Portobello '15

New Zealand Phylogenomics Conference

1-6 February, 2014
Portobello, Otago Peninsula
New Zealand

ALLAN WILSON CENTRE

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

1 Welcome

Welcome to the 19th Annual New Zealand Phylogenomics Meeting, the first to be held in Otago.

This meeting has been made possible through generation sponsorship from the Allan Wilson Centre and the Otago Department of Mathematics and Statistics.

We gratefully acknowledge help and support from

1. Wayne Cameron and Portobello Community Inc.
2. Peter Simkins and the staff of the Penguin Café
3. Marguerite Hunter, Dept. Mathematics and Statistics, University of Otago
4. Emersons Brewery
5. Dietrich Radel and the University of Canterbury Biomathematics Research Centre.

Coronation Hall

The Coronation Hall was built in Portobello in 1912. Constructed originally to house the Portobello Road Board a 250 grant was made for its construction from the Government Coronation Fund in 1911. A portion of the small lagoon area adjacent to the road and Latham Bay was reclaimed to create the site the hall sits on today. In 1921 the kitchen addition was added. Today the hall is operated by a local committee who manage the bookings, cleaning and renovations of this valuable community asset.

http://portobello.org.nz/our-community/portobello-coronation-hall/

2 Information

- All talks will be held in Coronation Hall.
- Standard length talks are 15 minutes + 5 minutes discussion.
- Short talks are 5 minutes + 5 minutes discussion.
- Load up all talks onto the conference laptop before your session: contact Gordon or Monika for help.
- An informal breakfast will be available in the Coronation Hall from 7:30am.
- Your registration includes breakfast, morning and afternoon teas, lunches, Fish and Chips (Monday), banquet (Tuesday) and Beer tasting + pizza (Thursday).
- Groceries etc. available from the Portobello store. Meals available from the pub or from 1908 Cafe Restaurant (make sure to reserve ahead).
- If you are in need of help, assistance, information, transport etc. talk to Mike Hendy, David Bryant or Marguerite Hunter.
- Limited wireless internet access is available in the hall. The password is “PHYLOBELLO15”. This is running through a mobile modem, so please don’t download large files.

Portobello 2015

CONFERENCE VENUE

Coronation Hall

(approx 250m to the store)

Store

Portobello Store : 7.30am-8pm

Eating places

Portobello Hotel: 11am - late

Penguin Cafe Portobello: 8am-4pm

1908 Cafe Restaurant: 5pm - late

Ric's galley: 5.30-8pm, Wed - Sun

Medical centre

Otago Peninsula Medical Centre

8.30am-5pm, Mo-Fr

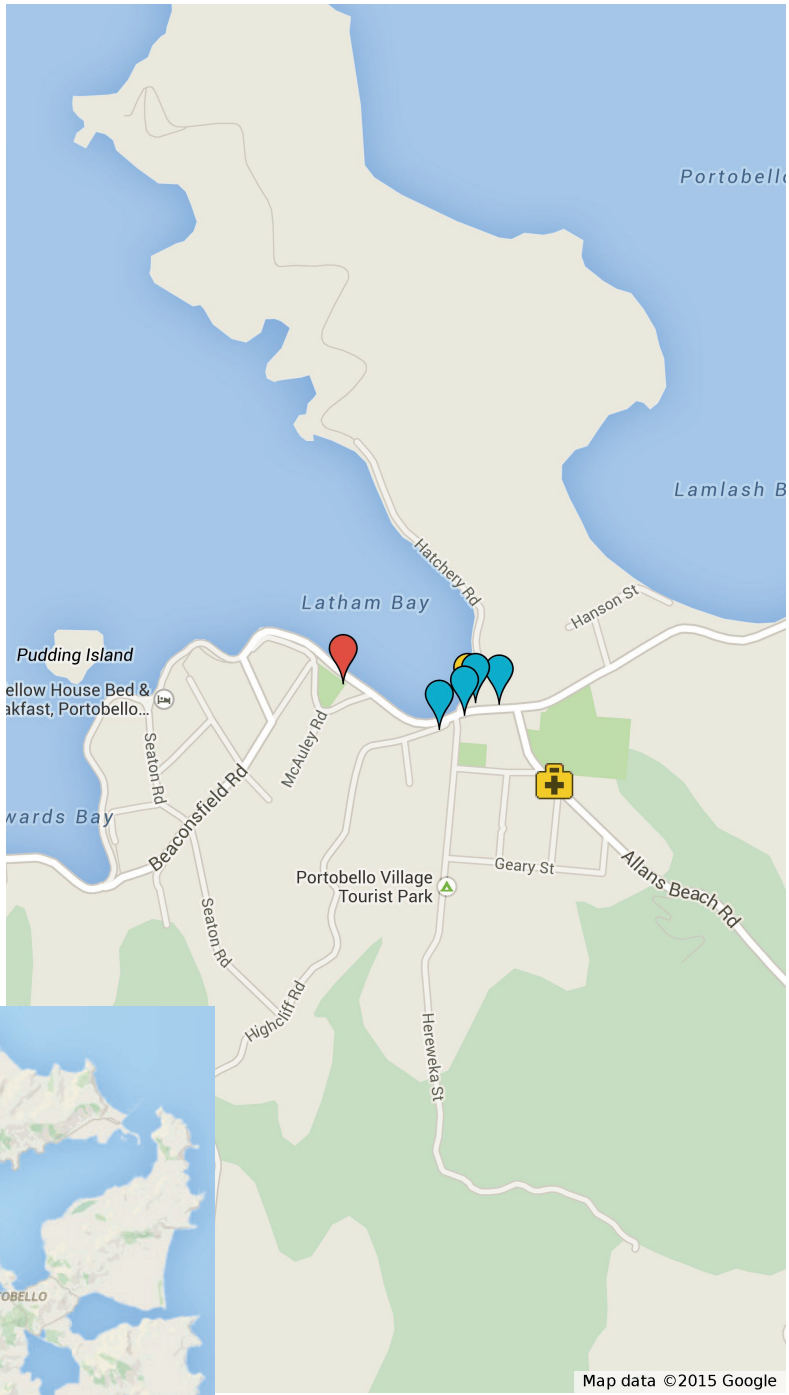
Tel: 03 4780880

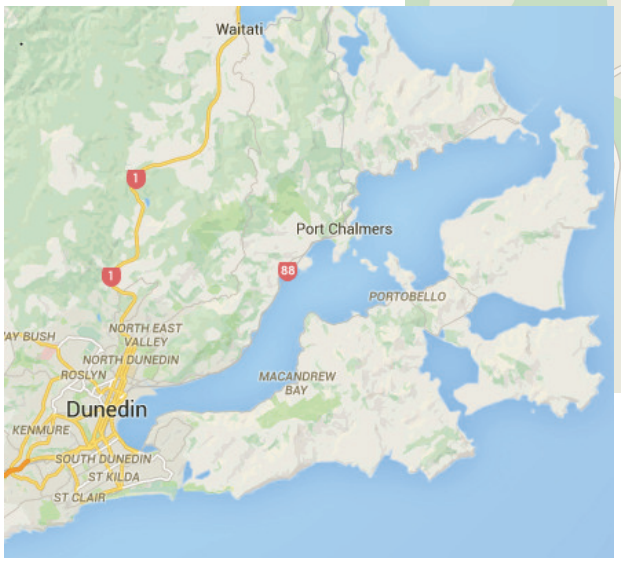
After hours: 03 4792900

IMPORTANT!!!

