

Doom 10



**The Annual New Zealand Phylogenetics Meeting
Tuesday 9th - Sunday 14th February 2010**

List of attendees

Alethea Rea	Auckland	NZ	alethea.rea@gmail.com
Alexei Drummond	Auckland	NZ	alexei@cs.auckland.ac.nz
Andreas Dress	Shanghai	China	andreas@picb.ac.cn
Andreea S. Calude	Reading	UK	a.s.calude@reading.ac.uk
Andrew Roger	Halifax	Canada	aroger@dal.ca
Ant Poole	Christchurch	NZ	anthony.poole@canterbury.ac.nz
Barbara Holland	PN	NZ	b.r.holland@massey.ac.nz
Barbara Schoenfeld	PN	NZ	b.schoenfeld@massey.ac.nz
Beata Faller	Christchurch	NZ	fallerbeata@gmail.com
Bennet McComish	PN	NZ	b.mccomish@massey.ac.nz
Brendan Bycroft	Christchurch	NZ	brb44@student.canterbury.ac.nz
Celine Scornavacca	Tuebingen	Germany	Celine.Scornavacca@lirmm.fr
Charles Semple	Christchurch	NZ	c.semple@math.canterbury.ac.nz
Chris Simon	Welly/Storrs	NZ/USA	chris.simon@uconn.edu
David Bryant	Auckland	NZ	d.bryant@auckland.ac.nz
David Penny	PN	NZ	d.penny@massey.ac.nz
Denise Kuhmert	Auckland	NZ	denise.kuehnert@gmail.com
Fabio Pardi	Montpellier	France	pardi@lirmm.fr
Gillian Gibb	PN	NZ	g.c.gibb@massey.ac.nz
Jack Sullivan	Moscow	USA	jacks@uidaho.edu
James Degnan	Christchurch	NZ	jamdeg@gmail.com
Jessica W. Leigh	Paris/Auckland	France/NZ	jessica.w.leigh@gmail.com
Joseph Heled	Auckland	NZ	jheled@gmail.com
Josh Collins	Christchurch	NZ	j.collins@math.canterbury.ac.nz
Klaus Schliep	Paris	France	klaus.schliep@gmail.com
Lakoa Fitina	Madang	PNG	lfitina@dwu.ac.pg
Leo van Iersel	Christchurch	NZ	l.j.j.v.iersel@gmail.com
Lesley Collins	PN	NZ	l.j.collins@massey.ac.nz
Marc Suchard	Los Angeles	USA	msuchard@ucla.edu
Mareike Fischer	Vienna	Austria	email@mareikefischer.de
Michael Charleston	Sydney	Australia	mcharles@it.usyd.edu.au
Michael DeGiorgio	Ann Arbor	USA	degiorimi@umich.edu
Michael Woodhams	Sydney	Australia	woodhams@it.usyd.edu.au
Mike Hendy	PN	NZ	m.hendy@massey.ac.nz
Mike Steel	Christchurch	NZ	mathmomike@gmail.com
Olivier Gascuel	Montpellier	France	gascuel@lirmm.fr
Peter Waddell	West Lafayette	USA	pwaddell@purdue.edu
Peter Lockhart	PN	NZ	p.j.lockhart@massey.ac.nz
Raazesh Sainudiin	Christchurch	NZ	r.sainudiin@math.canterbury.ac.nz
Remco Bouckaert	Auckland	NZ	remco@cs.auckland.ac.nz
Rob Lanfear	Canberra	Australia	rob.lanfear@gmail.com
Simon Greenhill	Auckland	NZ	s.greenhill@auckland.ac.nz
Simon Hills	PN	NZ	s.f.hills@massey.ac.nz
Stefan Grünewald	Shanghai	China	stefan@picb.ac.cn
Steffen Klaere	Auckland	NZ	steffen.klaere@gmail.com
Tal Dagan	Dusseldorf	Germany	tal.dagan@uni-duesseldorf.de

(dis)Organisers

Joy Wood	j.r.wood@massey.ac.nz
Barbara Holland	b.r.holland@massey.ac.nz (mobile # 021 0375887)
Pete Lockhart	p.j.lockhart@massey.ac.nz

Programme:

Tuesday 9th Feb

17:00 – 20:00	Registration and reception (Skotel bar)
----------------------	--

Wednesday 10th Feb

08:45 – 09:15	Registration
09:15 – 09:20	Greetings and opening
09.20 – 10:00	[Starters] Chair: Barbara Holland 09:20–09:40 Alexei Drummond, Biogeoinformatics 9:40–10:00 Mike Steel, Can we avoid 'SIN' in the 'House of No Common Mechanism'?
10:00 – 10.40	Morning break
10:40 – 12:00	[Networks] Chair: Jack Sullivan 10:40–11:00 Andrew Roger, Is a 'vertical' signal in deep prokaryote evolution recoverable given the 'horizontal' noise? Filtering discarded bathwater for babies. 11:00–11:20 Tal Dagan, A directed network of recent lateral gene transfers in prokaryotes reveals barriers for transfer between different taxonomic groups 11:20–11:40 Brendan Bycroft, Mutational Markov Chains on Ancestral Recombination Graphs 11:40–12:00 Josh Collins, A quick and dirty algorithm to calculate the hybridisation number
12:00 – 14:00	Lunch
14:00 – 15:00	[Phylogeny] Chair: Michael Woodhams 14:00–14:20 Mike Charleston, Resolving the origin and evolution of the major animal phyla 14:20–14:40 Gillian Gibb, New Zealand passerines help resolve basal oscine phylogeny 14:40–15:00 Chris Simon, Progress in the Systematics of NZ, Australian and World cicadas.
15:00 – 15.30	Afternoon break
15:30 – 16.30	[Likelihood & RNA] Chair: David Penny 15:30–15:50 Marc Suchard, Inferring likelihood-based phylogenies 100-fold faster 15:50–16:10 Bennet McComish, Multiple local maxima for

	likelihoods of phylogenetic trees constructed from biological data 16:10-16:30 Lesley Collins, ncRNA systems biology: fitting RNA genes into a protein-centric network
17:30	Dinner – your own arrangements

Thursday 11th Feb

07:00 – 09:00	Breakfast at the Chateau
09:15 – 09:20	Greetings and opening
09.20 – 10:00	[Migration] <i>Chair: Alexei Drummond</i> 09:20–09:40 Steffen Klaere, Mean ratio of segregating sites for sets of lineages in a migration model 9:40–10:00 Joseph Heled, Radiation with Migration: The effects of migration on species tree estimation.
10:00 – 10.40	Morning break
10:40 – 12:00	[Lots of Trees] <i>Chair: Mike Steel</i> 10:40–11:00 Jessica Leigh, Congruence in Massive Messy Multigene Datasets 11:00–11:20 Remco Bouckaert, DensiTree: making sense of sets of trees 11:20–11:40 Celine Scornavacca, Methods to include multi-labeled phylogenies in a supertree framework 11:40–12:00 Mareike Fischer, Using groves to find trees – a new supertree method and its drawbacks
12:00 – 14:00	Lunch
14:00 – 15:00	[PN escapees] <i>Chair: Charles Semple</i> 14:00–14:20 Pete Waddell, Fast Fitchstein and Better Things than the Bootstrap 14:20–14:40 Klaus Schliep, Catching networks with the LASSO 14:40–15:00 Michael Woodhams, The Mosaic Model: modelling changes in evolutionary process across deep phylogenies.
15:00 – 15.30	Afternoon break
15:30 – 16.30	[Bits and Bobs] <i>Chair: James Degnan</i> 15:30–15:50 Rob Lanfear, The relationship between mitochondrial protein structure and rates of molecular evolution 15:50–16:10 Charles Semple, New results on the subtree distance between phylogenies

	16:10-16:30 David Bryant, Testing Mr Bayes using Rorschach inkblots
17:30	Dinner – your own arrangements

Friday 12th Feb

All day	Excursions
18:30 – 20:00	BBQ Dinner

Saturday 13th Feb

09:15 – 09:20	Greetings and opening
09:20 – 10:20	<p>[Gene Trees and Species Trees] <i>Chair: David Bryant</i></p> <p>09:20–09:40 James Degnan, Distribution of tree balance for gene trees in species trees</p> <p>9:40–10:00 Jack Sullivan, Divergence with gene flow in chipmunks</p> <p>10:00–10:20 Michael DeGiorgio, Fast and consistent estimation of species trees using supermatrix rooted triples</p>
10:20 – 11:00	Morning break
11:00 – 12:00	<p>[Languages & Viruses] <i>Chair: Chris Simon</i></p> <p>11:00–11:20 Andreea Calude, Using phylogenetics methods to uncover deep linguistic relationships</p> <p>11:20–11:40 Simon Greenhill, The Evolution of Language Structure</p> <p>11:40–12:00 Olivier Gascuel, Phylotyping Pathogenic Viruses</p>
12:00 – 14:00	Lunch
14:00 – 15:00	<p>[Networks II] <i>Chair: Olivier Gascuel</i></p> <p>14:00–14:20 Raazesh Sainudiin, Lumpings of coalescents with multiple mergers</p> <p>14:20–14:40 Stefan Grunewald, A Quartet Version of Split Decomposition</p> <p>14:40–15:00 Leo van Iersel, Treewidth and the Genome of Eden</p>
15:00 – 15.30	Afternoon break
15:30 – 16.30	<p>[Bits and Bobs] <i>Chair: Barbara Holland</i></p> <p>15:30–15:50 Fabio Pardi, Robustness of phylogenetic inference based on minimum evolution</p> <p>15:50–16:10 Beata Faller, Using linear programming to optimize</p>

