Waiheke 2014

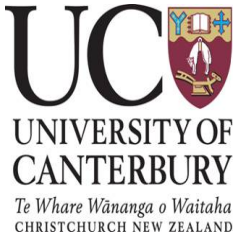
New Zealand Phylogenetics

-Onetangi beach-
2-7 February 2014

18th Annual New Zealand Phylogenomics Meeting
Waiheke Island 2014

2–7 February 2014

Many thanks to our "sponsors"



Organizers: Stéphane Guindon, Alexandra Miliotis, David Welch, Abhinav Chopra, Dietrich Radel & Steffen Klaere.

Contents

1 Programme

Sunday 2 February

Registration from 5:00 pm

Monday 3 February

Registration from 8:45 am

Networks I

9:10 Welcome and announcements

9:20 Charles Semple *Counting phylogenetic networks*

9:40 Fabio Pardi *Reconstructible phylogenetic networks: no need to distinguish the indistinguishable*

10:00 Monika Balvočiūtė *Computation of Planar Split Networks*

Networks II

10:50 David Welch *Phylogenetic transmission trees constrained to networks*

11:10 Paul Cordue *Characterising when a phylogenetic network displays a tree twice*

11:30 Simone Linz *Surprising properties of maximum parsimony on networks*

11:50 Giulio Valentino Dalla Riva *The web traits and the tree*

Actual phylogenies

14:00 Mait Metspalu (short talk) *Database of over 400 complete human genomes from Eurasia and beyond*

14:10 Bojian Zhong *Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes*

14:30 Gillian Gibb *Why fly when you can walk? Flightlessness in modern and extinct NZ rails*

14:50 Simon F. K Hills *Diversification patterns in molecular and paleontological data of the marine mollusc genus Alcithoe*

Not quite trees

15:40 Pete Lockhart *Hybridization may facilitate in situ survival of endemic species through periods of climate change*

16:00 Celine Scornavacca *A practical approximation algorithm for solving massive instances of hybridization number for binary and nonbinary trees*

16:20 Catherine Matias *Co-phylogeny reconstruction via an approximate Bayesian computation*

16:40 David Bryant *Getting excited about diversities*

Tuesday 4 February

Using Fossils

9:20 Michael Gemmell (short talk) *Conundrums caused by conflicting shapes and sequences: When mtDNA and morphometrics disagree*

9:30 Felix Vaux *Whelk phylogenetics as a test of punctuated evolution*

9:50 Tanja Stadler *Inferring phylogenies with fossils*

10:10 Alexandra Gavryushkina *Bayesian MCMC for Sampled Ancestor Trees: Application to Fossil Calibration*

Epidemics and population dynamics I

11:00 David Duchene (short talk) *Phylogenetic timescale estimation when molecular evolution and diversification rates are linked*

11:10 Denise Kühnert *Phylodynamics of Human influenza B virus in Australia and NZ*

11:30 Alex Popinga *Determinism and stochasticity in SIR epidemics: Extension of the coalescent for serially-sampled data*

11:50 Gabriel E. Leventhal *Exact tree likelihoods for nonlinear population dynamical models*

Epidemics and population dynamics II

14:00 Alexei Drummond (short talk) *Introducing Beast 2*

14:10 Barbara Holland *Assessing the fit of open and closed models of bacterial genome evolution*

14:30 Tim Vaughan *Recombination-aware phylogenetics for bacteria using BEAST*

14:50 Stephane Guindon *The coalescent with randomly fluctuating rate*

Origins

15:40 Velimir Gayevskiy (short talk) *ObStruct: A method to objectively analyse factors driving population structure using Bayesian ancestry profiles*

15:50 David Penny *Loss of information at deeper divergences*

16:10 Peter Wills *Spontaneous self-organisation of phenotypic meaning of genetic sequences; what should be the axioms of phylogenomic analysis?*

16:30 Mike Steel *Autocatalytic sets and the origin of life*

18:45 Dinner at Casita Miro See details on Page 24 for restaurant location. The meal is free for registered participants. You are welcome to bring guests at a cost of $70 per person. Please advise David or Stéphane on Monday if you wish to bring a guest or cannot come.

Wednesday 5 February

Free day

Thursday 6 February (Waitangi Day)

Phylogeography

9:20 Velimir Gayevskiy *Phylogenomic signals describe patterns of global microbial dispersal*

9:40 David C. Marshall *Global biogeography, phylogeny, and diversification of a cicada radiation from Australasia dependence of gene rates and dates on taxon sampling*

10:00 Chris Simon *The molecular evolution of insects that count, or, living life in four year jumps*

Distances

10:50 Andreas Spillner *Some observations on the UPGMA algorithm*

11:10 Mareike Fischer *The maximum parsimony score and its properties*

11:30 Stefan Grünewald *The quartet distance between phylogenetic trees and an application beyond phylogenetics*

11:50 Peter Waddell *Genomic distance-based techniques for quantifying pre Homo sapiens/neanderthalensis archaics interbreeding with more derived lineages of the Homo species complex.*

Speed

14:00 Arndt von Haeseler *Ultrafast Bootstrap*

14:20 Bui Quang Minh *Terrace-Aware Phylogenomic Inference*

14:40 Olivier Gascuel *Fast dating using least-squares criteria and algorithms*

15:00 Gordon Hiscott *The use of delayed acceptance sampling to increase efficiency of SNAPP*

Culture and language

15:50 Remco Bouckaert *Phylogeography of Australian Aboriginal Languages*

16:10 Elsa Guillot *Simulating social and genetic dynamics of human population*

16:30 Quentin Atkinson *The ghost of cultures past - testing for statistical non-independence between human cultures*

Friday 7 February

Rates and selection

9:20 Mike Steel (short talk) *Predicting the past: some new results for ancestral state reconstruction*

9:30 Sarah Parks *A new test for positive selection with greatly improved power*

9:50 David Duchene *Misleading estimates of evolutionary timescales driven by time-dependent rates of evolution*

10:10 Nicolas Privault *Closed form modeling of rates of evolution using geometric Brownian motion and the CIR process*

Sequences and mutation models

11:00 Arnt von Haeseler (short talk) *Coverage patterns in high throughput sequencing*

11:10 Rob Lanfear *Exploring the utility and limits of partitioning in phylogenetics*

11:30 Bennet McComish *Microsatellite evolution in ancient and modern penguins*

11:50 Michael Woodhams *Lie Markov DNA mutation models: when theory meets Data*

2 Abstracts

Quentin Atkinson

The ghost of cultures past — testing for statistical non-independence between human cultures

The social sciences commonly treat data from different cultures as statistically independent. This is despite the fact that statistical non-independence of cultures has been recognised since 1888, when Sir Francis Galton famously noted that the similarities between cultures could also arise due to common ancestry or more recent diffusion of ideas. Here we test the performance of two methods for quantifying the relative contribution of common ancestry (as measure by language phylogeny) and diffusion (as measured by geographic distance) to cultural variation. We apply these methods to several cross-cultural datasets.

Co-authors: Sam Passmore and Ties Coomber

Monika Balvočiūtė

Computation of Planar Split Networks

Planar split networks are a type of phylogenetic networks that can be drawn in the plane without edge crossings what facilitates data analysis. They can be used to detect and visualize potential relationships between sequence evolution and other aspects of a data set such as, for example, the geographic distribution of the corresponding organisms as well as indicate possible conflict in the data (for example, in case of recombination). General planar split networks have the advantage over circular (such as those produced by Neighbor-Net) that their labels are not constrained to lie on the outside of the network. This gives more flexibility to the range of splits that may be shown thus providing more detailed overview of the data than phylogenetic trees or circular split networks (note that trees and circular (outer-labeled) split networks are also planar, but more specific). We will discuss properties and applications of these split systems and describe a new method for constructing them from data.

