elizabeth Wade, Chris Owen, Geert Goemans, Emily Lemmon, Alan Lemmon, and John McCutcheon.
University of Connecticut, USA

Insects with amino-acid poor diets are well known for harboring bacterial endosymbionts that supply needed nutrients and vitamins. These endosymbionts have highly reduced genomes and are transmitted maternally. In plant-sucking bugs (Hemiptera), they often come in pairs. One member of the pair (the dominant endosymbiont) typically produces 7 or 8 essential amino acids while the other produces two or three. In all large plant sucking bugs (Auchenorrhyncha) the dominant endosymbiont, Sulcia, generally evolves slowly. The partner endosymbiont differs among auchenorrhychan families. The cicada partner endosymbiont is called Hodgkinia. It was notable for its small size. Suprisingly, subsequently sequenced Hodgkinia from different cicada species were found to undergo lineage splitting followed by genome degradation with single individual cicadas housing multiple lineages of Hodgkinia whose relationships were tree-like.

We surveyed Hodgkinia from across the Family Cicadidae to determine the distribution and abundance of this peculiar lineage splitting/degradation and evolution phenomenon. Hodgkinia DNA was extracted from individual cicada bacteriomes, amplified using 16S eubacterial primers, and cloned. Multiple clones from single individual cicadas were sequenced and analyzed phylogenetically using Neighbor-Net in SplitsTree. While the closest outgroup family, Tettigarctidae, and some Cicadidae housed only a single lineage of Hodgkinia, many cicada species were found to house multiple lineages, some of which were extremely divergent. The North American periodical cicadas and some of their relatives exhibited the most extreme Hodgkinia evolution. Lineage splitting and degradation seems to be common in and unique to cicadas. Multiple species in a single genus of cicadas may or may not share the same set of Hodgkinia lineages.

Seeing the trees for the tree-children

Jack Simpson
University of Canterbury, New Zealand

Historically phylogenetic trees have been the principle means by which evolutionary information has been expressed and explored. Extremely elegant and clear their use is natural. However in more modern times it has been found that trees are too restrictive. Unable to express certain evolutionary events, more general structures have been considered in place of trees. Despite this, some of the most fundamental aspects of evolution are still believed to be tree-like. Consequently, the tree-like substructure of a network remains of interest. Particularly for my purposes, the variety of ways in which multiple trees may be represented by a single network. In this talk I will discuss the combinatorial properties that govern how and when a set of binary phylogenetic trees may be simultaneously displayed by a tree-child or related phylogenetic network.

Reconstructing a dated tree of life using phylogenetic incongruence

Gergely J. Szöllősi[1], Adrián A Davín[2], Eric Tannier[2], Bastien Boussau[2], Vincent Daubin[2]
[1]Eötvös University, Budapest, Hungary, MTA-ELTE "Lendület" Evolutionary Genomics Research Group
[2]CNRS, UMR 5558, Laboratoire de Biometrie et Biologie Evolutive, Universie de Lyon, France

Past evolutionary events recorded in the DNA of living organism might be the key to study the Early History of Life, a long period of time for which the fossil record is scarce and unreliable. The exchange of genes among different species have left tell-tale footprints of phylogenetic incongruence in extant genomes that can be detected with modern phylogenetic techniques. These genomic fossils can tell us which ancient species lived at the same time and thus can be used for dating studies.

I describe our recent results that show that transfers detected using gene tree-species tree reconciliations carry a strong time signal resembling those of paleontological fossils across the three domains of life. I also show that different methods for relaxing the molecular clock produce varying fit with the dating information conveyed by transfers, hence transfer events

can potentially be used to choose among competing alternatives. Finally, I discuss our plans for developing genome-scale dating methods that exploit the genomic fossils recorded by phylogenetic incongruence.

Tree-Based Networks
Katherine St. John
City University of New York

We address a question of Francis and Steel about phylogenetic networks and trees. They give a polynomial time algorithm to decide if a phylogenetic network, N, is tree-based and pose the problem: given a fixed tree T and network N, is N based on T? We show that it is NP-hard to decide, by reduction from 3-Dimensional Matching (3DM), and further, that the problem is fixed parameter tractable.