	<p>phylogenetic diversity with ecological constraints</p> <p>16:10-16:30 Simon Hills, Can uneven extinction lead to basal branch contraction in molecular clock analysis?</p>
17:30	Dinner – your own arrangements

Sunday 14th (Valentine's Day):

08:55 – 09:00	Greeting and notices
09:00 – 10:00	<p>[Mystery session] <i>Chair: Peter Waddell</i></p> <p>09:00–09:20 Alethea Rea, Confidence sets for tree topology</p> <p>9:20–9:40 Ant Poole, A comparative genomic test of the introns-first theory for the origin of spliceosomal introns</p> <p>9:40–10:00 David Bryant, Sneak preview of next year's meeting</p>
10:00 – 10:30	Morning break
10:30 – 11:30	<p>[Big Ideas] <i>Chair: Barbara Holland</i></p> <p>10:30–11:00 Andreas Dress, Species, Genealogies, and the 'Tree of Life', from a graph-theoretical point of view.</p> <p>11:00–11:30 David Penny, The continuity of mind</p>
11:50	Closing comments, lunch

Abstracts:

Remco Bouckaert

Title: DensiTree: Making Sense of Sets of Trees

Abstract: Bayesian analysis through programs like BEAST and MrBayes provides a powerful method for reconstruction of evolutionary relationships. One of the benefits over many other methods is that well founded estimates of uncertainty in models can be made available so that for example not only the mean time of a most recent common ancestor is estimated, but also the spread. This distribution over model space is represented by a set of trees, which can be rather large and difficult to interpret. DensiTree is a tool that helps navigating these sets of trees.

The main idea behind DensiTree is to draw all trees in the set transparently. As a result, areas where a lot of the trees agree in topology and branch lengths show up as highly colored areas, while areas with little agreement show up as webs. This makes it possible to quickly get an impression of properties of the tree set such as dominant clades, distribution of tMRCA and areas of topological uncertainty. Thus, DensiTree provides a quick method for qualitative analysis of tree sets.

DensiTree is freely available from

<http://compevol.auckland.ac.nz/software/DensiTree/> (stable version) and
<http://www.cs.auckland.ac.nz/~remco/DensiTree> (developer version).

David Bryant

Title: Testing Mr Bayes using Rorschach inkblots

Abstract: Theory tells us that, if run for long enough, MCMC software like MrBayes and Beast will return trees sampled from the posterior distribution. Unfortunately it is often very difficult to determine whether your MCMC chains are long enough, even from synthetic data. Mossel and Vigoda have shown that, for a fairly simple posterior distribution, even gazzillion iterations might not be sufficient. Which all begs the question: are Bayesian phylogeneticists being prematurely optimistic?

In this talk I describe a cheap and dirty experiment for assessing convergence in MCMC phylogenetic software. Given an arbitrary alignment, I mangle it so as to produce a symmetry in the posterior distribution that can be tested statistically.

Brendan Bycroft

Title: Mutational Markov Chains on Ancestral Recombination Graphs

Abstract: We present a C++ program for a fully customisable simulation of repetitive and non-repetitive DNA on structured ancestral recombination graphs. This allows for species-specific and locus-specific mutation models to produce the sampling distributions of most population-genetic statistics. The talk will focus on the capabilities of the program and a user-friendly tour.

Joint-work with:

Raazesh Sainudiin and Kevin Thornton

Andreea S. Calude**Title:** Predicting Deep Linguistic Relationships

Abstract: Like genes, words retain signals of their ancestry dating back millennia, but how far back in time can they persist and are they likely to show deep conservation which may be identified in advance? Here we use recent results suggesting time depths of 20,000 or more years are plausible, to show that words carry far greater historical signal than previously believed, and that highly conserved words can be predicted independently of their particular form. The frequencies with which words are used in everyday speech can predict the words most likely to have been shared among the proto-forms of seven Eurasian language families thought to have evolved from a common ancestor around 15,000 years ago (Bomhard and Kerns, 1994). We use these results to derive phylogenetic tree of the Eurasian language families, indicating 'genetic' relationships in this group, and an estimated time depth consistent with a 15kya origin. The tree implies that some highly used words such as who, I and thou have remained unchanged since at least the last ice age. Our results point to a remarkable fidelity in the transmission of some highly-used lexical elements, provide a predictive framework to underpin efforts to establish deep relationships among languages, and give a theoretical justification to the search for words left over from the human Mother Tongue.

Joint-work with: Quentin D. Atkinson, Andrew D. Meade, Mark Pagel

Mike Charleston**Title:** Resolving the origin and evolution of the major animal phyla

Abstract: In dealing with large heterogeneous data sets for phylogenetic inference there are many challenges, and there have been many approaches to overcome them.