Remco Bouckaert

Phylogeography of Australian Aboriginal languages

Over 300 Australian aboriginal languages are known and they are fairly uniformly spread over the continent. However, their origin and history of dispersal is unknown. In this exploratory study, we attempt to apply phylogeographical methods developed for virology to the circa 300 languages for which data is available to get better insights in the geographical development of these languages. This talk will highlight some of the stumbling blocks that we ran into. This includes MCMC convergence issues and the fact that phylogeographical methods tend to place the origin around the centre of the sample locations, suggesting that the language group originates from the rather implausible location of the far outback of Australia. In this talk, some solutions to these issues will be presented.

David Bryant

Getting excited about diversities

For the past five years I've been working with Paul Tupper (Simon Fraser University) and others on the mathematics of 'diversities', a concept originating in work on evolutionary diversity and the tight span theory of Dress. The field has proved far richer and diverse than we would have ever anticipated. In this talk I'll give a quick introduction to the basic ideas, discuss recent discoveries relating to diversity embeddings, and speculate wildly about future directions.

Paul Cordue
Characterising when a phylogenetic network displays a tree twice
Evolutionary [phylogenetic] trees have successfully described evolution since Darwin's 1859 paper Origin of Species. The reconstruction of evolution now faces new challenges as evolution may not be fully represented by an evolutionary tree but by an entwined network that reflects how a species inherits its DNA from more than one ancestor. Such a network is a better representation of the evolution of the species when there are a significant number of events such as hybridization, lateral gene transfer, and recombination. New questions and concepts arise when studying evolutionary [phylogenetic] networks, and one such question is: When does a phylogenetic network display a phylogenetic tree twice? This talk will introduce the concepts that are needed to answer the above question, and an efficient algorithm will be presented that decides the question for a class of phylogenetic networks that lies strictly between tree-child networks and tree-sibling networks.

David Duchene
Misleading estimates of evolutionary timescales driven by time-dependent rates of evolution
Several factors can produce a time-dependent pattern in molecular rates of evolution. In particular, short-term rates can be elevated due to mutations that are not observed on longer timescales due to loss. Additionally, there are methodological factors that can lead to an inflation of rate estimates. This time-dependent pattern poses a challenge for the estimation of evolutionary timescales. In the context of this time-dependent rate variation, it is unclear whether the age of calibrations has an effect on the estimated rates and timescales, assuming that calibrations are accurate sources of timescale. We explore the impact of this phenomenon by simulating data sets with a pattern of time-dependent rates and re-estimating the phylogenies with calibrations in different sections of the tree. We reveal that the age of the calibration dictates the estimate of the rate and the overall timescale of the phylogeny. We suggest that in empirical data sets with suspected time-dependent rates it is necessary to include calibrations at both deep and shallow sections of the phylogeny, to avoid misleading estimates of the evolutionary timescale.
Co-authors: Sebastian Duchene and Simon Ho

David Duchene
(5 minute talk) Phylogenetic timescale estimation when molecular evolution and diversification rates are linked
Estimates of the tempo and mode of diversification often rely on molecular phylogenies with branch lengths proportional to time. Current methods for phylogenetic estimation of substitution rates assume that the branching process is independent of the rate of molecular evolution. An increasing body of evidence, however, shows that molecular evolution is often correlated with net diversification rate of lineages. If the rates of substitutions estimated along branches are influenced by diversification, node age estimates might be misleading, which could further complicate the inference of macroevolutionary patterns from molecular phylogenies. We use simulations to explore the effect of the link between diversification rate and the rate of molecular evolution on phylogenetic reconstruction of evolutionary timescales.
Co-authors: Xia Hua, Sebastian Duchene, and Lindell Bromham

Mareike Fischer
The maximum parsimony score and its properties
Within the field of phylogenetics there is great interest in distance measures to quantify the dissimilarity of two trees. In my talk, based on an idea of Bruen and Bryant, I propose and analyze a new distance measure: the Maximum Parsimony (MP) distance. This is based on the difference of the parsimony scores of a single character on both trees under consideration, and the goal is to find the character which maximizes this difference. I will introduce this new metric and show some of its interesting properties, including its relationship to the well-known SPR distance. On the complexity side, I will show that calculating the MP distance is in general NP-hard, and identify an interesting island of tractability in which the distance can be calculated in polynomial time.
Co-author: Steven Kelk

Olivier Gascuel
Fast dating using least-squares criteria and algorithms
Very large virus sequence datasets are available nowadays, typically from HIV, with thousands of dated sequences that have been sampled trough time. Such serial data can be used to study the origin and diffusion of epidemics, estimate the dates of major events (e.g. introduction in a given country or continent), decipher the variation of substitution rate (e.g. due to drug treatments), and reveal the forces shaping the tree (e.g. speciation and extinction rates). We also have more and more paleo-sequences from ancient DNA, for which estimating the rates and dates is an essential task. Current sophisticated probabilistic methods hardly deal with such large datasets comprising thousands of sequences, and even existing distance-based methods (much faster than probabilistic ones) may be faced to computational difficulties. We thus revisited previous approaches and models in a least-squares framework that allows for very fast linear algebra algorithms. We describe two main algorithms, both assuming a strict molecular clock and inspired from the Langley-Fitch model. The first algorithm estimates in linear time the clock-like substitution rate of the input tree and the dates of the root and all interior nodes, but do not impose any precedence constraints on the dates of nodes with ancestry relationships. The second algorithm achieves the same task, but accounts for these temporal constraints using a quadratic programming approach. Moreover, our least-squares model is able to accommodate for some violation of the strict molecular clock. We compare these two algorithms to existing standard programs (Pathogen, r8s and BEAST) and show on simulated data that their performance is similar to that of the most sophisticated probabilistic methods, while their computing time is much faster. Then, we apply these algorithms on two real data sets: a small data set comprising 17 strains of dengue virus and a large data set comprising 1,195 strains of Influenza (H1N1) virus. Again the results show that these algorithms provide a very fast alternative with results similar to those of other probabilistic programs.
Co-authors: Thu Hien To, Matthieu Jung, Samantha Lycett

Alexandra Gavryushkina
Bayesian MCMC for Sampled Ancestor Trees: Application to Fossil Calibration
Probabilistic models that are used for phylogenetic analysis where all samples come from the same time point do not allow for direct ancestry within the sample. However, using serial sampled data or fossils suggests the need for models that allow one sampled individual to be an ancestor of another sampled individual. These models produce trees with sampled ancestors or sampled ancestor trees. We developed a Bayesian MCMC sampler for sampled ancestor trees.
It is common to employ fossils to estimate absolute values of species divergence times. Placing fossils on the tree correctly and using them efficiently is a longstanding problem. Recently, Heath et. al. proposed a method where fossilization events are explicitly modelled as a part of the tree process. One of the main advantages of this method is that it can incorporate all the available fossils into the inference. In this model, fossil samples may be direct ancestors of extant species or other fossils. Being a general tool, the sampler can be applied to inference of species divergence times when fossils may be direct ancestors.
Co-authors: David Welch and Alexei Drummond

Velimir Gayevskiy
Phylogenomic signals describe patterns of global microbial dispersal
Our fundamental question of interest is to investigate the phylogenomics of microbes, and how global distributions have been affected by anthropomorphic actions. New Zealand, being the last major landmass colonized by humans, represents an interesting and unique environment to investigate these questions. We focused on the research model and commercially important yeast Saccharomyces cerevisiae. Previous work has shown that New Zealand harbors a diverse but globally genetically distinct population of this species. Our questions address the origin and age of this population. Was this species present in NZ before humans arrived $\tilde{7}00$ years ago, or did humans unwittingly introduce this species? What is the origin of this group? Are some or all remnants of a more ancient population, or have they invaded from other areas more recently? To address these questions we have obtained the whole genome sequences of 52 strains chosen to represent the diversity of this species in New Zealand, and compared them to the 60 genomes from other S. cerevisiae isolates from varied global locations, and 36 genomes of its sister species S. paradoxus.