Chasing the tail: mathematical aspects of collectively autocatalytic sets
Mike Steel
University of Canterbury, New Zealand

A key step in the emergence of early life is the formation of a set of reactions that is simultaneously (i) self-sustaining (all reactions involve reactants that are either produced by other reactions or are available in the environment) and (ii) autocatalytic (each reaction is catalyzed by some molecule produced by the system). The notion of a 'collectively autocatalytic set' was pioneered by Stuart Kauffman, who demonstrated that such networks will appear with high probability in a simple binary polymer model that allows for sufficiently long polymers. In this talk, I describe a variety of mathematical, computational and experimental results that have been inspired by Kauffman's early ideas. In particular, I will highlight some recent work with Wim Hordijk (J. Math. Chem. 2016) on variable catalysis rates, and inhibition.

Investigating seasonal New Zealand influenza dynamics using genomic and mobile phone data
Tim Vaughan, André Lichtsteiner, Richard Hall, Sue Hang, John Graves, David Welch, Alexei Drummond

Seasonal influenza (H3N2) imposes a large annual health burden across the world. Phylogeographic inference plays an increasingly critical role in unravelling the global migration patterns followed by new strains that emerge from the tropics on a yearly basis. On smaller scales, however, there is much that remains a mystery. For instance, how many strains enter New Zealand each year? How do they circulate within the country? What are the main drivers of this movement?

In this talk we will present results of an ongoing project to address these and other questions through both separate and joint analysis of spatially sampled genetic data and mobile phone movement data. We show that these analyses provide evidence for multiple epidemic-spawning strains entering the country each year and for spatial structure in the genetic data. We consider the degree to which this is explained by the mobile phone-derived movement data and further use this data to help infer the seasonal movement of the virus.

Beta splitting as a model for transmission trees on contact networks

David Welch[1], Raazesh Sainudiin[2]
1 Centre for Computational Evolution and Dept of Computer Science, University of Auckland
2 Uppsala University, Uppsala, Sweden

In an SI epidemic where indiviudals can be either Susceptible or Infectious, the transmssion tree is a binary tree showing who infected whom and when. We consider the shape of a tranmission tree in the spread of an SI epidemic over a contact network – a simple graph where nodes represent individuals and edges represent the potential for transmission between two nodes. We show that for a few very simple networks, the distribution of the transmssion trees is exactly modeled by a biparametric Beta- splitting model.
As the underlying contact networks are interpolated smoothly across the three example networks, the maximum likelihood estimator of the Beta-splitting parameters obtained from the corresponding transmission trees change smoothly. This suggests that the Beta-splitting family of transmission trees can be thought of as being generated by contact networks that smoothly span over a large and rich class of networks beyond the three illustrative examples.

The end of the RNA World

Peter R Wills[1], Charlie Carter[2]
[1]Centre for Computational Evolution & Dept of Physics, University of Auckland, New Zealand
[2]Department of Biochemistry and Biophysics, University of North Carolina, USA

The now standard scenario for the origin of life is that it began with an autocatalytic system of RNA ribozymes that started making peptides in a coded fashion, leading to its own demise with a nearly complete takeover of catalytic functions by superior protein enzymes. However a new calculation indicates that proteins could not take over the key molecular biological functionalities of genetic information replication and translation without destroying them, necessitating an answer to the question of how uncoded proteins could have bootstrapped such processes, including exerting selection pressure adequate to identify and preserve the genes they needed to catalyse the creation of new generations with improved functionalities. The answer lies in the dynamics of gene-replicate-translatase systems, mixed RNA-protein autocatalytic sets that have an intrinsic tendency to self-organise in a stepwise fashion from primitive beginnings, progressively optimizing a map of amino acid functionality into the final orderly form of the universal genetic code. No equivalent mechanism exists in the hypothetical RNA World, which would have to rely on the blunt instrument of unguided protocell selection to create such a map, like finding a gold atom in a haystack. Recent experimental work lends strong support to these conclusions.

Lie-Markov models in the main stream: IQ-TREE implementation and phylogenetic accuracy

Michael Woodhams
University of Tasmania, Australia

The Lie-Markov DNA models have the property that their transition/Markov matrices are closed under multiplication. This means that (unlike most time reversible models) a time-varying process on an individual lineage can accurately be modelled as a homogeneous process.

Although we have been experimenting with Lie-Markov models for several years, there has been no production quality phylogenetic software implementing them. This deficiency has been rectified as Lie-Markov models are now in IQ-TREE.
We generate simulated alignments on a five taxon tree under an inhomogeneous (time-reversible rate matrix) DNA mutation process. We reconstruct tree topology using IQ-TREE under a homogeneous implementation of various models (both time reversible and Lie-Markov) and show that Lie-Markov models are superior to time reversible models in the presense of substitution process inhomogeneity.