No petrol stations on the Peninsula

No ATMs either; EFTPOS usually available





3 Timetable

Sunday, February 1

17h00–20h00	Registration, Coronation Hall

Monday, February 2nd

8h30	Registration, Coronation Hall		
9h00	Announcements	*Mike Hendy*	
9h10	Phylogeography of six co-distributed New Zealand cicadas and their relationship to multiple biogeographic boundaries suggests a re-evaluation of the Taupo Line	*Chris Simon*	22
9h30	Birth and death of microsatellites across the avian phylogeny	*Bennet McComish*	19
9h50	Dynamic data structures and their applicability for coevolutionary analysis	*Ben Drinkwater*	11
10h10	*Morning tea*		
11h00	Encoding X-Trees with tree preorder	*Yaokun Wu*	24
11h20	What's all the fuss about Amborella?	*Pete Lockhart*	17
11h40	A heterotachy model for phylogenetics	*Remco Bouckaert*	9
12h00	*Lunch break*		
14h00	Flexible exact phylodynamic inference using the PMMH algorithm	*Timothy Vaughan*	23
14h20	Does the fungal community in kauri leaf litter change after invasion by a pathogen?	*Mahajabeen Padamsee*	21
14h40	Sifting through tangled trees: A particle-filtering method for Bayesian reconstruction of cophylogenies	*Arman Bilge*	9
14h50	From phylogenetics to multidimensional scaling	*Monika Balvočiūtė*	8
15h00	*Afternoon tea*		
16h00	The space of ultrametric phylogenetic trees	*Alex Gavruskin*	13
16h20	Oh what a tangled web we weave, when first we practice to misspecify our evolutionary models	*Stephen Crotty*	10
16h40	The role of Polynesian Outliers in the settlement of Eastern Polynesia	*Phillip Endicott*	12
18h00	Fish and chip dinner on Allans beach. Meet at Coronation Hall.		

Tuesday, February 3rd

9h00	Keep minding the gap: further evidence of bias in estimates of multiple sequence alignment	*Angela McGaughran*	19
9h20	Software testing in phylogenetics	*Michael Charleston*	10
9h40	Reassortment in influenza A virus	*Catherine Macken*	18
10h00	*Morning tea*		
11h00	Running faster with pruned trees	*Joseph Heled*	14
11h20	The (evolving) web of life	*Gulio Dalla Riva*	11
11h40	Potts models for detecting change-points in multiple sequence alignments	*Volkmar Liebscher*	17
12h00	*Lunch break*		
1400	Statistically sound structural phylogenetics - I	*Jane Allison*	8
1410	Statistically sound structural phylogenetics - II	*Ashar Malik*	18
14h30	Phylogenetic uncertainty can bias comparative analyses	*Robert Lanfear*	16
14h50	Simulating hybrid evolution and summary statistics for sets of trees	*Barbara Holland*	14
15h10	*Afternoon tea*		
16h00	Insights into the early epidemic spread of ebola in Sierra Leone provided by viral sequence data	*Denise Kühnert*	16
16h20	Holistic vineyard ecology	*Steffen Klaere*	16
16h40	How many characters to capture an evolutionary tree?	*Charles Semple*	22
18h00	Conference Banquet, Larnach Castle. Bus leaves Coronation Hall at 18h00. Bus leaves Larnach Castle at 21:30.		

Wednesday, February 4th

Excursion day. Collect packed lunch from Coronation Hall between 8h00 and 9h30.

Thursday, February 5th

9h00	Estimating the marginal likelihood	*Patricio Maturana*	18
9h20	Parametrization of Lie Markov models	*Michael Woodhams*	24
9h40	A new metric for the comparison of phylogenetic trees	*Michelle Kendall*	15
10h00	*Morning tea*		
11h00	RNA-seq and the Pitman Sampling formula	*Arndt von Haeseler*	23
11h20	Using conditional clades to pick better topologies for summary trees	*Huw Ogilvie*	20
11h40	Phylogenetic lassos and the triple conjecture	*Stefan Grünewald*	13
12h00	Comparing population history inferred from genetic and linguistic data in Central Asia	*Frederic Austerlitz*	8
12h10	*Lunch break*		
14h00	Parsing approximate tandem repeat sequences	*Michael Hendy*	14
14h20	Constructing a phylogenetic network that displays a given number of trees	*Paul Cordue*	10
14h40	Reconstructing past demography from complete genome sequence	*Flora Jay*	15
14h50	Fly-relationships: higher-level transcriptomics of early dipteran lineages	*Karen Meusemann*	20
15h00	*Afternoon tea*		
16h00	Tree-like reticulation networks	*Andrew Francis*	12
16h20	Nuclear versus mitochondrial trees: discordance is consistent with higher interspecific hybridization in freshwater fishes	*Graham Wallis*	24
16h40	IQ-TREE: Fast and effective stochastic algorithm for estimating maximum likelihood trees	*Minh Bui*	10
17h30	Beer Tasting, followed by pizza. Coronation Hall		

Friday, February 6th (Waitangi Day)

9h00	The Goldilocks Zone for finding novel RNA genes is perversely narrow.	*Ant Poole*	21
9h20	A simple model that predicts the statistical 'shape' of trees	*Mike Steel*	23
9h40	The total evidence approach with sampled ancestors	*Alexandra Gavryushkina*	12
10h00	*Morning tea*		
11h00	Nearest neighbor interchange meets phylogenetic networks	*Simone Linz*	17
11h20	Toward a global language phylogeny	*Luke Maurits*	19
11h40	A question about protein evolution	*David Penny*	21
12h00	*Lunch*		

4 Abstracts

Jane Allison
(Massey University, j.allison@massey.ac.nz)

Statistically sound structural phylogenetics I

Phylogenetic tools have been widely used to reconstruct evolutionary history. Until recently, phylogenetic relationships have largely been based on genetic sequence comparisons. It is proteins, however, that largely determine the phenotypic properties of an organism. Millions of the proteins discovered by high-throughput sequencing show little or no sequence-level similarity to known proteins. On the other hand, approximately two-thirds of proteins with no known function show structural similarity to known proteins. Recently, it has been shown that protein structural phylogenies allow clearer evolutionary relationships and functional roles to be established than those available in databases such as CATH and SCOP, particularly in cases where sequence divergence is too great for meaningful sequence-level comparisons, such as in the characterisation of deep evolutionary relationships. We are now aiming to take this one step further by including protein conformational dynamics in the generation of phylogenetic relationships, as a means of providing statistical rigour to structural phylogenetics.

Frederic Austerlitz
(Centre National de la Recherche Scientifique, austerlitz@mnhn.fr)

Comparing population history inferred from genetic and linguistic data in Central Asia

Genetic and linguistic data can be used jointly to infer the cultural and biological history of human populations. We compared the history of Central Asian populations as inferred from such kind of data. These populations belong to two distinct linguistic groups: Indo-Iranian or Turkic. For linguistic data, we used Swadesh word lists recording the basic vocabulary. Words were classified into cognates, i.e. homologous words related by common ancestry. For genetic polymorphism data, we used mitochondrial DNA sequences, Y-chromosome and autosomal microsatellites. For both linguistic and genetic data, we used the program starBeast to infer the genealogical tree of the populations. We compared the trees obtained for each kind of genetic data with the tree obtained for linguistic data. We found that the autosomal microsatellite tree had the best congruence with the linguistic tree. This may reflect the information gained by using many independent loci. Furthermore, the mitochondrial tree shows more congruence with the linguistic tree than the Y-chromosome tree, a surprising result in these patrilineal populations. Finally, we find several populations from one linguistic group to cluster genetically with the other linguistic group, which likely reflects their specific linguistic replacements.