Velimir Gayevskiy
(5 minute talk) ObStruct: A method to objectively analyse factors driving population structure using Bayesian ancestry profiles
Bayesian population genetics inference methods implemented in software packages such as Structure and InStruct are extensively used to infer the presence of population structure within genetic data. The primary outputs of these methods are ancestry profiles of individuals among a number of computationally inferred populations. Presently, these ancestry profiles are visualized using distruct, and there are no objective tests conducted to determine to what extent the factor of interest (usually geographic origin) correlates with inferred population structure, and if so, which populations are driving the structure.
I would like to present ObStruct, a novel application of a classic statistical method to objectively analyze these ancestry profiles to estimate the amount of structure in a data set, provide information on the relative contributions of sampled and inferred populations to this structure, test the statistical significance of differences between populations and crucially determine whether the factor of interest correlates with the inferred population structure. ObStruct works alongside existing analysis methods in population genetics and provides a useful final step in the pipeline for analyzing population genetics data.
Co-author: Steffen Klaere

Michael Gemmell
(5 minute talk) Conundrums caused by conflicting shapes and sequences: When mtDNA and morphometrics disagree
Morphology is an invaluable tool for inferring the evolutionary history of New Zealands marine gastropods from the fossil record through to the modern fauna and has played an important role in species delimitation. Molecular techniques enable robust inferences of the relationships among taxa but this is revealing anomalies between genetic phylogenies and relationships implied by taxonomy. An example of this is the species complex consisting of Buccinulum colensoi, B. vittatum vittatum and B. vittatum littorinoides. These divisions are based on the shell morphology and distribution of the taxa. A new geometric morphometric analysis shows the groups inhabiting overlapping morphospace with a stronger similarity between the two vittatum sub species and a separation of B. colensoi, as reflected in the species delimitation. An analysis of the mitochondrial DNA indicates that the genetics are not consistent with the taxonomy. The mtDNA groups B. colensoi with B. vittatum littorinoides leaving B. vittatum vittatum as sister to this cluster. Mismatch of morphology and mtDNA could be the result of a number of processes: 1. Recent introgression (mtDNA capture). 2. Environmentally induced plasticity. 3. Rapid morphological evolution driven by selection. We aim to test these alternative explanations and to clarify the phylogeny of this group by obtaining nuclear data from next generation sequencing using a RAD approach and utilising population genetic tools.
Co-authors: Mary Morgan-Richards, Steve Trewick, Simon Hills and Felix Vaux.

Gillian Gibb

Why fly when you can walk? Flightlessness in modern and extinct NZ rails.

It is a curious phenomenon that many birds are flightless, when for most of us birds are quintessentially flyers. For a flightless species to evolve from a flying ancestor there must be good ecological reasons and pathways for selection: options for being flightless must exist. The rails (Rallidae) are a cosmopolitan group of birds with a broad distribution through the world. Flightlessness has evolved independently multiple times within the group, often in association with the colonisation of islands and there are numerous examples in the New Zealand/Pacific region. Nearly all islands of the Pacific have been colonized by one or more rail lineages, and before human contact it is likely that most islands had endemic flightless rails. Evolution of flightlessness in rails has been shown to occur rapidly, and may be measured in generations rather than millennia. While ecological reasons for loss of flight may be obvious, what sort of genetic adaptations might be underlying the evolution of flightlessness? We use both modern and ancient DNA to investigate whether there has been relaxed selection in metabolic genes in flightless New Zealand rails.

Co-author: Steve Trewick

Stefan Grünewald

The quartet distance between phylogenetic trees and an application beyond phylogenetics

In phylogenetic data analysis, using different methods or different data sets frequently results in different trees with the same label set. Further, some heuristic optimization algorithms require the construction of a different but not too different tree from a given one, in order to find a (locally) optimal tree. Therefore, we have to quantify the dissimilarity between two phylogenetic trees with identical leaf sets.

One common measure for this is the quartet distance. It is defined to be the number of sets of exactly four taxa for which the trees have different restrictions. Bandelt and Dress showed in 1986 that the maximum distance between two binary trees, when normalized by the number of all 4-sets, is monotone decreasing with n. They conjectured that the limit of this ratio is 2/3 which corresponds to the distance between two random trees.

I will summarize recent improvements to prove a generalization of the conjecture to arbitrary X-trees. In addition, our results show that quartets can be used to quantify the dependence between two real- valued random variables or between two data sets. In some sense, this measure of dependence corresponds to Kendall's tau to measure correlation.

Elsa Guillot

Simulating social and genetic dynamics of human population

Social practices, such as marriage rules and caste systems, have long been known to impact patterns of genetic diversity in the communities that practice them. In turn, these patterns are easily confused with signals of demography or natural selection. This study presents new methods based on simulations, to detect and measure the impact of social practices on population genetics.

Stephane Guindon

The coalescent with randomly fluctuating rate

Kingman's coalescent (1982) model is a simple stochastic process that describes a genealogy under several assumptions, including that of a large and constant population size. When the population size changes following a deterministic function, the probability density of a genealogy is given by the standard coalescent with the time unit rescaled in an appropriate manner. The same goes for the situation where the population size evolves according to a stationary Markov process. However, more realistic models that describe how population size fluctuates, such as the linear birth and death process, are non-stationary. A solution to calculating the distribution of the time to coalescence of two lineages under such model is presented. Comparing coalescent trees with stochastically varying population sizes and birth death trees

Co-authors: Spencer Enesa and Kelly Lu

Simon F. K Hills

Diversification patterns in molecular and paleontological data of the marine mollusc genus Alcithoe

The rich Cenozioc marine mollusc fossil record of New Zealand provides a wealth of paleontological data for integration and comparison with molecular datasets. We have utilised fossil occurrences of living species together with sampling probabilities (of fossils) to generate calibration distributions for the inference of a time-calibrated molecular phylogeny of the volute genus Alcithoe (Gastropoda). Our selection of calibration species eliminates the taxonomic uncertainty associated with the assignment of extinct species to internal nodes in the phylogeny. Calibration prior distributions are derived from measured uncertainty from the fossil record. This approach yielded a well-supported tree with divergence times that are consistent with the fossil record. The pattern of diversification implied by the molecular phylogeny indicates that most Alcithoe speciation has occurred relatively recently. However, this appears to be at odds with the fossil record, in which more taxonomic diversity is observed to have occurred in the past. This observation leads to the question: to what extent does the diversification pattern seen for the living species reflect the history of diversification for the whole clade? Using our time-calibrated molecular phylogeny and diversification rates for New Zealand Mollusca calculated from paleontological data, we examine the diversification process in the Alicthoe lineage during the Cenozoic. By comparing the Alcithoe molecular phylogeny with trees simulated using various diversification models we test for scenarios that can reproduce trees similar to the inferred molecular tree. The models tested are derived from measurements of origination and extinction from the fossil record data and capture increasingly complex parameterisation of the diversification process. Our results show that for Alcithoe, diversification of the living taxa is not representative of the complete evolutionary history of the lineage. Appropriate modelling using the paleontological data is however sufficient to recover a diversification process that can generate phylogenies consistent with the molecular tree.

Co-authors: James Crampton, Steve Trewick, Mary Morgan-Richards

Gordon Hiscott

The use of delayed acceptance sampling to increase efficiency of SNAPP

SNAPP, which is short for SNP and AFLP Phylogenies, is a program for inferring species trees from unlinked SNPs, or single nucleotide polymorphisms. Unfortunately, as the number of individuals increases, the complexity of the likelihood calculation causes SNAPP to run very slowly. Our current approach to this problem is finding techniques which carry out exact sampling more efficiently through the use of approximations. We use delayed acceptance sampling in order to accelerate SNAPP analysis, initially for problems in which there are only two populations.