We have developed new methods with a philosophy of practicality and necessary compromise. Since molecular data are generated by heterogeneous mechanisms, we should accommodate them with heterogeneous models, yet more complex models are more susceptible to sampling error. While small numbers of taxa have smaller tree space, larger ones permit a better global picture.

Herein we discuss some of the developments and results we have gathered, in particular, the effects of model estimation and data concatenation on quality of phylogenetic signal, the enduring problem of multiple sequence alignment and why Clustal Looks Great (until you look closely), and the removal of "problem" taxa by strangulation.

Josh Collins:**Title:** A quick and dirty algorithm to calculate the hybridisation number

Abstract: The interleaving algorithm described in a submitted paper by Linz and Semple depends on a technique that can be used to generate a very small program that will calculate the maximum acyclic agreement forest of a set of trees. This talk will elucidate this method and talk about some of the results that 'fall out' of this formulation and compare it with the times required via other similar methods to solve real world data.

Lesley Collins:

Title: ncRNA systems biology: fitting RNA genes into a protein-centric network

Abstract: ncRNAs do not translate into proteins but they do interact with them in large cellular networks. However, ncRNA genes are largely missing from gene interaction networks and essential information is lost when they are fitted into protein-centric models. This talk will go over some of the issues involved in networking ncRNAs both at an interaction and regulatory level, and how we can begin to work around these issues to model RNP complexes.

Tal Dagan

Title: A directed network of lateral gene transfer reveals barriers and bypaths to gene acquisition in prokaryotes.

Abstract: Lateral gene transfer is an important mechanism of natural variation among prokaryotes, but the extent of genomic exchange among different species and possible barriers to it are still debated. Here we report the use of phylogenetic networks to capture both vertical inheritance and recent lateral gene transfer among 657 prokaryotes. Recently acquired genes whose nucleotide content has not yet reached the characteristic genomic GC content are identified by statistical methods. Among the 2,129,548 genes in our dataset, 446,854 genes are identified as recent acquisition. This corresponds to an average of $20 \pm 9\%$ recently acquired genes in each genome, with similar proportions observed across different taxonomic groups. For 9% of the acquired genes a candidate donor gene was identified among the available genomes by a combination of statistical inference and phylogenetic heuristics. For the remaining acquired genes no candidate donor could be inferred on the basis of the current genome sample. All lateral gene transfer events are then superimposed as a lateral network upon a reference species tree. By comparison of orthologous acquired genes within monophyletic groups we are able to reconstruct transfer events to ancestral nodes of the phylogenetic tree. Most of the detectable gene transfers occur between species within the same taxonomical groups. Moreover, species having similar genomes are more frequently connected than species having dissimilar genomes. This suggests that donor-recipient dissimilarity is a barrier for lateral gene transfer. However, species having the proteins required for non-homologous end-joining (NHEJ) are connected more frequently with dissimilar donors than species lacking that mechanism. This suggests that NHEJ has a possible role in gene acquisition in prokaryotes.

Joint with Ovidiu Popa, Einat Hazkani-Covo, Giddy Landan, William Martin

James Degnan

Title: Distribution of tree balance for gene trees in species trees

Abstract: Studies of tree shape have attempted to find appropriate null models for phylogenetic trees, where these trees are often interpreted as trees of species, yet phylogenetic trees are typically inferred using aligned sequences from genes (gene trees). We find that there is some effect from the multispecies coalescent on the distribution of tree shape for gene trees within a species tree when the species tree is generated under a Yule model, and that an appropriate null distribution for gene tree

shape (averaged over species trees) is generally not the same as the null distribution for the shape of the species tree. In particular, there can be some increased probability that the gene tree is less balanced than is predicted under the Yule model.

Andreas Dress

Title: Species, Genealogies, and the 'Tree of Life', from a graph-theoretical point of view.

Abstract: The binary "parent-child" relation defines a very natural directed graph structure on the set V of all organisms that ever lived on earth. I will discuss various natural options for defining "species", "genera", "families" and so on as subsets of V based on this di-graph so that these subsets form a nice "hierarchy", i.e., a system of non-overlapping sets, this way trying to address the simple, yet fundamental question "What really is a species?"

Alexei Drummond:

Title: Bioinformatics

Abstract: Bioinformatics is the organization and analysis of biological data; especially biomolecular sequence data such as that deposited in the public database Genbank. Increasingly life scientists want to store a broader range of data types alongside this sequence data and incorporate these new sources of data into rigorous model-based statistical analyses. This is especially the case in the fields of molecular ecology and evolution. In this talk I will highlight two such datatypes: geographical location of samples, and geo-located/mapped ecological niche data. I will outline the types of analyses that combine these data types with sequence data in a range of organisms, including some work from our research group. Finally I will make the case for the development of a bioinformatic research infrastructure, that expands on Genbank, and provides the necessary impetus for further developments in this burgeoning area of science.