Monika Balvočiūtė
(University of Otago, mokana@gmail.com)

From phylogenetics to multidimensional scaling

We show how a well known approach for computing phylogenetic trees and networks can be applied to a similar problem of multidimensional scaling (MDS). Neighbour identification and agglomeration strategy has been used in such popular algorithms as NeighbourJoining and Neigh-

bourNet, also in a new method for computing planar split networks FlatNJ. In the talk we will explain how point configurations in 2D are related to the planar split networks and how agglomeration can be applied for positioning points in a low-dimensional space.

Arman Bilge
(The University of Auckland, abil933@aucklanduni.ac.nz)

Sifting through tangled trees: A particle-filtering method for Bayesian reconstruction of cophylogenies

Systems of interdependent evolutionary histories, often represented by cophylogenies, are important models for genome evolution (i.e., gene trees versus species trees) and the coevolution between a symbiont and its host organism. A cophylogeny consists of a ”host” phylogeny, a ”guest” phylogeny, and a reconciliation, the mapping of ancestral guests to their hosts, and is determined by a series of events, including cospeciation, duplication, loss, and host-switching/horizontal transfer. Existing methods for reconstructing cophylogenies generally assume that the host phylogeny is known and that its population dynamics are deterministic and cannot consider uncertainties in dating and geographic location. We are developing an approach that uses Bayesian Markov chain Monte Carlo to perform joint inference on the host and guest phylogenies, reconciliation, divergence dates, and event rates. Because the model complexity makes it impossible to compute analytically the likelihood of a cophylogeny, we employ a particle filtering method that uses piecewise simulation and resampling to approximate the likelihood. We are implementing the sampler as a plugin for BEAST, an existing MCMC framework for Bayesian evolutionary analysis. Furthermore, the use of simulation makes it straightforward to extend our code to consider more complex models.

Remco Bouckaert
(University of Auckland, remco@cs.auckland.ac.nz)

A heterotachy model for phylogenetics

Most methods for performing a phylogenetic analysis based on sequence alignments of gene data assume that the mechanism of evolution is constant throughout time. The fact that some sites may evolve somewhat faster than others can be captured using a (gamma) rate heterogeneity model, and that some branches evolve faster than others by a relaxed clock model, but the underlying substitution models are typically assumed to be constant and time reversible. However, it has been observed that evolutionary constraints on proteins that make up some protein complexes appear to evolve over time, thus violating the assumption of constant evolution. Misspecification of models that assume constant evolution can lead to both errors in topology as well as biased date estimates. We propose a new model that allows gamma rate heterogeneity to change on branches, thus offering a more realistic model of heterotachy, and adding negligible computational cost to likelihood calculations. We illustrate its effectiveness with an example of green algae and land-plants.

Minh Bui
(University of Vienna, minh.bui@univie.ac.at)

IQ-TREE: Fast and effective stochastic algorithm for estimating maximum likelihood trees

(joint work with Lam Tung-Nguyen, Heiko A. Schmidt and Arndt von Haeseler)

We present IQ-TREE, a stochastic algorithm to infer ML trees. IQ-TREE combines different heuristic optimization searches such as evolution strategy, hill-climbing, and random perturbation of locally optimal trees that lead to a more efficient exploration of the likelihood surface.

For large biological DNA/amino acids alignments IQ-TREE found trees with higher likelihoods compared with likelihoods computed by other likelihood based methods (RAxML and PhyML) while requiring similar computing times. IQ-TREE is available as open source code at http://www.cibiv.at/software/iqtree.

Michael Charleston
(University of Tasmania, michael.charleston@sydney.edu.au)

Software testing in phylogenetics

Software that has not been tested cannot be regarded as reliable, yet there is a mass of bioinformatics software available whose testing is unclear. It is common to see software "testing" by pitting programs against each other and using secondary measures of quality: score of a function, memory footprint, time taken, number of contigs, etc. Software testing can be broadly divided into *verification* — are we building the product right? and *validation* – are we building the right product? It is more common to see validation of software, by such means as direct comparison across implementations, yet comparatively rare to see verification, which must be based on tests of output correctness. Perhaps that's not so surprising: computationally complex bioinformatics problems must often be solved with heuristics, which may even include a stochastic element. Testing programs that can have varying output given the same input clearly, therefore, presents significant challenges. This talk outlines some practices that we can use to test correct execution of bioinformatics software, and describes some work we have completed on some published and very commonly used phylogenetics software.

Paul Cordue
(University of Canterbury, paul.cordue@pg.canterbury.ac.nz)

Constructing a phylogenetic network that displays a given number of trees

This talk looks at the following question: Given a positive integer n, does there exist a binary phylogenetic network that displays exactly n trees?

Stephen Crotty
(University of Adelaide, stephen.crotty@adelaide.edu.au)

Oh what a tangled web we weave, when first we practice to misspecify our evolutionary models

Given the complexity of the evolutionary process, attempts to model it for the purposes of phylogenetic analyses will invariably result in model misspecification. Phylogeneticists have always been aware of this source of error but relatively little has been done to examine in detail the impact of specific types of misspecification and their interactions. I'll present models and simulated datasets that have been carefully constructed to exhibit specific types of model misspecification, illuminating their effects on ML estimation.

Giulio Dalla Riva
(Biomathematics Research Centre, gvd16@uclive.ac.nz)

The (evolving) web of life

The role played by evolutionary processes in the development of (unipartite) food webs is a major open problem in network ecology [1]. Conversely, building evolutionary models integrating the (food-web) niche construction process is challenging [2]. We propose a rigorous approach to quantify the effect of phylogenetic correlations on food-webs global structure [3]. The model is based on a radical take on the randomness-structure dichotomy in food webs. Interestingly, in two large empirical food webs we found evidences for a stronger phylogenetic signal in food-webs stochastic backbones rather than in food-webs fine wirings. We will review this results in the light of the broader discussion about eco-evo entanglement and its deep consequences for phylogenetic comparative methods [4].

[1] Rossberg, A. G. (2013). Food webs and biodiversity: foundations, models, data. John Wiley & Sons. [2] Barker, G., & Odling-Smee, J. (2014). Integrating Ecology and Evolution: Niche Construction and Ecological Engineering. In Entangled Life (pp. 187-211). Springer Netherlands. [3] Dalla Riva, Giulio V., & Stouffer, Daniel B. (submitted). Exploring the evolutionary signature of food webs' backbones using functional traits. [4] Nuismer, S. L., & Harmon, L. J. (2015). Predicting rates of interspecific interaction from phylogenetic trees. Ecology letters, 18(1), 17-27.