Barbara Holland

Assessing the fit of open and closed models of bacterial genome evolution

Multi-locus sequence typing (MLST) has long been a popular means of typing bacterial species. Traditionally 7 housekeeping genes are sequenced and each strain of bacteria is summarized by 7 numbers recording which unique sequence the strain has for each of the genes. With the advent of cheap genome sequencing this concept can easily be extended to larger numbers of genes giving allele profiles that may relate to many 100s of genes. The allele profile data has some interesting properties from a phylogenetic standpoint. Bacteria evolve by a mixture processes that act vertically on the tree such as inheritance and mutation as well as processes that act horizontally across the tree. A mutation event or recombination event both have the same effect on the allele profile of changing the unique identifier for one of the genes. One question that arises is the extent to which the allele profile data fits the infinite alleles model. If sequences for each gene are long then mutation should usually introduce a sequence type that has not been seen before. With recombination it will depend on the extent to which the bacteria under consideration form a closed or open system. In this talk I investigate the fit of closed and open models to allele profile data derived from 46 Campylobacter jejuni genomes.

Co-authors: Nigel French, Patrick Biggs, Shoukai Yu

Denise Kühnert
Phylodynamics of Human influenze B virus in Australia and New Zealand
We analyze the epidemiological dynamics of human influenza B viruses using DNA sequence data from Australia and New Zealand sampled between 20022013. The BDSIR analysis of single seasonal epidemics reveals similarities and differences be- tween the two major groups, B/Yamagata/16/88-like and B/Victoria/2/87-like strains. The two lineages are believed to have co-circulated on a global scale since about 40 years. We discover a higher transmission potential of the Victoria strains, with larger variation among single years. Interestingly, the joint analysis of samples spanning the whole sampling period of 11 years under a coalescent skyride tree prior results in significantly longer branch lengths within the seasonal subtrees than the analysis assuming a birthdeath skyline tree prior.

Rob Lanfear
Exploring the utility and limits of partitioning in phylogenetics
Partitioning is a routine step in most attempts to infer phylogenies. It involves the inference of independent parameters for different groups of sites in a sequence alignment. The ideal solution to the partitioning problem is Bayesian - integrating over all possible assignments of model parameters to sites. But for many datasets, Bayesian approaches to partitioning are not feasible. In these cases, researchers must find a way to choose a partitioning scheme, often relying heavily on a priori information about gene boudaries, codon positions, etc. It is well known that these approaches to partitioning produce models that fit the data better than not partitioning at all, and we think this is because our a priori information reflects meaningful differences in patterns of molecular evolution between genes and codon positions. But can we go further than this? Are there ways to optimise partitioning schemes in the absence of a priori information on patterns of molecular evolution, or ways to build on that information? Can we build methods that will work on genome-scale datasets in reasonable amounts of time? And, finally, does all of this make any difference to the trees that we infer? In this talk, I'll have a go at answering these questions.

Gabriel E. Leventhal
Exact tree likelihoods for nonlinear population dynamical models
Tree likelihood expressions that go beyond the conventional constant-rate birth-death or coalescent models can account for complex population dynamical models. This allows for for more precise inference of population dynamic parameters, such as density-dependence in species diversification (Etienne 2012) and susceptible-infected-susceptible (SIS) epidemic models (Leventhal 2014). I will present a modeling framework that can compute the tree likelihood under a variety of non-linear time-homogeneous population dynamic models with large state spaces. In addition to more precise process reconstruction, certain parameter degeneracies (Stadler 2013) can be resolved using only slightly more complex dynamical models.
Etienne et al. Proc R Soc B, 2012 - Leventhal et al. Mol Biol Evol, 2014 - Stadler et al. Proc Nat Acad Sci U.S.A., 2014

Simone Linz
Surprising properties of maximum parsimony on networks
Maximum parsimony is a popular technique to reconstruct phylogenetic trees from a set of characters. Since trees are not best suited to represent non-tree-like processes such as horizontal gene transfer, hybridization, and recombination, two different definitions for calculating a parsimony score of a phylogenetic network have recently been proposed. In this talk, we will review these concepts and explain some surprising implications for the big parsimony problem, i.e. finding a most parsimonious network among all possible phylogenetic networks.
This is joint work with Christopher Bryant and Charles Semple.

Peter Lockhart
Hybridization may facilitate in situ survival of endemic species through periods of climate change
Predicting survival and extinction scenarios for climate change requires an understanding of the present day ecological characteristics of species and future available habitats, but also the adaptive potential of species to cope with environmental change. Hybridization is one mechanism that could facilitate this. Here we report statistical evidence that the transfer of genetic information through hybridization is a feature of species from the plant genus Pachycladon that survived the Last Glacial Maximum in geographically separated alpine refugia in New Zealand's South Island. We show that transferred glucosinolate hydrolysis genes also exhibit evidence of intralocus recombination. Such gene exchange and recombination has the potential to alter the chemical defence in the offspring of hybridizing species. We use a mathematical model to show that when hybridization increases the adaptive potential of species, future biodiversity will be best protected by preserving closely related species that hybridize rather than by conserving distantly related species that are genetically isolated.
Co-authors: Matthias Becker, Nicole Gruenheit, Mike Steel, Claudia Voelckel, Oliver Deusch, Peter B. Heenan, Patricia A. McLenachan, Olga Kardailsky, Jessica Leigh

David C. Marshall
Global biogeography, phylogeny, and diversification of a cicada radiation from Australasia — dependence of gene rates and dates on taxon sampling.
The globally distributed cicada tribe Cicadettini is a rare example of a global terrestrial animal group with an Australasian origin. We sequenced nuclear and mitochondrial DNA from 200 species from 76 genera plus undescribed material obtained from intensive sampling in Australia and global sampling through a network of world collaborators. Relaxed-clock divergence times were calibrated mainly with literature-derived COI molecular clock estimates, because fossils for the group are few and uninformative. Phylogenetic analyses nested all non-Australasian genera in two clades within the main tribe radiation, and geographic range reconstruction indicated an Australasian ancestor for the group. COI-calibrated molecular clock analyses at the tribe level recovered extremely old dates that conflicted with published estimates for the well-studied, nested Maoricicada-Rhodopsalta-Kikihia (MRK) clade of New Zealand. Relaxed-clock analyses of subclade datasets combined with maximum-likelihood analyses of the same showed that COI rates and branch lengths became progressively inflated relative to EF1-α intron and exon rates as phylogenetic depth increased. Final divergence-times were estimated using multiple methods, including an extrapolated-COI clock for the EF1-α coding partition, relative-time-scaling from analyses of the MRK clade, and exploitation of a geological calibration point from New Caledonia. These results, while variable, were consistent with diversification commencing in the early-to-mid-Cenozoic and continuing with the development of mesic and arid Australian habitats. Extreme and unexpected among-lineage diversification rate contrasts were found within the Cicadettini and between Cicadettini and its trans-Australian sister clade.
Co-authors: Kathy B. R. Hill, D. Vanderpool, John R. Cooley, and Chris Simon

Catherine Matias
Co-phylogeny reconstruction via an approximate Bayesian computation
Despite an increasingly vaster literature on co-phylogenetic reconstructions for studying host-parasite associations, understanding the common evolutionary history of such systems remains a problem that is far from being solved. Most algorithms for host-parasite reconciliation use an event-based model, where the events include in general (a subset of) co-speciation, duplication, loss, and host-switch. All known event-based methods then assign a cost to each type of event in order to find a reconstruction of minimum cost. The main problem with this approach is that the cost of the events strongly influence the reconciliation obtained. To deal with this problem, we developed an algorithm, called Coala, for estimating the frequency of the events based on an approximate Bayesian computation approach. The benefits of this method are twofold: (1) it provides more confidence in the set of costs to be used in a reconciliation, and (2) it allows to estimate the frequency of the events in cases where the dataset consists of trees with a large number of taxa. We evaluate our method on simulated and on real datasets.
Co-authors: C. Baudet, B. Donati, B. Sinaimeri, P. Crescenzi, C. Gautier and M-F. Sagot

Bennet McComish
Microsatellite evolution in ancient and modern penguins
Microsatellites are short tandem repeat sequences that have been widely used as genetic markers in a variety of population genetic studies. The number of repeats at a locus is thought to change by slippage of DNA polymerases during replication, and these loci exhibit high levels of length polymorphism. Several models of microsatellite evolution have been developed, some of which take into account both replication slippage and point mutation. These models vary widely in complexity, but it is not clear that any of them succeed in capturing all of the relevant biological processes. I will present some new results from modern and ancient genome sequences of Adelie penguins, and discuss how these might be used in testing existing models of microsatellite evolution and developing new ones.