Beáta Faller

Title: Using linear programming to optimize phylogenetic diversity with ecological constraints

Abstract: When conservation decisions require prioritizing some species over others, one solution is to select the set of species with maximum phylogenetic diversity (PD). Introduced by Faith in 1992, the PD of a set of species is the sum of the branch lengths of the phylogenetic tree spanned by the species. An optimal subset of a set of species of a given size that maximizes PD can be obtained by a greedy algorithm. However, when species depend on other species, such as in a food web (a map of who eats who in an ecological community), maximizing PD alone may not lead to a viable subset of species being protected. Therefore, we require that the set of selected species is viable (every non-source species has some prey and is connected to a source). This additional constraint makes the problem computationally hard, except in the unrealistic case of a star phylogeny and a tree-like food web. To find exact solutions to realistically large problems, we formulate the optimization problem as an integer linear program. We apply this technique to a 249-node empirical food web and are able to find subsets of any size that maximize PD while preserving a viable food web.

Joint work with Travis Ingram and Charles Semple

Mareike Fischer

Title: Using groves to find trees – a new supertree method and its drawbacks

Abstract: A major challenge in biological sciences is the reconstruction of the Tree of Life. To this effect, large genomic databases are being mined for clusters from which phylogenies can be inferred. Systematists and comparative biologists commonly combine such phylogenies into informative supertrees that reveal information which was not explicitly displayed in any of the original phylogenies. However, whether a supertree is informative depends on particular overlap properties among the clusters from which it originates. In this context, Ané et al. formally introduced the concept of phylogenetic groves (Ané et al.: “Groves of Phylogenetic Trees”, *Annals of Combinatorics*, DOI 10.1007/s00026-009-0017-x), which are sets of clusters with the potential to construct informative supertrees. They stated that maximal potential candidate clusters for informative supertree construction can be identified in large databases through groves, prior to inferring trees for each cluster. However, they left one crucial question open: if two groves intersect, is their union a grove? The authors conjectured that this was true. In fact, one would need this property in order to identify all maximal groves in a database. Unfortunately, using the definition of groves introduced by the authors, it is comparatively easy to construct counterexamples. I will present this construction in my talk. Moreover, I will introduce the concept of strictly informative groves for which the required property does hold.

Oliver Gascuel

Title: Phylotyping Pathogenic Viruses

Abstract: We present simple ideas, criteria and algorithms to identify pathogenic virus phlotypes, that is, clusters of virus strains with common ancestry and epidemiological importance. This latter is assessed using user-supplied extrinsic factors, such as geographic location, population group, or risk behaviour. Given a phylogeny of the strains, phlotypes are obtained using a tree searching algorithm. This combines character mapping to account for the extrinsic factors, and statistical criteria that express the genetic separation and homogeneity of the clusters. Fast linear computing times are involved, which enable the exploratory analysis of very large phylogenies. Appropriate extrinsic factors and cluster selection criteria are combined and adjusted until a clear picture is obtained from the data. We illustrate this approach with HIV data sets, where complex transmission chains are identified from one country to another or among population groups. A web server is available to use this tool online.

Joint work with F. Chevenet, A.-M. Arigon, H. Lehvaslaiho, T. de Oliveira

Gillian Gibb

Title: New Zealand passerines help resolve basal oscine phylogeny

Abstract: Despite the increasing amount of sequence information and improving phylogenetic methods, the basal nodes of Passeriformes, the largest avian radiation, are still poorly resolved. The basal phylogeny of oscine passerines is important for

interpreting the radiation of this group out of the Australasian region. We report five new complete mitochondrial genomes from basal oscine lineages (tui, saddleback, hihi, grey warbler and NZ robin). Our well-resolved phylogeny is in agreement with recent analyses of certain nuclear exons and introns, but partly disagrees with phylogenies based on Rag-1, one of the most commonly used nuclear exons in passerine phylogenetics. We find several so-called 'basal corvid' taxa are more likely basal to Passerida. This study also reports 1.8 Kb of mitochondrial DNA from the extinct New Zealand piopio. We identify that the previously published piopio *cytb* is incorrect, and show the piopio is not a basal bowerbird, but falls within the core Corvoidea.

Stefan Grünewald

Title: A Quartet Version of Split Decomposition

Abstract: Split decomposition is one of the first and most widely used methods to reconstruct not necessarily compatible weighted split systems which can be visualized as undirected phylogenetic networks (split networks). Its input is a dissimilarity function (usually a metric) but it can also be considered as a quartet-based method where, for all 4 taxa, the weight of at least one of the three possible quartets is zero. I will introduce some work in progress where we develop a variant of split decomposition such that the weights of all quartets (which might be computed directly from the raw data) can be positive. The method can reconstruct more general than weakly compatible split systems. Further, a systematic bias towards too long quartets can be recognized from the output while too long distances will cause long pending edges in the split decomposition network. Some first experiments show that the output of the new method depends heavily on how the quartet weights are computed, and it tends to be significantly different from the (distance-based) split decomposition and other methods like NeighborNet and QNet.