Ben Drinkwater
(School of Information Technologies, benjamin.drinkwater@sydney.edu.au)

Dynamic data structures and their applicability for coevolutionary analysis

The topology of evolutionary trees, and in particular their unbalanced nature, is a well studied problem within phylogenetics. Coevolutionary analysis, which considers the evolutionary relationships between two or more phylogenetic trees, has to date not considered the unbalanced nature of phylogenies as a means to reduce the complexity of algorithms used in coevolutionary analysis. In this work we leverage previous analyses to improve the asymptotic efficiency of existing coevolutionary analysis algorithms. Our technique is proven to provide an asymptotic reduction in the time and space required to infer coevolutionary events compared to previous algorithms by storing coevolutionary events within a dynamic data structure. Our proposed technique is validated using both synthetic and previously published biological data sets and is shown to provide a decrease in the actual running time similar to the asymptotic reduction achieved. We extend this result further by considering the reduction in terms of both time and space which is possible by using a combination of dynamic data structures and randomised sampling. Using this approach we are able to derive a linear time heuristic capable of handling larger coevolutionary data sets than has previously been possible.

Phillip Endicott
(Museum of Natural History Paris, phillip.endicott@gmail.com)

The role of Polynesian Outliers in the settlement of Eastern Polynesia

There are more than 20 islands where Polynesian languages are spoken which are traditionally referred to as "Outliers". Historically, this was because they lie outside the region considered to be the Polynesian homeland, which encompasses Tonga, Samoa, and several smaller islands including Niue, Tuvalu, Tokelau and Uvea. The "Outliers" lie to the west of the "homeland" region and are widely considered to be the result of a "backwash", associated with the geographical expansion of Polynesian speaking peoples sometime around $\sim$1000-1200 years ago. As such, "Outliers" are often treated as superfluous to understanding the major episodes of Polynesian prehistory, but there are data from comparative linguistics linking the more northerly islands with the languages of remote Eastern Polynesia. Here I present genetic evidence that might be used to test hypotheses generated from linguistics.

Andrew Francis
(University of Western Sydney, a.francis@uws.edu.au)

Tree-like reticulation networks

Hybrid evolution and horizontal gene transfer (HGT) are processes where evolutionary relationships may more accurately be described by a reticulated network than by a tree. In such a network, there will often be several paths between any two extant species, reflecting the possible pathways that genetic material may have been passed down from a common ancestor to these species. These paths will typically have different lengths but an average distance can still be calculated between any two taxa. In this talk, I will show that in some circumstances this average distance is unable to distinguish reticulate evolution from pure tree-like evolution. That is, there are reticulated networks in which the average distances satisfy the four-point condition. Another way to say this is that a metric satisfying the four-point condition does not imply the taxa are from a tree: it merely gives the existence of a tree.

Joint work with Mike Steel

Alexandra Gavryushkina
(The University of Auckland, sasha.gavryushkina@auckland.ac.nz)

The total evidence approach with sampled ancestors.

The 'total-evidence' approach, where molecular and morphological data of extant and fossil species are jointly used to infer species divergence times and macroevolutionary parameters, is a promising method. However current implementations of the method need to be improved in several aspects. In a Bayesian framework, an important component is the tree prior model which describes the tree branching process. Previous attempts to apply the total-evidence approach have used tree prior models that do not account for the possibility of fossil samples to be direct ancestors of other samples, that is, other fossils or extant species. Recently, Heath et al and Gavryushkina et al applied the fossilized birth-death model that explicitly models the sampling process and naturally allows for sampled ancestors to estimate divergence times based on molecular data and fossil

occurrence dates. Here we present a method to analyse morphological and molecular data in a unified Bayesian framework with models that account for sampled direct ancestors. We apply this method to extant penguins and their fossil ancestors.

Alex Gavruskin
(University of Auckland, alex@gavruskin.com)

The space of ultrametric phylogenetic trees

We consider two standard parameterisations of the space of ultrametric phylogenetic trees: (1) using lengths of coalescent intervals and (2) using times of divergence events. Using these parameterisations, we introduce two metric spaces on ultrametric phylogenetic trees, called τ-space and t-space, and compare them with existing models of tree space. We formulate several formal requirements that a metric space on phylogenetic trees must possess in order to be a satisfactory space for statistical analysis, and justify them. We show that only a few known constructions of the space of phylogenetic trees satisfy these requirements. However, our results suggest that these basic requirements are not enough to distinguish between the two metric spaces we introduce and that the choice between metric spaces requires additional properties to be considered. Despite their similarity, the two parameterisations have significantly different geometric and algorithmic properties. We proved that shortest paths, Fréchet mean, standard variance, and some other geometric and statistical characteristics are efficiently computable in τ-space but not in t-space.

I will introduce the basic notions and ideas of this approach, formulate the results, and suggest potential implications of the results on phylogenetic inference.

This is a joint work with Alexei Drummond.

Stefan Gruenewald
(CAS-MPG Partner Institute for Computational Biology, Shanghai, stefan@picb.ac.cn)

Phylogenetic lassos and the triple conjecture

It is a classical result in phylogenetics that a tree can be reconstructed efficiently from its induced metric. More recently, Dress, Huber, and Steel studied the situation where some of the distance information is missing. Given a phylogenetic tree, they call a subset of the set of unordered pairs of taxa a strong phylogenetic lasso, if there is no other phylogenetic tree having identical distances on every pair in the subset. In my talk, I will give examples and some sufficient as well as some necessary conditions for a set of pairs to be a lasso, and I will solve the most prominent open question that the authors above asked: In a binary phylogenetic tree, every interior vertex defines a tripartition of the taxa set, as removing the vertex disconnects the tree into three components. The triple cover conjecture says that a collection L of unordered pairs is a strong phylogenetic lasso, if for every interior vertex, we can find one representative from every part of its tripartition such that all three pairs from those three taxa are in L.

Joseph Heled
(Auckland University, jheled@gmail.com)

Running faster with pruned trees

A large analysis of one tree (hundreds of taxa) and a large number (hundreds) of "features", where each feature is a length one sparse alignment can be sped up by using an alignment specific "pruned" tree which is identical to the main tree but with all non informative leaves removed. I will show how to quickly determine how to determine if the pruned tree has changed given the nodes which changed on the main tree, and how to reconstruct the pruned tree when there is a change

Michael Hendy
(University of Otago, mhendy@maths.otago.ac.nz)

Parsing approximate tandem repeat sequences

The $\mathbf{tandem\ repeat}$ (TR)

$$\xi = \cdots \texttt{GACAG CTACG CTACG CTACG CTACG CTACG CTACG CAAGA} \cdots$$

consists of 6 copies of the $\mathbf{motif}$ $\texttt{CTACG}$. ξ can be expressed as

$$\xi = \cdots \texttt{GACAG } (\texttt{CTACG})^6 \texttt{ CAAGA} \cdots .$$

Other representations of ξ are obtained by shifting the motif boundaries one nucleotide left or right,

$$\xi = \cdots \texttt{CGACA } (\texttt{GCTAC})^6 \texttt{ GCAAG} \cdots = \cdots \texttt{ACAGC } (\texttt{TACGC})^6 \texttt{ AAGAT} \cdots ,$$

with 6 copies of the motifs $\texttt{GCTAC}$ and $\texttt{TACGC}$ respectively. We refer to these as different $\mathbf{parsings}$ of ξ. For exact TRs the data does not offer any objective reason for preferring one parsing.

In approximate TRs, the motif copies are not identical, but are similar and can be aligned with a template motif. $\mathbf{Duplication\ History\ Trees}$ (DHTs), depicting the evolution of the TR from a single ancestral motif, can vary among different parsings. The parsimony score of a DHT can be a linear combination of the numbers of nucleotide substitutions and the numbers of duplication and deletion events. We propose the parsing(s) which gives a minimal DHT provides an objective criterion to discriminate among the parsings.