Mait Metspalu
(5 minute talk) Database of over 400 complete human genomes from Eurasia and beyond
We have generated a database of over 400 complete human genomes (ca. 40X) from Eurasia and beyond with a strategy to cover many populations with ca. three genomes per population. Together with the publicly available genomes sequenced with the same platform (Complete Genomics) the number of genomes available for analysis is over 700. This database will be publicly available but we are open to collaborative projects even before we can release the data to the public. I will outline the major lines of research we are undertaking using these data and encourage interested parties to join in and help utilize the data to the fullest.

Bui Quang Minh
Terrace-aware phylogenomic inference
Recently Sanderson, McMahon, and Steel (Terraces in phylogenetic tree space, Science 333:448-450, 2011) showed that due to missing data many trees on a terrace have the same likehoods under a full partition model. Here, we exploit this fact to derive a terrace-aware data structure that avoids likelihood computation for trees belonging to the same terrace. The data-structure speeds up phylogenomic tree inference (e.g., by a factor of 5 for an alignment with 73% missing data) and reduces memory consumption. We highlight further developments of the IQ-TREE software in the light of phylogenomics.
Co-authors: Olga Chernomor and Arndt von Haeseler

Fabio Pardi
Reconstructible phylogenetic networks: no need to distinguish the indistinguishable
Most proposed methods for phylogenetic network reconstruction evaluate candidate networks on the basis of the trees they display. This is certainly true for all methods based on data consisting of clusters of taxa, triplets, quartets, or trees with any number of leaves, but also for sequence-based approaches such as the first formalisations of maximum parsimony and maximum likelihood for networks.
This poses a problem: from the perspective of these methods, all networks that display the same set of trees are "indistinguishable", as the objective function assigns the same score to all networks displaying the same set of trees. This problem is partially solved by accounting for branch lengths, although this merely reduces the size of the classes of indistinguishable networks.
In this talk we propose a novel definition of what constitutes a "uniquely reconstructible" network: for each class of indistinguishable networks, we define a canonical form. Under mild assumptions, the canonical form is unique. Given data coming from any phylogenetic network, only its canonical equivalent can be uniquely reconstructed. This is a fundamental limitation that implies a drastic reduction of the solution space in phylogenetic network inference.
Co-author: C. Scornavacca.

Sarah Parks

A new test for positive selection with greatly improved power

Finding positive selected genes, or sites in genes, is a key question in biology. A variety of maximum likelihood and Bayesian methods for testing for positive selection using omega, the ratio of the fixation rates of non-synonymous and synonymous mutations, have been developed. These are all very conservative when applied to real genes, and hence achieve a lower power than desired.

One of these tests, embodied in the SLR method (Massingham and Goldman, 2005), estimates the maximum likelihood value of omega for each site, and then tests whether this value is greater than one. This test is conservative because the null hypothesis assumes that all sites are either neutral or positively selected ($\omega \geq 1$), whereas in reality the majority of sites in genes are under purifying selection ($\omega < 1$).

I will present a new method to test for positive selection that has greatly improved power whilst retaining control of the false positive rate. This involves a new sitewise likelihood ratio test, designed to have power and control when many of the sites in the gene are under purifying selection (as is typically the case), and a diagnostic for detecting certain situations in which the original SLR test should be preferred. The new test obtains improved power by fitting the null hypothesis to the data and then performing parametric bootstraps.

The method has been tested using simulations over a wide range of realistic conditions, including standard comparisons used in previous studies, and larger and more realistic examples modelled on real-life studies. I will show that, for those rare cases where all sites are either strictly neutral ($\omega = 1$) or positively selected ($\omega > 1$), the new method performs as well as SLR. More importantly, for genes where many of the sites are conserved, this method has much better power than other tests and a controlled false positive rate.

Co-author: Nick Goldman

David Penny

Loss of information at deeper divergences

It has been shown but Mossel and Steel (2004) that simple Markov models lose information at the deepest divergences; and that the fall off is exponential at deeper times. However, that does not mean that there is no information left, for example, the three-dimensional structure of proteins may still retain information at deeper divergences. Biologists still want to estimate these deeper divergences and thus it is a significant question to find additional sources of information. Several suggestions are offered for a formal analysis. Firstly, we probably expect that where there is a real Gamma distribution of rates that information may be retained for longer. Secondly, if there is really a bimodal distribution of rates, then identifying, and eliminating, these faster evolving sites should help. Thirdly, the inference of ancestral sequences at deeper divergences appears quite robust, and there is some evidence that this may help recover deeper divergences. Fourthly, it is increasingly possible to infer three-dimensional structures, and these should retain information longer. Fifthly, there may be differences between the loop regions of Akaryote and Eukaryote genes, and only taking the regions crossing the central 3D region might help. Sixthly, an approach of weighting, not of characters, but of the partitions they are consistent with, might help. Seventhly, possibly gene order information might be helpful. Several examples of such approaches will be presented, and a challenge issued to mathematicians to solve some of these fundamental issues.

Alex Popinga
Determinism and stochasticity in SIR epidemics: Extension of the coalescent for serially-sampled data
In the interest of understanding the origins, patterns, and driving forces critical to the progression of epidemic outbreaks, a conjunct approach is employed to utilise the information embedded in both evolutionary and environmental processes. Key epidemiological parameters, such as the rates of pathogen transmission and death, can be estimated by tying together genealogical and epidemiological models.
In this study, we implement and extend a published coalescent method [Volz (2012)] for Bayesian inference of key epidemic parameters drawn from a Susceptible-Infected-Removed (SIR) branching scheme. Our extensions exploit the additional information embedded in time-stamped (heterochronous) sequence data and operate on both stochastically and deterministically described trees and corresponding trajectories of the underlying host populations. We compare the performance of these extensions with that of a recent birth-death-sampling method [Kuhnert et al. (2014)].
Co-authors: Tim Vaughan, Alexei Drummond

Nicolas Privault
Closed form modeling of rates of evolution using geometric Brownian motion and the CIR process
In this talk we are interested in the statistical properties of average substitution rates Z when the rate itself is modeled according to a geometric Brownian motion or a CIR process. We discuss the availability of closed form expressions for various statistical parameters such as mean, variance and probability density functions, and we compare them with currently used approximations based in particular on the gamma density.

Celine Scornavacca
A practical approximation algorithm for solving massive instances of hybridization number for binary and nonbinary trees
Reticulate events play an important role in determining evolutionary relationships. The problem of computing the minimum number of such events to explain discordance between two phylogenetic trees is a hard computational problem. Even for binary trees, exact solvers struggle to solve instances with reticulation number larger than 40-50.
In this talk we present CycleKiller and NonbinaryCycleKiller, the first methods to produce solutions verifiably close to optimality for instances with hundreds or even thousands of reticulations. Using simulations, we demonstrate that these algorithms run quickly for large and difficult instances, producing solutions that are very close to optimality.
Co-authors: Leo van Iersel, Steven Kelk and Nela Lekic

Charles Semple
Counting phylogenetic networks
The number of binary phylogenetic networks on ℓ taxa is a classical result in mathematical phylogenetics dating back to Schröder's work in 1870. This result also gives the number of such trees on n labelled vertices. In contrast, the number of binary phylogenetic networks on n labelled vertices is unknown. In this talk, we provide some answers to the problems of counting the numbers of phylogenetic networks.