Joseph Heled

Title: Radiation with Migration: The effects of migration on species tree estimation.

Abstract: Several methods for inferring multispecies phylogenies are available today, and surely more will come in the future (BEST, *BEAST, etc). Those models typically assume that speciation occurs at a specific point in time, after which the two lineages evolve in total isolation. However, speciation may occur over a period of time where sister lineages remain in partial contact. Inference of multispecies phylogenies under such conditions is not available yet, but it is relatively easy to simulate a gene history while including the effects of possible contact. In my talk I will present the theory and some results of the effects of the model mismatch between speciation with migration and *BEAST strict speciation model.

Simon Hills

Title: Can uneven extinction lead to basal branch contraction in molecular clock analysis?

Abstract: Molecular clock analysis of the New Zealand marine snail genus *Alcithoe* have inferred tree-root dates that are younger than currently accepted taxonomy

suggests. This discrepancy may be the result of contraction of a long basal branch in the phylogeny. In *Alcithoe* this long basal branch is likely the result of uneven extinction processes across the lineage. If this is the case, including calibrated nodes closer to the root should recover more accurate tree-root age estimates. However, all known divergent lineages between the root of the tree and the oldest divergence in the modern taxa are extinct and cannot be sampled for molecular analysis. A sequence simulation approach was implemented to examine the effect of extinction processes on the inference of node ages in molecular clock analysis. These analyses utilised molecular and paleontological data to inform model parameters.

Steffen Klaere

Title: Mean ratio of segregating sites for sets of lineages in a migration model

Abstract: We regard a model of trait evolution which is governed by mutation (single lineage-single site event) and migration (pairs of lineages-single site event). Using this model we infer the mean and variance of the ratio of segregating sites for a set of lineages. In addition we present a way to infer the covariances between sets of lineages.

These values provide the possibility to investigate the changes of evolutionary distances between lineages under changing migration rates.

Rob Lanfear:

Title: The relationship between mitochondrial protein structure and rates of molecular evolution

Abstract: A number of recent studies have suggested that the structure of mitochondrial membrane proteins directly determines the rate of damage to mitochondrial DNA. Specifically, some amino acids have been suggested to increase DNA damage rates, while others have been suggested to reduce them. Here, we examine this hypothesis by testing for links between the structure of mitochondrial membrane proteins and the neutral mutation rate of mitochondrial DNA. We use a large dataset of complete mitochondrial genomes for 180 mammals and 3D homology modelling to estimate the fine-scale structural composition of the mitochondrial membrane proteins, and test whether these structural characteristics are related to the underlying mutation rate.

Jessica Leigh

Title: Congruence in Massive Messy Multigene Datasets

Abstract: A variety of congruence assessment methods have been proposed for multi-locus phylogenetic analysis over the years. These have ranged from Farris' parsimony based incongruence length difference test and Huelsenbeck and Bull's likelihood ratio test to adaptations of likelihood-based topology tests and statistical methods such as principal component analysis. For various reasons, existing congruence methods are particularly ill-suited to datasets with: a) a very large number of genes; b) a high level of incongruence; and c) different taxon distributions among markers. These characteristics apply to most truly phylogenomic datasets, particularly those composed of prokaryote or virus sequences. These genomes encode the vast majority of available sequence data, not to mention the majority of as-yet unsequenced

genomic data on Earth. We have developed a new method for assessing congruence that uses clustering of markers based on similarity of bipartition posterior distributions. Our method scales nearly linearly with the number of markers, accommodates high levels of incongruence, and offers different methods for treating missing taxa. Results with simulated data show that our congruence-based clustering method performs very well across a broad range of conditions.

Bennet McComish

Title: Multiple optima of likelihood on trees from real data

Abstract: It is known that the maximum likelihood function can have multiple local optima on a given tree. Some simulation studies suggest that this is not likely to affect tree-building, but these simulations used data generated on a single tree. In contrast, it has been shown that simple mixture models can generate data where multiple optima can occur even on the tree with the highest likelihood. I will present some preliminary results on how often multiple optima occur with real biological sequence data.