However the determination of a minimal DHT is computationally expensive, so we have proposed three easily determined heuristic criteria which correlate with the length of a DHT. We will introduce the rationale for each of these heurisitcs and report the result of our simulation study.

Barbara Holland
(University of Tasmania, Barbara.Holland@utas.edu.au)

Simulating hybrid evolution and summary statistics for sets of trees

We introduce a simulator of hybrid evolution. It is designed for use in conjunction with approximate Bayesian computation approaches in order to determine the relative importance of hybrid speciation and introgression compared to incomplete lineage sorting in producing patterns of incongruence across gene trees.

Important features of the new simulator are (1) a choice of models to capture the decreasing probability of successful hybrid species formation or introgression as a function of genetic distance between potential parent species; (2) the ability for hybrid speciation to result in asymmetrical contributions of genetic material from each parent species; (3) the ability to vary the rates of hybridization, introgression and speciation in different epochs; and (4) incorporation of the coalescent, so that patterns of incongruence due to incomplete lineage sorting can be compared to those due to hybrid evolution.

I will also discuss a suite of summary statistics we have developed for sets of gene trees in order to make use of the simulator in an Approximate Bayesian Computation framework.

Flora Jay
(National Museum of Natural History, Paris, flora.jay@mnhn.fr)

Reconstructing past demography from complete genome sequence

Approximate Bayesian Computation (ABC) has proven to be useful for inferring demography from microsatellite or SNP data. Whole-genome data are expected to be extremely rich in information about past demography but, because simulations were, until recently, computationally too costly, ABC methods have not been thoroughly tested on such very long sequences. Dense polymorphism data contain extra information that is not available from unlinked site polymorphisms, and should improve the reconstruction of demographic history. We investigate how summary statistics computed from sequences, such as the lengths of haplotypes shared between individuals, or the decay of linkage disequilibrium with distance, can be combined with classical statistics (eg heterozygosity, Tajima D) and efficiently integrated to an ABC framework. We then quantify their influence on the inference of demographic parameters, particularly in the presence of expansion. Furthermore, we describe how errors that are usually more frequent in sequence than SNP data impact the inference, and show that modeling the error process in the ABC framework increases accuracy.

Michelle Kendall
(Imperial College London, m.kendall@imperial.ac.uk)

A new metric for the comparison of phylogenetic trees

We propose a new metric for rooted, labelled trees which has some advantages over existing phylogenetic metrics. In particular, our metric captures something of the difference in the 'biological story' told by the trees, detecting differences in shape, lineages and, if desired, branch lengths. We will outline some of the applications where our metric has already proven useful. These include the comparison of tree inference methods, sorting the trees from an MCMC chain, and visualising tree space, which enables us to look for 'islands' (sets of trees with similar likelihood and topology). Finally we will describe some work in progress where we use the metric to find an average or consensus tree, and outline a possible method for detecting recombination. This is joint work with Caroline Colijn.

Steffen Klaere
(University of Auckland, s.klaere@auckland.ac.nz)

Holistic vineyard ecology

Winemaking has become one of the major industries in New Zealand. However, there are many challenges, environmentally, biologically and economically, that the industry is facing. One word that binds all of these challenges is ”sustainability”, which incidentally lacks a standard definition. Since labelling your wine ”sustainable” permits an instant increase in value to the product, the industry has a strong interest in a marketable definition of the term.

In this project, researchers from microbiology, virology, chemistry, environmental sciences, entomology, and a lone statistician (me) intend to holistically study a set of vineyards to provide this type of definition and provide novel ways of describing an ecosystem.

I will present some research from microbiology and sensory sciences projects that I was involved with, and describe my role, the challenges, and opportunities I see within this seven-year-project.

Unfortunately, I cannot provide sensory samples of this project, yet.

Denise Kühnert
(ETH Zurich, denise.kuehnert@env.ethz.ch)

Insights into the early epidemic spread of ebola in Sierra Leone provided by viral sequence data

The current Ebola virus epidemic in West Africa has been spreading at least since December 2013. The first confirmed case of Ebola virus in Sierra Leone was identified on May 25. Based on viral genetic sequencing data from 72 individuals in Sierra Leone collected between the end of May and mid June, we utilize a range of phylodynamic methods to estimate the basic reproductive number (R_0). A structured birth-death model allows explicit incorporation of the incubation period. Hence, we also estimate the expected lengths of the incubation and infectious periods of the virus. Finally, we use phylogenetic trees to examine the role played by population structure in the epidemic.

The median estimates of R_0 based on sequencing data alone range between 1.65-2.18, with the most plausible model yielding a median R_0 of 2.18. We estimate the expected length of the infectious period to be 2.58 days and an incubation period of just under 5 days.

Overall we show that sequencing data can robustly infer key epidemiological parameters. Such estimates inform public health officials and help to coordinate effective public health efforts.

Robert Lanfear
(Macquarie University, robert.lanfear@mq.edu.au)

Phylogenetic uncertainty can bias comparative analyses

Ancestral state reconstruction (ASR) is often used to explore the evolutionary history of traits that leave little or no trace in the fossil record. Many ASR studies have suggested that the number of evolutionary origins of complex traits is higher than was previously thought. The scope of such

inferences is increasingly rapidly, thanks to the availability of huge phylogenies and life-history databases. In this talk, I'll show how standard approaches to ASR can be misled by phylogenetic inaccuracy. Specifically, when the phylogeny is inaccurate, the number of transitions tends to be overestimated. The bias can be surprisingly large in realistic simulations, and cannot be overcome for by the most common approach to accounting for phylogenetic uncertainty: namely repeating the analysis on a large collection of possible trees.

Volkmar Liebscher
(Ernst-Moritz-Arndt University Greifswald, volkmar.liebscher@uni-greifswald.de)

Potts models for detecting change-points in multiple sequence alignments

Complexity penalised maximum likelihood is a versatile tool for detecting change points in time-series like data. Yet, for longer sequences the quadratic or cubic running time may prohibit direct application.

We discuss several approaches to make this method work for the detection of change points in multiple sequence alignments of moderate length.

Simone Linz
(University of Tübingen, simone_linz@yahoo.de)

Nearest neighbor interchange meets phylogenetic networks

The graph-theoretic nearest neighbor interchange (NNI) operation is one of the most popular tools to quantify the dissimilarities between two phylogenetic trees. In the first part of this talk, we will review the NNI operation which induces a metric on phylogenetic trees. In the second part, we will then introduce an NNI inspired operation on phylogenetic networks and investigate some of its properties for classes of relatively simple networks.

This is joint work with Katharina Huber, Vincent Moulton, and Taoyang Wu.

Peter Lockhart
(Massey University, p.j.lockhart@massey.ac.nz)

What's all the fuss about Amborella?

Genome Science and the combined efforts of a research community are enabling great progress in our understanding of biology. However, if there is one irksome thing about how Science is currently being done, it might be that some earlier lessons have already been forgotten. In phylogenetic reconstruction the issue of data-model fit still remains as problematic as ever. This situation is illustrated with green plant phylogeny, and perhaps the best example of this concerns the story of Amborella a shrubby plant from New Caledonia, and the youngest poster child for plant genome science. Touted for over a decade as the sister species for all other flowering plants, a recent exchange in Systematic Biology makes it clear why we need independent approaches to plant molecular evolution.