Chris Simon

The molecular evolution of insects that count, or, living life in four year jumps

Three distinct periodical cicada lineages occur in the eastern U.S. (Decim, Cassini, and Decula) and have genetically programed 13- or 17-year synchronized life cycles. It has always been assumed that these three lineages evolved in parallel and simultaneously because 1) they co-occur through most of their ranges, 2) they are locked together ecologically and evolutionarily by their predator satiation strategy, and 3) they have peculiar life cycles that are unlikely to have evolved multiple times independently. Recently we provided molecular phylogenetic evidence that, the 13- or 17-year life cycle has arisen at least eight times within the extant representatives of the three lineages. We analyzed nuclear and mitochondrial DNA markers for samples collected throughout the range of all three lineages between 1978 and 2008. We found that the earliest divergence occurred 3.9 million years ago (mya) with one branch forming the Decim species group, followed by the subsequent splitting of the other branch 2.5 mya into the Cassini and Decula species groups. We suggest that the earliest common ancestor was already periodical with either a 13- or 17-year life cycle. Our phylogeny suggests that the three lineages independently diverged into 13- and 17-year life cycles, with the earliest extant split occurring in the Decim lineage 0.5 mya and all three lineages experiencing at least one instance of life-cycle divergence since that time. The presence of repeated, independent 4-year life cycle shifts in all three lineages, implies that the species groups share a common genetic basis for the 13- and 17-year life cycles and that these cycles originated prior to the species groups. Extensive geographic sampling and Ecological Distribution Modeling suggests that 13-year cicadas are found in areas with longer growing seasons, i.e., the southern US and Mississippi Valley region. Our phylogeographic studies suggest that northern 17-year populations of the Cassini and Decula lineages invaded southern 13-year Decim populations during and after Pleistocene glacial cycles and that synchronization of life cycles between invading and resident periodical cicada populations may have been favored by natural selection because it helped invaders escape predation. I speculate on the causes of evolution of 13-17-year periodical life cycles.

Co-authors: Teiji Sota (Kyoto University), Satoshi Yamamoto (Kyoto University), John R. Cooley U. Connecticut), Kathy B.R. Hill (U. Connecticut), and Jin Yoshimura (Shizuoka U.)

Andreas Spillner

Some observations on the UPGMA algorithm

Given a distance matrix on a set of taxa, we consider a natural rooted variant of the minimum evolution problem and note that the well-known UPGMA-algorithm can be viewed as a bottom-up greedy heuristic for it. Moreover, we describe how the minimum evolution score of any binary rooted phylogenetic tree T results as the average length of a minimum spanning tree compatible with the set of clusters induced by T. This is the rooted analogon of the known interpretation of the balanced minimum evolution score of an unrooted phylogenetic tree as the average length of a spanning cycle compatible with that tree. Finally, we show that a local search of the space of all binary rooted phylogenetic trees using rooted nearest neighbor interchanges will yield the unique optimal tree for any input distance matrix that is a generic ultrametric.

Tanja Stadler

A unified framework for inferring phylogenies with fossils

Molecular and fossil data are observations from the same speciation and extinction process. However, when dating (time-calibrating) phylogenies, we typically assume different models for the generation of the extant species phylogeny (obtained from the molecular data) and for the fossil data. While extant species phylogenies might be modelled through a speciation and extinction process, fossil occurrence times are modelled through arbitrary calibration densities. I will discuss our new approach assuming a speciation and extinction model simultaneously for both data types. We show in a simulation study the improvement in accuracy of dating phylogenies, while avoiding arbitrary calibration densities.The method is implemented in DPPDiv (http://phylo.bio.ku.edu/content/tracy-heath-dppdiv) and the manuscript is available on arXiv:1310.2968.

(Joint work with Tracy Heath and John Huelsenbeck)

Mike Steel

Autocatalytic sets and the origin of life

A key step in the emergence of early life was the formation of a set of reactions that is (i) self-sustaining (all reactions involve reactants that are either produced by other reactions or are available in the environment) and (ii) autocatalytic (each reaction is catalyzed by some molecule produced by the system). Mathematical, algorithmic and stochastic techniques can help define, analyse, classify and search for these self-sustaining autocatalytic system in large chemical reaction networks. In this talk, I will provide a brief overview of some of our earlier and recent result in this area.

Joint work with Wim Hordijk and others.

Tim Vaughan

Recombination-aware phylogenetics for bacteria using BEAST

Bayesian phylogenetic methods such as those implemented in BEAST are often used to infer evolutionary and demographic model parameters from genetic sequence data. In this context, the phylogeny itself is merely the glue which ties the sequence data to the population genetics/dynamics models that describe the larger-scale behaviour of the population.

This style of inference has been used to great effect in the study of viral evolution. However, its application to the study of bacterial populations is currently impeded by the limited availability of methods capable of properly and efficiently dealing with horizontal gene transfer events, which are a characteristic feature of the molecular evolution of bacterial genomes.

In this talk we present an inference framework for bacterial evolution in BEAST 2 based on the ClonalOrigin model (Didelot et al., Genetics, 2010), the aim of which is to facilitate both parametric and non-parametric demographic model inference using bacterial sequence data. We will discuss the inference scheme and the technical details of its implementation. We will also discuss our intended application: to infer the demographic history of populations of bacteria belonging to the genus Campylobacter in New Zealand using full-genome data.

Co-authors: Alexei Drummond, Nigel French

Felix Vaux

Whelk phylogenetics as a test of punctuated evolution

New Zealands marine molluscs have long been proposed as an ideal scenario to investigate species evolution. This is due to unique geological preservation where the fossil record reveals a wide variety of lineages that can be tracked over significant lengths of evolutionary time. Crucially there is also a sufficient level of sampling available to follow morphological change in-depth in many species. Furthermore, the diversity of extant taxa permits for large-scale molecular reconstructions that can be used for rigorous comparison.

In our study we aim to use this dataset to comprehensively test assumptions derived from punctuated equilibrium. While debate over punctuated evolution has persisted over the last four decades, with major revisions, empirical tests of the key assumptions (stasis, rapid speciation with significant evolutionary change, cladogenesis versus anagenesis) have been wanting. This is largely due to an inadequate fossil record or limited genetic data. In particular we aim to focus on the neogastropod family Buccinidae (true whelks) and plan to use molecular, morphological, and integrated phylogenies to test these assumptions. Buccinids exhibit very high intra- and interspecific morphological diversity in both living and fossil species, but little information is known about the genetic relationship between numerous clades. This therefore means that in addition our research shall also improve true whelk taxonomy for New Zealand and abroad, and it should clarify areas of conflict between past morphological and molecular species classifications.

Co-authors: Mary Morgan-Richards, Steve Trewick, Simon Hills, James Crampton

Arndt von Haeseler
Ultrafast Bootstrap
We present an very fast approximation of the classical nonparametric bootstrap in phylogenetic inference. Key ingredients of the approximation are the resampling estimated log-likelihood method and an effective candidate tree collection scheme. With this we obtain a median speed up of 3.1 to 10.2 for DNA and amino acid alignments, respectively. We will also discuss the robustness of our ansatz.
Co-authors: Bui Quang Minh and Minh Anh Thi Nguyen

Peter Waddell
Genomic distance-based techniques for quantifying pre Homo sapiens/neander-thalensis archaics interbreeding with more derived lineages of the Homo species complex
The lines of descent in the genus/species complex Homo are becoming much clearer. Reasons for this include a range of near complete genomes for both modern humans and two distinct groups of archaics, Neanderthals and Denisovans. Another reason has been steadily improving and expanding methods of analysis. These include the P2D2 tests of non-tree descent (e.g. Waddell et al. 2001, Reich et al. 2010, Waddell 2011) and full maximum likelihood modeling under the coalescent with reticulation (Waddell et al. 2011, 2012). Here I will look at how distance matrices can be used to further tease apart reticulate patterns of descent, including from unknown ancestors. Planar graphs generated by the NeighborNet selection algorithm (Bryant and Moulton 2004), combined with constrained weighted least squares edge length estimation, provide a useful starting point to uncover splits in autosomal genomic data. One of the most interesting of these is a Denisovan + Chimp split, that suggests introgression from an archaic member of Homo (erectus-like?) that diverged prior to the modern/Neanderthal split. Here I will discuss techniques to further interrogate this and another very interesting signal of Papuan + Neanderthal + Denisovan + Chimp split, that is consistent with some particularly archaic derived alleles going into the Papuan genome and/or all other modern humans exchanging alleles excluding papuans. A decomposition of the distance-based split spectrum is generated under such a population-genetic scenario and it is fitted to the observed splits in order to estimate the fraction of the Denisovan genome that came from a pre modern/Neanderthal archaic. Analyses of morphological and morphometric data on Homo skulls suggest that the archaic that mixed with Denisovan may have looked like an example of specimens sometimes called H. heidelbergensis.