Scholar Relational Network

Walter Xie
Centre for Computational Evolution and Dept of Computer Science, University of Auckland

Which research fields are connected to Genomics by research terms in Google Scholar? We analysed more than 33,000 researcher profiles in Google scholar to build a relational network started from "Genomics", and use the size of nodes and edges to display the relations among those common terms. Furthermore, a market analysis about users of BEAST and Geneious evolved out of this network will be demonstrated.

Participants

Samuel	Bloomfield	Massey University	s.bloomfield@massey.ac.nz
Veronika	Boskova	ETH Zurich	veronika.boskova@bsse.ethz.ch
Remco	Bouckaert	University of Auckland	remco@cs.auckland.ac.nz
Minh	Bui	University of Vienna	minh.bui@univie.ac.at
Reed	Cartwright	Arizona State University	cartwright@asu.edu
Michael	Charleston	University of Tasmania	Karen.Bradford@utas.edu.au
Olga	Chernomor	University of Vienna	olga.chernomor@univie.ac.at
Mana	Dembo	Simon Fraser University	mana.dembo@sfu.ca
Jordan	Douglas	University of Auckland	jdou557@aucklanduni.ac.nz
Alexei	Drummond	University of Auckland	alexei@cs.auckland.ac.nz
Sebastian	Duchene	University of Melbourne	sduchene@unimelb.edu.au
Matthew	Egbert	University of Auckland	m.egbert@auckland.ac.nz
Evan	Floden	Centre for Genomic Regulation	evan.floden@crg.eu
Jasmin	Franke	Massey University	frankejasmin@aol.com
Jemma	Geoghegan	University of Sydney	jemma.geoghegan@sydney.edu.au
Russell	Gray	University of Auckland / Max Planck Institute for the Science of Human History	rd.gray@auckland.ac.nz
Stefan	Gruenewald	CAS-MPG Partner Institute for Computational Biology, Shanghai	gruenewald@sibs.ac.cn
Momoko	Hayamizu	Institute of Statistical Mathematics	hayamizu@ism.ac.jp
Lina	Herbst	Ernst-Moritz-Arndt University, Greifswald	lina.herbst@web.de
Barbara	Holland	University Of Tasmania	Barbara.Holland@utas.edu.au
Anica	Hoppe	University Of Canterbury	mathmomike@gmail.com
Steffen	Klaere	University of Auckland	s.klaere@auckland.ac.nz
Jonathan	Klawitter	University of Auckland	jo.klawitter@gmail.com
Simone	Linz	University of Auckland	s.linz@auckland.ac.nz
Catherine	Macken	Independent researcher	c.macken@auckland.ac.nz
Nicholas	Matzke	Australian National	nick.matzke@anu.edu.au

		University	
Huw	Ogilvie	Australian National University	huw.ogilvie@anu.edu.au
Maj	Padamsee	Landcare Research	padamseem@landcareresearch.co.nz
Zoe	Patterson Ross	University of Sydney	zoe.pattersonross@sydney.edu.au
Joan Carles	Pons	University of Baleanric Islands	joancarles.pons@uib.es
Leonie	Raijmakers	University of Oxford	leonie.raijmakers@merton.ox.ac.uk
Andrew	Ritchie	University of Sydney	andrew.ritchie@sydney.edu.au
Hilde	Schneemann	Australian National University	Hilde.Schneemann@anu.edu.au
Guillaume	Scholz	University Of East Anglia	g.scholz@uea.ac.uk
Helene	Schulz	University of Manitoba	Helene.Schulz@umanitoba.ca
Julia	Shore	University of Tasmania	julia.shore@utas.edu.au
Chris	Simon	University of Connecticut	chris.simon@uconn.edu
Jack	Simpson	University of Canterbury	jrs149@uclive.ac.nz
Alexander	Skeels	Australian National University	alexander.skeels@anu.edu.au
Katherine	St. John	Lehman College-- City University Of New York	stjohn@lehman.cuny.edu
Mike	Steel	University Of Canterbury	mathmomike@gmail.com
Gergely	Szollosi	MTA TKI	sszolo@gmail.com
Sonja	Tuerpitz	University Of Canterbury	mathmomike@gmail.com
Tim	Vaughan	University of Auckland	t.vaughan@auckland.ac.nz
Arndt	Von Haeseler	University of Vienna and Medical University of Vienna	arndt.von.haeseler@univie.ac.at
David	Welch	University of Auckland	david.welch@auckland.ac.nz
Peter	Wills	University of Auckland	p.wills@auckland.ac.nz
Anja	Wilsdorf	Massey University	anjawilsdorf@gmx.net
Michael	Woodhams	University of Tasmania	michael.woodhams@utas.edu.au
Walter	Xie	University of Auckland	walter@cs.auckland.ac.nz
Stuart	Kauffman		