Catherine Macken
(University of Auckland, c.macken@auckland.ac.nz)

Reassortment in influenza A virus

Reassortment among the segments of the influenza A virus can generate novel viruses than cause influenza pandemics in humans. While the process of reassortment has been recognized for several decades, fundamental questions about reassortment remain. In this talk, we outline our investigations of the following questions: Do segments reassort as cassettes? How frequently do segments reassort? Does the influenza virus population exist as a mixture of reassorting and non-reassorting viruses? We developed a novel statistical analysis for detecting reassortant viruses, called the two-time test. Using this methodology, first on the four segments coding the replication complex of avian influenza viruses of all serotypes, and, second, on the six segments coding the internal proteins of avian influenza A (H5N1) viruses (bird flu) we found that circulating genotypes are replaced by novel reassortant genotypes, with no obvious co-ordination among segments when reassorting. Further, for avian H5N1 viruses, we found that reassortment of just three of the internal genes had the most importance, and HA and the constellation of internal genes may be jointly important in the emergence of dominant variants.

Ashar Malik
(Massey University, asharjm@gmail.com)

Statistically sound structural phylogenetics - II

Phylogenetic analysis allows evolutionary relationships to be studied, looking at the differences between sequences of genes or proteins. The sensitivity of this analysis is highly dependent on the degree of difference; which beyond a certain degree of divergence loses its ability to generate meaningful results. It has long been known that structure of a protein is more conserved than the sequence that generates it. Even with a high degree of divergence between the sequences being probed, the similarity of the structure can still reveal more meaningful results. The use of structure for studying evolutionary relations has previously been probed and shows a high degree of promise. I will present some preliminary results of a study that uses molecular dynamics simulations to generate multiple conformations of a structure, allowing us to generate statistically significant phylogenetic relationships. Haemoglobin, a protein that has a very well characterized evolutionary relationship, is used in the development of this method. The high degree of difference in the sequence of the alpha and the beta units making up the tetramer, in contrast to the high degree of structural similarity, make it an ideal model system for use with this investigation.

Patricio Maturana
(University of Auckland, p.russel@auckland.ac.nz)

Estimating the marginal likelihood

Bayes factors are a common method for statistical model selection. The computation of such factors is based on the marginal likelihood, an integral that can be hard to estimate depending on the model complexity. The models employed in phylogenetic inference are examples of high complexity motivating the development of different methods. In this talk we present a statistical

example where traditional methods do not work and we present a more general method called Nested sampling, a Bayesian algorithm which provides the means to estimate the marginal likelihood and also yields posterior samples. We apply this method for model selection and parameter inference under different phylogenetic models considering a data set analysed previously in the literature.

Luke Maurits
(University of Auckland, luke@maurits.id.au)

Toward a global language phylogeny

This talk will give a high-level overview of a project currently underway at the University of Auckland aimed at building a phylogenetic tree relating the world's roughly 7000 currently spoken languages. This is an ambitious project which involves combining several different types of data from multiple sources, processing the combined data to ensure suitability, compatibility and consistency, and modelling several different evolutionary processes, both linguistic and otherwise. I will introduce and motivate the problem the project is aimed at solving, discuss some of the problems faced so far and the solutions the project team has developed and discuss anticipated future problems and directions.

Bennet McComish
(University of Tasmania, bennet.mccomish@utas.edu.au)

Birth and death of microsatellites across the avian phylogeny

The birth-death process by which microsatellites appear and disappear is an important component of microsatellite evolution, but is often confounded with the mutation process of extant microsatellite loci. With a view to examining this birth-death process separately, we have identified microsatellite loci in an alignment of 48 complete avian genomes, and have developed a pipeline for mapping the locations of microsatellites in each genome to coordinates in any other. This enables us to identify homologous microsatellite loci across the alignment, and any changes in repeat motif along the tree. We can then use ancestral state reconstruction to estimate when each microsatellite was born on the avian phylogeny, as well as estimating the likelihood of a microsatellite loci being lost.

Angela McGaughran
(CSIRO, Canberra, angela.mcgaughran@csiro.au)

Keep minding the gap: further evidence of bias in estimates of multiple sequence alignment

Multiple sequence alignments (MSAs) are pivotal in phylogenetics but known to be error-prone. In particular, gap placement is central to MSA inference but the propensity of alignment methods to insert erroneous gaps is unclear. We tested ten MSA methods (ClustalO, DIALIGN-TX, GramAlign, MAFFT, MUSCLE, PASTA, PRANK, Probalign, ProbCons, T-Coffee) using nucleotide and amino acid sequences simulated on trees of various topology and length. Sequence data were

generated without gaps, so inferred alignments containing gaps represented MSA program errors. For each test, we quantified the amount and distribution of erroneous gaps. All methods inserted minimal gaps at shorter tree lengths but at higher lengths ($>$16), inserted gaps in a manner dependent on data type, tree topology, and tree length. Overall, different aligners appear suitable for different alignment problems, depending on the characteristics of the dataset involved, and users may wish to adjust parameters of their chosen MSA method to reflect this. Combined with information on computational resource requirements, our results provide MSA software users with a detailed basis for examination of gap placement behaviour under the tested conditions.

Karen Meusemann
(National Research Collections Australia, CSIRO, karen.meusemann@csiro.au)

Fly-relationships: higher-level transcriptomics of early dipteran lineages

Phylogenetic relationships between the earliest evolving lineages of Brachycera are not resolved while there is no doubt that this is a monophyletic group. We present the first results of our phylogenomic analysis using $\sim$2000 orthologous single copy genes (approx. 500 kb of aligned amino acid sites) derived from transcriptome data from the 1KITE project. Using $\sim$40 terminal taxa spanning the phylogenetic region of interest, we performed partitioned analyses of our dataset based on protein domains and optimized partitioning schemes to improve modelling during phylogenetic inference. We reduced the amount of missing data by selecting only partitions for phylogenetic inference that have the necessary taxon coverage for the question at hand. Furthermore, we assess whether our primary dataset matches the compositional and time-reversible conditions assumed in phylogenetic inference.

Huw Ogilvie
(Australian National University, huw.ogilvie@anu.edu.au)

Using conditional clades to pick better topologies for summary trees

Reducing posterior samples of phylogenetic trees to a summary tree is necessary to communicate results and conduct downstream analyses (for example, inference of macroevolutionary rates and modes). Better summary trees would improve taxonomies and those downstream analyses. Existing methods to construct summary trees identify the most likely tree topology directly from the posterior sample. However, because the number of possible topologies increases hyperexponentially with the number of taxa, the probability of any given topology being present in a posterior sample declines as the number of taxa increases. The posterior sample of conditional clades (which are bifurcations conditional on the monophyly of the parent node) may be more complete and representative, as the number of conditional clades grows merely exponentially with the number of taxa. Because topology probabilities can be calculated from conditional clade probabilities, even for topologies absent from the posterior sample, an improved method to construct summary trees uses conditional clade probabilities to identify the most likely topology from the set of all possible topologies. We present a novel algorithm to implement this method, the relative performance of the algorithm and future research directions.
Co-authors: Alexei Drummond, Alex Gavryushkin

Mahajabeen Padamsee
(Landcare Research, padamseem@landcareresearch.co.nz)

Does the fungal community in kauri leaf litter change after invasion by a pathogen?