David Welch
Phylogenetic transmission trees constrained to networks
The pattern of infectious contacts in a host population is frequently modelled as a contact network. Characterising these networks is central to understanding and controlling epidemics. I will discuss work to estimate the parameters of exponential-family random graph models (ERGMs) for these networks using epidemiological data (such as recovery time data) and ongoing work to extend these methods to viral genomic data to aid in the estimation.
This is joint work with David Hunter, Chris Groendyke and Matthew Jones.

Peter Wills
Spontaneous self-organisation of phenotypic meaning of genetic sequences; what should be the axioms of phylogenomic analysis?
The information flow in biological systems is said to be one-way from DNA to RNA to protein, never the reverse. Thus, the accumulation of biologically useful information in genetic sequences is attributed to selection within populations of individuals whose survival is stochastically determined by the nucleic acid sequences they carry. However, the genotype-phenotype mapping cannot be the product solely of the process of selection which requires it. In contrast to standard phylogenetics with its essentially neo-Darwinian assumption, a logically consistent analysis of natural selection based on genetic sequences requires a model of self-organisation processes that operate at the level of the phenotype and species interactions. By using such a method we are getting close to seeing whether the mutual self-organisation of genetic coding and information has left a genomic palimpsest in the amino acid sequences of enzymes.

Michael Woodhams
Lie Markov DNA Mutation Models: When Theory Meets Data
The Lie Markov DNA models allow us to model time-inhomogeneous mutation processes in a consistent way, unlike (for example) the GTR model. A hierarchy of 37 models have been developed which satisfy the Lie Markov symmetry constraints while also allowing a distinction between transitions and transversions. I will present likelihood tests of these models on a selection of real world data sets. In addition, I will demonstrate that "Fourier-Motzkin Elimination" is more than just a good name for a heavy metal band.

Bojian Zhong
Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes
The phylogenetic branching order of the green algal clades that gave rise to land plants remains uncertain despite its fundamental importance to understanding plant evolution. Previous studies have demonstrated that land plants evolved from streptophyte algae, but different lineages of streptophytes have been suggested to be the sister group of land plants. To better understand the evolutionary history of land plants, and to determine the potential effects of "long-branch attraction" in phylogenetic reconstruction, we analysed a plastid genome dataset including three new chloroplast genomes from streptophyte algae: Coleochaetae orbicularis (Coleochaetales), Nitella hookeri (Charales), and Spirogyra communis (Zygnematales). We further applied a site pattern sorting method together with site- and time-heterogeneous models to investigate the branching order among streptophytes and land plants. Our chloroplast phylogenomic analyses are similar to nuclear data in placing Zygnematales alone, or a clade consisting of Coleochaetales and Zygnematales, as the closest living relatives of land plants.
Co-authors: Zhenxiang Xi, Vadim V. Goremykin, Richard Fong, Patricia A. Mclenachan, Philip M. Novis, Charles C. Davis and David Penny

3 List of participants

Aydar Aliev	a.aliev@massey.ac.nz
Patricio Andres Maturana Russel	pmat747@aucklanduni.ac.nz
Quentin Atkinson	q.atkinson@auckland.ac.nz
Monika Balvočiūtė	mokana@gmail.com
Remco Bouckaert	remco@cs.auckland.ac.nz
David Bryant	david.bryant@otago.ac.nz
Michael Charleston	mcharleston@it.usyd.edu.au
Paul Cordue	paul.cordue@pg.canterbury.ac.nz
Giulio Dalla Riva	giulio.dallariva@googlemail.com
Alexei Drummond	alexei.drummond@gmail.com
David Duchene	david.duchene@anu.edu.au
Phillip Endicott	phillip.endicott@gmail.com
Mareike Fischer	email@mareikefischer.de
Tomas Flouri	Tomas.Flouri@h-its.org
Olivier Gascuel	gascuel@lirmm.fr
Alexander Gavruskin	alexander@gavruskin.com
Sasha Gavryushkina	sasha.gavryushkina@auckland.ac.nz
Velimir Gayevskiy	vel@vel.co.nz
Michael Gemmell	m.r.gemmell@massey.ac.nz
Gillian Gibb	gillian.c.gibb@gmail.com
Stefan Grünewald	stefan@picb.ac.cn
Elsa Guillot	e.guillot@massey.ac.nz
Stéphane Guindon	s.guindon@auckland.ac.nz
Simon Hills	s.f.hills@massey.ac.nz
Gordon Hiscott	ghiscott@maths.otago.ac.nz
Barbara Holland	Barbara.Holland@utas.edu.au
Steffen Klaere	s.klaere@auckland.ac.nz
Denise Kuhnert	denise.kuehnert@env.ethz.ch
Robert Lanfear	rob.lanfear@anu.edu.au
Gabriel Leventhal	gabriel.leventhal@env.ethz.ch
Simone Linz	simone_linz@yahoo.de
Peter Lockhart	p.j.lockhart@massey.ac.nz
David Marshall	david.marshall@uconn.edu
Catherine Matias	catherine.matias@genopole.cnrs.fr
Bennet McComish	bennet.mccomish@utas.edu.au
Mait Metspalu	mait@ebc.ee
Bui Quang Minh	minh.bui@univie.ac.at
Fabio Pardi	Fabio.Pardi@lirmm.fr
Sarah Parks	roffice@ebi.ac.uk
David Penny	D.Penny@massey.ac.nz
Alexandra Popinga	apop146@aucklanduni.ac.nz
Nicolas Privault	nprivault@ntu.edu.sg
Céline Scornavacca	celine.scornavacca@univ-montp2.fr
Charles Semple	charles.semple@canterbury.ac.nz

Chris Simon	chris.simon@uconn.edu
Andreas Spillner	anspillner@googlemail.com
Tanja Stadler	tanja.stadler@env.ethz.ch
Mike Steel	mathmomike@gmail.com
Timothy Vaughan	tgvaughan@gmail.com
Felix Vaux	f.vaux@massey.ac.nz
Neeraj Verma	verma.neeraj.9229@gmail.com
Arndt von Haeseler	arndt.von.haeseler@univie.ac.at
Peter Waddell	pwaddell.new@gmail.com
David Welch	david.welch@auckland.ac.nz
Peter Wills	p.wills@auckland.ac.nz
Michael Woodhams	michael.woodhams@utas.edu.au
Bojian Zhong	b.zhong@massey.ac.nz