Fabio Pardi

Title: Robustness of phylogenetic inference based on minimum evolution

Abstract: Minimum evolution is the guiding principle of an important class of distance-based phylogeny reconstruction methods, including neighbor-joining (NJ), which is the most cited tree inference algorithm to date. The minimum evolution principle involves searching for the tree with minimum length, where the length is estimated using various least-squares criteria. Since evolutionary distances cannot be known precisely but only estimated, it is important to investigate the robustness of phylogenetic reconstruction to imprecise estimates for these distances. The safety radius is a measure of this robustness: it consists of the maximum relative deviation that the input distances can have from the correct distances, without compromising the reconstruction of the correct tree structure. Answering some open questions, I will show the safety radius of two popular minimum evolution criteria: balanced minimum evolution (BME) and minimum evolution based on ordinary least squares (OLS+ME). Whereas BME has a radius of 1/2, which is the best achievable, OLS+ME has a radius tending to 0 as the number of taxa increases. This difference may explain the gap in reconstruction accuracy observed in practice between OLS+ME and BME (which forms the basis of popular programs such as NJ and FastME).

Joint with Sylvain Guillemot and Olivier Gascuel

David Penny

Title: The continuity of mind, from great apes to humans

Abstract: It is a central prediction from mainstream evolutionary biology that there is a continuous set of intermediate states in the mental abilities from an ape-human ancestor to modern humans. A quick analysis is given to the problem introduced by Rene Descartes of the supposed gap between mind matter and physical matter, the former distinguishing humans from all other animals. A three-pronged analysis is then given that revolves around 1) the continuity of mind from young children to adults, 2) the similarity of measured mental abilities between young great apes and young children at similar stages of development, and 3) the observation that there appear to

be no unique genes in the human genome for mental abilities (including for ‘wisdom and intelligence’). The components that are required for language are then analysed, and important precursors found in the great apes. The conclusion is that the best supported hypothesis, especially in a Bayesian framework, is for a continuum in mental states between an ancestral ape and modern humans. There are a range of predictions from this model that can, and have been, tested.

Alethea Rea

Title: Confidence set for tree topology

Abstract: Often phylogenetic trees are reported with their bootstrap probabilities. A branch with a high bootstrap probability indicates that given the sequence and evolutionary model, support is high for that branch being in the true tree. A branch with low bootstrap support indicates that at least one taxon could be placed on the other side of the tree. Bootstrap probabilities alone give no indication as to which taxon or taxa are causing the uncertainty and what alternatives are plausible given the data. The approach we take is to use a candidate set of splits and the tools of the approximately unbiased test of Shimodaira 2002 to get approximately unbiased p-values relating to the hypothesis that the true gene tree contains that split. The set of splits not rejected can be used to construct a confidence set of trees; an exhaustive set trees that can be formed given those splits. Examination of this set can assist in determining where in the tree uncertainty lies. We demonstrate the method on simulated and real data.

Raazesh Sainudiin

Title: Lumpings of coalescents with multiple mergers

Abstract: We present some results on Markov lumpings of coalescents with multiple mergers at six coalescent resolutions. This work extends the Markov lumpings of the Kingman-Tajima coalescent with binary mergers and leads to a more general unified multi-resolution coalescent.

Joint work with Amandine Véber

Klaus Schliep

Title: Catching networks with the LASSO

Abstract: We will describe an algorithm for estimating phylogenetic network from a matrix of pairwise distances. Finding the optimal tree using Least-Squares (LS) is known to be NP-hard (Day 1986) and it is related to the problem of finding the optimal subset in regression analysis. For high dimensional problems a penalized Least-Squares (LS) using a LASSO penalty (Tibshirani 1996, Osborne et al. 2000) offers the advantage that many variables (in our case splits) will be set to zero, i.e. a variable selection is automatically performed. The Least Angle Regression (LARS) algorithm (Efron et al. 2004) gained popularity as it provides a fast implementation for the whole path of LASSO solutions up to a specified number of variables (splits). We adopted the LARS algorithm to fit a phylogenetic network from distance data given a set of splits with non-negative LARS or equivalent penalized (weighted) non-negative Least-Squares (suggestions for an acronym welcome). We can derive the distance Hadamard (Hendy and Penny 1993) is a special case of the LARS

algorithm. Therefore we can estimate a phylogenetic network which nicely minimizes a well-defined penalized Least-Squares criterion. This allows to choose the number of edges with goodness-of-fit criterion or to explore data-sets visually.

We will compare phylogenetic networks built with this new approach using splits computed from a bootstrap sample and from a NeighborNet for different species of darters and shags.

Joint with Barbara Holland, Mike Hendy, Martyn Kennedy

Celine Scornavacca

Title: Methods to include multi-labeled phylogenies in a supertree framework

Abstract: Gene trees are leaf-labeled trees inferred from molecular sequences. Due to duplication events arising in genome evolution, gene trees usually have multiple copies of some labels, i.e., species. Inferring a species tree from a set of multi-labeled gene trees (MUL trees) is a well-known problem in computational biology. We propose a novel approach to tackle this problem, mainly to transform a collection of MUL trees into a collection of evolutionary trees, each containing single copies of labels. To that aim, we provide several algorithmic building stones and describe how they fit within a general species tree inference process. Most algorithms have a linear-time complexity, except for an FPT algorithm proposed for a problem that we show to be intractable.