Kauri (*Agathis australis*, Araucariaceae) is restricted in distribution to the Northern tip of the North Island of New Zealand. Living to 1,500 years or more and having trunks up to 3 m diam., *A. australis* exerts enormous influence on surrounding forest composition and structure. Individual kauri trees produce massive amounts of litter, up to 2m deep that represents a significant store of C and N and provide varying habitat niches for complex fungal communities. However, information on the diversity of fungi under *A. australis* involved in nutrient recycling is sparse. Additionally, since the 1970s these trees have been under threat from the exotic invasive pathogen, *Phytophthora* taxon Agathis (PTA) that causes kauri dieback. We used pyrosequencing and soil analyses to investigate the kauri leaf litter fungi under asymptomatic trees and diseased trees. QIIME and UPARSE were used to assign sequence reads to OTUs. The OTUs were analysed to understand the impact of the invasion of PTA on fungal diversity at a site, and to increase basic bio-systematic knowledge of kauri-influenced biodiversity.

David Penny
(Massey University, d.penny@massey.ac.nz)

A question about protein evolution

There is the very important question about the origin of proteins, and here we need testable models about several issues; including reasons firstly for a triplet code, and then secondarily for the origin of the code itself. One testable hypothesis is given in terms of the Eigen limit for the length of sequence that can be copied before error catastrophe occurs. The error rate should be reduced if triplets of nucleotides are added, although there is simultaneously an increase in the number of possible triplets. However, this would still lead to much longer sequences being possible. It could be that tRNAs were initially important to add these triplets to the growing chain, and this could predict that the code existed before it was used to make proteins. Under this hypothesis there may have been a simple switch from RNA to proteins, and would also fit with the original Gilbert model. Under this hypothesis, short peptides would be important for catalysis, even though they initially might not have been coded. Coding may have evolved later?

Ant Poole
(University of Canterbury, anthony.poole@canterbury.ac.nz)

The "Goldilocks Zone" for finding novel RNA genes is perversely narrow.

Noncoding RNAs are integral to a wide range of biological processes. While genome sequencing is now routine, our capacity to identify noncoding RNA (ncRNA) elements remains limited. With transcriptomics fast becoming a standard tool, it may be possible to identify ncRNAs directly from transcriptional outputs. However, it is non-trivial to distinguish functional RNAs from transcriptional noise. Comparative analyses may help to identify expressed, evolutionarily-conserved noncoding regions of a genome. To that end, we screened for unannotated ncRNAs across $>$400 public transcriptomes from 37 strains of Bacteria and Archaea. We identified over 900 putative RNAs, but find

that the capacity to identify ncRNAs from this data is strongly dependent on phylogenetic sampling. In stark contrast to protein-coding genes, the phylogenetic window for effective use of comparative methods is perversely narrow: aggregating public datasets only produced one phylogenetic cluster where these tools could robustly separate unannotated noncoding RNAs from a null hypothesis of transcriptional noise. Thus, for RNA element discovery, a phylogeny-informed sampling approach is more effective than analyses of individual species. Our work reveals a narrow "Goldilocks Zone" (where species are not too similar and not too divergent) for RNA identification.

Ref: Lindgreen et al. (2014) PLoS Comp Biol doi:10.1371/journal.pcbi.1003907

Charles Semple
(University of Canterbury, charles.semple@canterbury.ac.nz)

How many characters to capture an evolutionary tree?

Evolutionary trees illustrate the ancestral relationships of present-day species. Characters describe attributes of the species and are the data typically used to infer such trees. If we are to recover the correct tree, amongst the exponentially many possible evolutionary trees, then it is necessary that the character data is consistent with just a single tree, that is, the characters capture a tree. How many characters does it take to capture an evolutionary tree?

Chris Simon
(University of Connecticut, chris.simon@uconn.edu)

Phylogeography of six co-distributed New Zealand cicadas and their relationship to multiple biogeographic boundaries suggests a re-evaluation of the Taupo Line.

We compare phylogeographic patterns across ecologically divergent, co-distributed taxa in the light of NZs palaeohistory and test the significance of proposed biogeographic boundaries. MtDNA from six co-distributed cicada species were analysed using phylogenetic methods, molecular dating techniques and AMOVA. Five species (Kikihia ochrina, K. cutora, K. laneorum, K. cauta, and K. scutellaris) show various degrees of intraspecific concordance with biogeographic boundaries found in previously studied taxa. Clade splits of forest species correlate with the Kauri Line and/or Northland Line, while splits of scrub/hill species correlate with the NW-SE Line. Four species (Kikihia ochrina, K. cutora, K. laneorum, and K. cauta) diversified before the last glacial maximum (LGM, 20,000 ya), while two species (K. scutellaris, and K. dugdalei) show only post-LGM diversification. Despite species idiosyncrasies, we see the imprint of shared palaeoclimatic/geological events. We distinguish between the importance of biogeographic lines as the demarcation between older genetically diverse and newer genetically depauperate populations versus biogeographic boundaries between sister-clades. We also stress the importance of dating clade splits to ensure consistency with explanations for biogeographic lines in question. We suggest that the Taupo Line has been overemphasized as a biogeographic boundary while the importance of the NW-SE mountain axis has been overlooked.

Mike Steel
(University of Canterbury, mathmomike@gmail.com)

A simple model that predicts the statistical 'shape' of trees

A statistical analysis of large numbers of evolutionary trees built from present-day sequence data reveals two puzzling features. Empirical trees are typically less balanced, and the distribution of their branch lengths are less extreme, than would be expected under various standard models of speciation and extinction. In this talk, I describe a simple model that appears to fit both features of observed tree shape (Joint work with Tanja Stadler, Klaas Hartmann and Oskar Hagen).

Timothy Vaughan
(University of Auckland, tgvaughan@gmail.com)

Flexible exact phylodynamic inference using the PMMH algorithm.

Since its introduction by Grenfell et al. in their 2004 Science paper, the term "phylodynamics" has been used to describe a wide variety of relationships between population and evolutionary dynamics, almost exclusively in the context of epidemiology. One of the primary goals of phylodynamic *inference* is to exploit these relationships in order to learn about the population dynamics of rapidly evolving organisms from sampled genetic sequence data.

In a Bayesian setting, this goal is often hindered by the difficulty of computing the marginal likelihood of the population model parameters given a sampled genealogy. As a result, approximations to this likelihood (e.g. coalescent likelihoods) are often used.

In this talk, I will discuss some early results from our recent investigations into the use of the particle marginal Metropolis-Hastings (PMMH) algorithm (Andrieu et al., 2010). In our context, this algorithm promises to allow exact phylodynamic inference under a broad class of stochastic population dynamics models by allowing us to avoid explicitly evaluating the marginal likelihood. This class includes standard epidemiological compartmental models such as SI, SIS and SIR, but the method in principle allows for inference under many other (not necessarily pathogen-related) birth-death models; including models with spatial structure.