Useful information

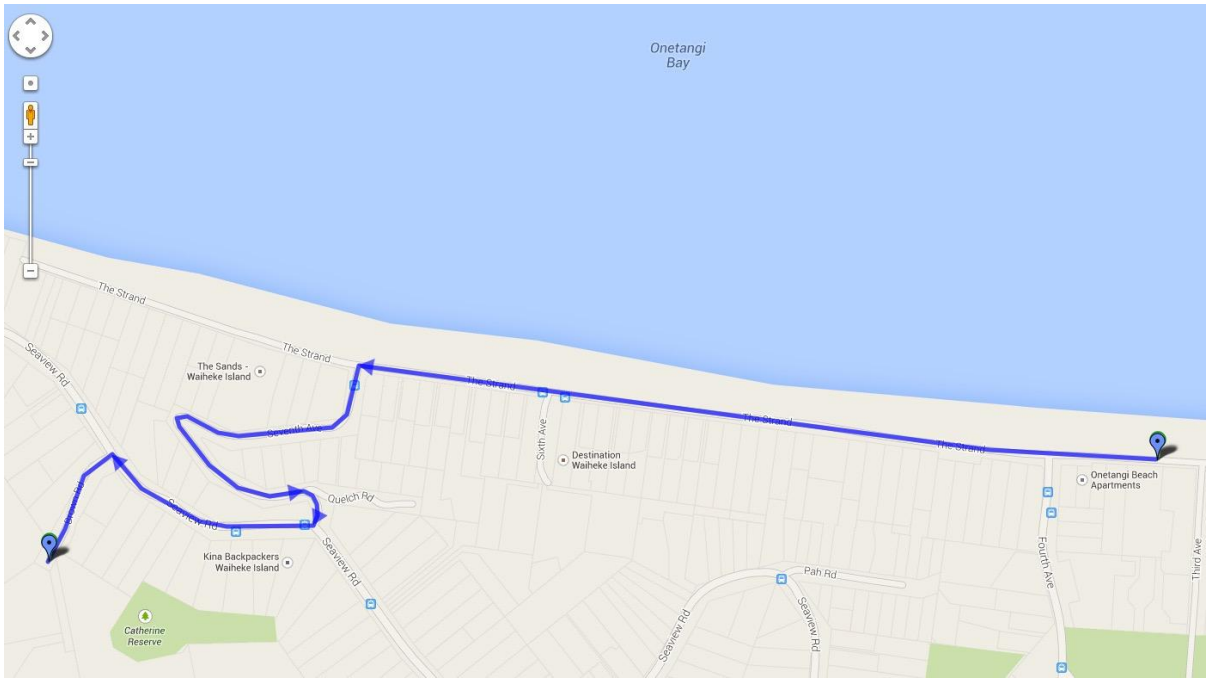
Local shops

Shopping is limited in the immediate area, the closest being the service station a couple of hundred metres away on Onetangi Road (take Fourth Ave). For a larger selection, the closest supermarket is in Ostend about 4km away. There are also lots of shops in Oneroa back near the Ferry terminal.

Conference dinner

The conference dinner is Thursday evening from 18:45 at Casita Miro (www.casitamiro.co.nz), 3 Brown Road, Onetangi, just above the western end of Onetangi Beach. It is about a 20-30 minute walk from the conference venue following the map below.

The meal is free for registered participants. You are welcome to bring guests at a cost of $80 per person. Please advise David or Alexei on Monday if you wish to bring a guest or cannot come.



Next year: Portobello 2018

The 22nd New Zealand Phylogenomics Meeting will be held in the lovely harbour-side setting of Portobello on the Otago Peninsula.

Dates: Feb 4 — Feb 9, 2018

Organisers

David Bryant (david.bryant@otago.ac.nz)
Mike Hendy (mhendy@maths.otago.ac.nz)
Marguerite Hunter (mhunter@maths.otago.ac.nz)

Otago Harbour, Photo by Steffen Klaere