Charles Semple

Title: New results on the subtree distance between phylogenies

Abstract: The subtree prune and regraft (SPR) distance is commonly used in phylogenetics for comparing two phylogenies. At the previous phylogenetics meeting held at Whakapapa, there were at least five talks centred around this distance---from the purely theoretical to ones describing applications in the study of reticulation. In this talk, we survey some new results and implementations of this distance and its variants.

Chris Simon

Title: Progress in the Systematics of NZ, Australian, and World Cicadas

Abstract: Cicadas are excellent organisms for studying biogeographic and phylogeographic patterns due to their unusual life history, widespread distribution, and acoustic sexual signals. Long-lived fossorial juveniles and short-lived above-ground adults lead to low dispersal rates, high levels of phylogeographic structure within species and clear differentiation among species. Low dispersal also enhances the value of geological events as calibrations for dating evolutionary trees. Cicadas' drought tolerance has equipped them to persist through challenging environmental shifts. Cicadas occupy a broad range of habitats and are distributed on all continents except Antarctica. They are known from a few Mesozoic and ~35 Cenozoic fossils. Songs of cicadas are highly species-specific, often diverge first in evolution, and can reveal cryptic species. Cicada songs also facilitate specimen capture and rapid gathering of distributional data. New Zealand has been a particularly fertile area for our work because NZ cicadas are found in nearly every habitat type and represent a

large radiation of species from just one colonizing ancestor (plus a much smaller radiation from a second independent lineage). For the past 17 years, my laboratory and our collaborators have been studying various aspects of NZ cicada evolution from populations to genera. We have sampled every species from most parts of their respective ranges and have constructed phylogenetic trees for five genera. For all the species we have studied geographic variation and for many of them we have conducted detailed phylogeographic studies. Moving out of NZ we have traced the relatives of NZ cicada to Australia and New Caledonia and studied the worldwide systematics and biogeography of the tribe Cicadettini to which NZ cicadas belong. Focusing on four areas in which we have made progress in the last year, I will present new findings from my group on the Cenozoic biogeography of all 38 tribes in the family Cicadidae, the diversification of Australian cicadettine genus *Pauropsalta*, the largest genus of this tribe in Australia, and finally returning to the population level, present preliminary evidence of differential mitochondrial and nuclear gene flow across contact zones in one species of New Zealand cicada.

Mike Steel

Title: Can we avoid 'SIN' in the 'House of No Common Mechanism'?

Abstract: In 'no common mechanism' (NCM) models of character evolution, each character can evolve on a phylogenetic tree under a partially or totally separate process (e.g. with its own branch lengths). In such cases, the usual conditions that suffice to establish the statistical consistency of tree reconstruction by methods such as maximum likelihood (ML) break down, suggesting that such methods may be prone to statistical inconsistency (SIN). In this paper we ask whether we can avoid SIN for tree topology reconstruction when adopting such models, either by using ML or any other method that could be devised. We prove that it is possible to avoid SIN for certain NCM models, but not for others, and the results depend delicately on the tree reconstruction method employed. We also describe the biological relevance of some recent mathematical results for the more usual 'common mechanism' setting. Our results are not intended to justify NCM, rather to set in place a framework within which such questions can be formally addressed.

Jack Sullivan

Title: Divergence with gene flow in chipmunk

Abstract: The study of rapid radiations is of central importance to evolutionary biology because much of the diversity of life has been generated in great bursts of speciation. However, the causes and consequences of rapid radiations are major unresolved issues in evolutionary biology, in part because the phylogeny and processes such as stochastic lineage sorting and hybridization are confounded. Because these processes are expected to be heterogeneous across the genome, comparison among marker classes may provide a means of disentangling these elements. Here we use nuclear-encoded reproductive protein genes that may be resistant to introgression to estimate the phylogeny of the western chipmunks (*Tamias*: subgenus: *Neotamias*), a rapid radiation that has experienced introgressive hybridization of mitochondrial DNA.

Leo van Iersel**Title:** Treewidth and the Genome of Eden

Abstract: Lateral gene transfer (LGT) refers to biological processes in which genetic material is transferred from one species to another. The extent of lateral gene transfer in the history of life is controversial. Molecular biologists have recently argued that, if one assumes that no LGT has occurred, some ancestral species must have had far more genes than their current-day descendants. Such an all-encompassing ancestral genome, referred to as a 'genome of Eden', is considered unlikely. Using similar arguments, the pattern of presence and absence of genes across currently-living species provides clues as to how many LGT events have occurred in their evolutionary history. I will show how this phenomenon is related to the mathematical concepts of treewidth and network flow and how this relationship can be used to formally bound the number of LGT events in an evolutionary history.