Arndt von Haeseler
(Max F. Perutz Laboratories, Vienna, arndt.von.haeseler@univie.ac.at)

RNA-seq and the Pitman Sampling formula

How deep is deep enough? While RNA-sequencing represents a well-established technology, the required sequencing depth for detecting all expressed genes is not known. If we leave the entire biological overhead and meta-information behind we are dealing with a classical sampling process. Such sampling processes are well known from population genetics and thoroughly investigated. Here we use the Pitman Sampling Formula to model the sampling process of RNA-sequencing. By doing so we characterize the sam- pling by means of two parameters which grasp the conglomerate of different sequencing technologies, pro- tocols and their associated biases. We differ between two levels of sampling: number of reads per gene and respectively, number of reads starting at each position of a specific gene. The latter approach allows us to evaluate the theoretical expectation of uniform coverage and the performance of sequencing protocols in that respect. Most importantly, given a pilot sequencing experiment we provide an estimate for the size of the

underlying sampling universe and, based on these findings, evaluate an estimator for the number of newly detected genes when sequencing an additional sample of arbitrary size. (Joint work with Stefanie Tauber, Luis F.P. Paz)

Graham Wallis
(Univeristy of Otago, g.wallis@otago.ac.nz)

Nuclear versus mitochondrial trees: discordance is consistent with higher interspecific hybridization in freshwater fishes

Discordance among gene trees can result from hybridization, paralogy, lineage sorting and selection. It is usually possible to exclude paralogy, lineage sorting is unlikely to influence well-separated nodes, and selection is unlikely to result in wholesale convergence. Hybridization, however, is proving to be a widespread feature of plant and animal evolution, and can occur between quite deep lineages. Integration of even low rates of hybridization over many millions of years can lead to extensive discordance among gene trees.

Among animals, freshwater fishes are particularly prone to hybridization, in stark contrast to their marine counterparts. If hybridization were a major cause of discordance among gene trees, we would expect significantly more discordance among freshwater fish gene trees than is seen in marine groups. Using papers published in Molecular Phylogenetics & Evolution since 2003 presenting mitochondrial and nuclear gene trees, we computed the proportion of nodes shared between gene trees for the two different genomes. For 15 marine and 17 freshwater species, node sharing was significantly higher in marine trees, on average by nearly two-fold. We suggest that phylogenies for many freshwater groups may be conflated by hybridization.

Michael Woodhams
(University of Tasmania, michael.woodhams@utas.edu.au)

Parametrization of Lie Markov models

The Lie Markov DNA mutation models are designed to be mathematically consistent when rate matrices change over time. A Lie Markov rate matrix is constructed as a linear combination of basis matrices. Most of these basis matrices include negative off-diagonal elements. For maximum likelihood hill-climbing or for MCMC, we need to be able to propose rate matrices from the model, and require that those rate matrices be stochastic, i.e. no mutation rates are negative. The problem of parametrizing the Lie Markov models in such a way as to simply produce all of the stochastic rate matrices and no non-stochastic rate matrices is not trivial. I will describe an elegant and efficient parametrization.

Yaokun Wu
(Shanghai Jiao Tong University, ykwu@sjtu.edu.cn)

Encoding X-Trees with tree preorder

Let X be a finite set. An X-tree is a tree T together with an injective map ϕ from X to $V(T)$ so that every vertex of degree at most two lies in the image of ϕ. Two X-trees (T, ϕ) and (T', ϕ') are

isomorphic if there is a graph isomorphism f from T to T' such that $\phi' = f \circ \phi$.

The tree preorder corresponding to an X-tree $\mathcal{T}$, denoted by $p_{\mathcal{T}}$, is the unique ordered partition of $\binom{X}{2}$ into the ordered sequence of nonempty sets $L_1, \ldots, L_d$, such that for any positive integers $i, j \leq d$, $\{x, y\} \in L_i$ and $\{u, v\} \in L_j$, it holds $dist_T(\phi(x), \phi(y)) < dist_T(\phi(u), \phi(v))$ if and only if $i < j$.

We show that two X-trees $\mathcal{T}$ and $\mathcal{T}'$ are isomorphic if and only if the tree preorders generated by them are the same, namely $p_{\mathcal{T}} = p_{\mathcal{T}'}$. In general, we discuss how a network structure is encoded by an ordered partition of $\binom{X}{2}$.

This is joint work with Ziqing Xiang, Zeying Xu and Yinfeng Zhu.

5 Participants

Jane Allison	j.allison@massey.ac.nz
Frederic Austerlitz	austerlitz@mnhn.fr
Monika Balvočiūtė	mokana@gmail.com
Arman Bilge	abil933@aucklanduni.ac.nz
Remco Bouckaert	remco@cs.auckland.ac.nz
David Bryant	david.bryant@otago.ac.nz
Minh Bui	minh.bui@univie.ac.at
Michael Charleston	michael.charleston@sydney.edu.au
Paul Cordue	paul.cordue@pg.canterbury.ac.nz
Stephen Crotty	stephen.crotty@adelaide.edu.au
Gulio Dalla Riva	gvd16@uclive.ac.nz
Ben Drinkwater	benjamin.drinkwater@sydney.edu.au
Phillip Endicott	phillip.endicott@gmail.com
Andrew Francis	a.francis@uws.edu.au
Alex Gavruskin	alex@gavruskin.com
Alexandra Gavryushkina	sasha.gavryushkina@auckland.ac.nz
Ole Geldschlaeger	og101912@uni-greifswald.de
Stefan Grünewald	stefan@picb.ac.cn
Joseph Heled	jheled@gmail.com
Mike Hendy	mhendy@maths.otago.ac.nz
Gordon Hiscott	ghiscott@maths.otago.ac.nz
Barbara Holland	Barbara.Holland@utas.edu.au
Georgi Hudjashov	hudja@ebc.ee
Flora Jay	flora.jay@mnhn.fr
Michelle Kendall	m.kendall@imperial.ac.uk
Steffen Klaere	s.klaere@auckland.ac.nz
Denise Kühnert	denise.kuehnert@env.ethz.ch
Rob Lanfear	robert.lanfear@mq.edu.au
Volkmar Liebscher	volkmar.liebscher@uni-greifswald.de
Simone Linz	simone_linz@yahoo.de
Pete Lockhart	p.j.lockhart@massey.ac.nz
Catherine Macken	c.macken@auckland.ac.nz
Ashar Malik	asharjm@gmail.com
Patricio Maturana	p.russel@auckland.ac.nz
Luke Maurits	luke@maurits.id.au
Bennet McComish	bennet.mccomish@utas.edu.au
Angela McGaughran	angela.mcgaughran@csiro.au
Karen Meusemann	karen.meusemann@csiro.au
Huw Ogilvie	huw.ogilvie@anu.edu.au
Maj Padamsee	padamseem@landcareresearch.co.nz
David Penny	d.penny@massey.ac.nz
Ant Poole	anthony.poole@canterbury.ac.nz
Alex Popinga	alexpopinga@gmail.com
Charles Semple	charles.semple@canterbury.ac.nz
Chris Simon	chris.simon@uconn.edu
Mike Steel	mathmomike@gmail.com

Valentin Thouzeau	valentin.thouzeau@agroparistech.fr
Tim Vaughan	tgvaughan@gmail.com
Arndt von Haeseler	arndt.von.haeseler@univie.ac.at
Graham Wallis	g.wallis@otago.ac.nz
Kristina Wicke	kristina.wicke@web.de
Michael Woodhams	michael.woodhams@utas.edu.au
Yaokun Wu	ykwu@sjtu.edu.cn
Walter Xie	walter@cs.auckland.ac.nz

Conference organisers: David Bryant, Mike Hendy, Marguerite Hunter