4 Practicalities

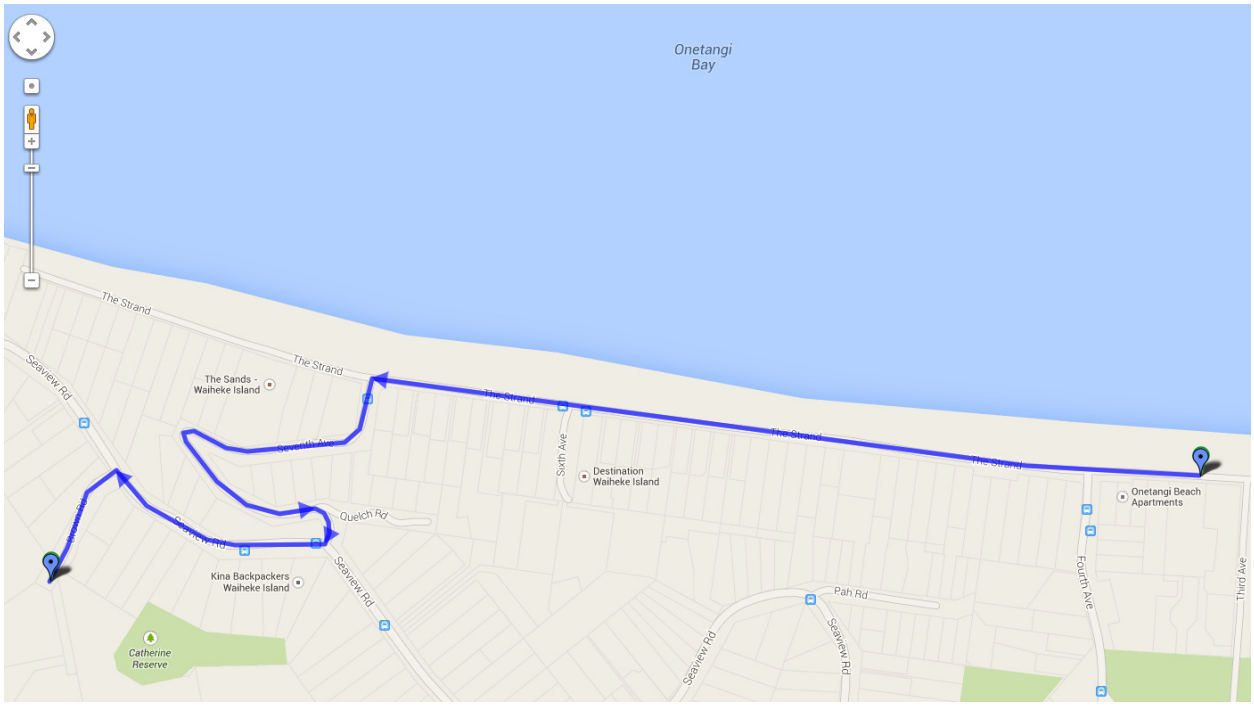
Local shops

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

Restaurant details

The conference dinner is Tuesday evening from 18:45 at Casita Miro (www.casitamiro.co.nz/), 3 Brown Road, Onetangi, just above the western end of Onetangi Beach. It is about a 20-30 minute walk from the conference venue following the map below. If anyone needs a ride, Stéphane is driving a van up there with room for 7 others.

The meal is free for registered participants. You are welcome to bring guests at a cost of $70 per person. Please advise David or Stéphane on Monday if you wish to bring a guest or cannot come.



Travel on and off the island

The ferries from Auckland to Waiheke and the buses on the island are run by Fullers.

The nearest bus stops to the conference venue are at the bottom of Fourth Ave. There are also stops at the bottom of Sixth and Seventh Aves.

Ferry and Onetangi bus timetables are overleaf. Full details of timetables and fares are at `http://www.fullers.co.nz/`. Note that Thursday 6th February is a public holiday so transport will run according to the Sunday timetable that day.

Onetangi bus timetable

Onetangi • Ostend • Surfdale • Blackpool to Matiatia Ferry Terminal

	Route	Notes	Onetangi	Ostend	Surfdale	Blackpool	Oneroa	Matiatia	Ferry Departs	Arrives Auckland
Monday to Friday									1,1a	3 4
AM	3		**5.20**	5.30	5.45	5.48	5.50	5.55	**6.05**	6.40
	1a	JW, #	**6.00**	6.10	6.18	6.22	6.25	6.30	**6.40**	7.15
	4		**6.05**	6.15	6.19	-	6.23	6.28	**6.40**	7.15
	1a	JW, #	**6.35**	6.45	6.52	6.56	6.59	7.04	**7.15**	7.50
	4		**6.40**	6.50	6.55	-	7.00	7.05	**7.15**	7.50
	1a	JW, #	**7.15**	7.25	7.33	7.38	7.40	7.45	**8.00**	8.35
	4		**7.25**	7.35	7.40	-	7.44	7.47	**8.00**	8.35
	1	JW	-	**8.30**	8.38	8.42	8.45	8.50	**9.00**	9.35
	4		**8.25**	8.35	8.40	-	8.44	8.47	**9.00**	9.35
	1	JW, D	**9.00**	9.10	9.18	9.22	9.25	9.30	**10.00**	10.40
	1	JW	**10.00**	10.10	10.18	10.22	10.25	10.35	**11.00**	11.35
	1	JW, D	**11.00**	11.10	11.18	11.22	11.25	11.30	**12.00**	12.40
PM	1	JW	**12.00**	12.10	12.18	12.22	12.25	12.30	**1.00**	1.35
	1	D	**1.00**	1.10	1.18	1.22	1.25	1.30	**2.00**	2.40
	1		**2.00**	2.10	2.18	2.22	2.25	2.30	**3.00**	3.35
	1	D	**3.00**	3.10	3.18	3.22	3.25	3.30	**4.00**	4.40
	1		**4.00**	4.10	4.18	4.22	4.25	4.30	**4.45**	5.20
	1		-	**5.00**	5.08	5.12	5.15	5.20	**5.40**	6.15
	4		**5.00**	5.10	5.15	-	5.19	5.22	**5.40**	6.15
	1		-	**5.45**	5.53	5.57	6.00	6.05	**6.15**	6.50
	4		**5.40**	5.50	5.55	-	5.59	6.08	**6.15**	6.50
	1		**6.35**	6.45	6.53	6.57	7.00	7.05	**7.15**	7.50
	1		**7.15**	7.25	7.33	7.37	7.40	7.45	**8.00**	8.35
	1		**8.45**	8.55	9.03	9.07	9.10	9.15	**9.30**	10.05
	1		**10.15**	10.25	10.33	10.37	10.40	10.45	**11.00**	11.35
	3		**11.45**	11.55	12.10	12.13	12.16	12.21	**12.30**	1.05
Saturday										1 3
AM	3		**6.35**	6.45	7.00	7.03	7.06	7.11	**7.15**	7.50
	3		**7.10**	7.20	7.35	7.38	7.41	7.46	**8.00**	8.40
	1	JW	**8.20**	8.30	8.38	8.42	8.45	8.50	**9.00**	9.35
	1	JW, D	**9.00**	9.10	9.18	9.22	9.25	9.30	**10.00**	10.40
	1	JW	**10.00**	10.10	10.18	10.22	10.25	10.30	**11.00**	11.35
	1	JW, D	**11.00**	11.10	11.18	11.22	11.25	11.30	**12.00**	12.40
PM	1	JW	**12.00**	12.10	12.18	12.22	12.25	12.30	**1.00**	1.35
	1	D	**1.00**	1.10	1.18	1.22	1.25	1.30	**2.00**	2.40
	1		**2.00**	2.10	2.18	2.22	2.25	2.30	**3.00**	3.35
	1	D	**3.00**	3.10	3.18	3.22	3.25	3.30	**4.00**	4.40
	1		**4.00**	4.10	4.18	4.22	4.25	4.30	**5.00**	5.35
	1	D	**5.00**	5.10	5.18	5.22	5.25	5.30	**6.00**	6.40
	1		**6.30**	6.40	6.45	6.49	6.52	7.00	**7.15**	7.50
	1		**7.25**	7.35	7.40	7.45	7.47	7.50	**8.00**	8.40
	1		**8.45**	8.55	9.05	9.10	9.12	9.17	**9.30**	10.05
	1		**10.15**	10.25	10.33	10.37	10.40	10.45	**11.00**	11.35
	3		**11.45**	11.55	12.10	12.13	12.16	12.21	**12.30**	1.05
Sunday and Public Holidays										1 3
AM	3		**7.10**	7.20	7.35	7.38	7.41	7.46	**8.00**	8.40
	1	JW	**8.20**	8.30	8.38	8.42	8.45	8.50	**9.00**	9.35
	1	JW, D	**9.00**	9.10	9.18	9.22	9.25	9.30	**10.00**	10.40
	1	JW	**10.00**	10.10	10.18	10.22	10.25	10.30	**11.00**	11.35
	1	JW, D	**11.00**	11.10	11.18	11.22	11.25	11.30	**12.00**	12.40
PM	1	JW	**12.00**	12.10	12.18	12.22	12.25	12.30	**1.00**	1.35
	1	D	**1.00**	1.10	1.18	1.22	1.25	1.30	**2.00**	2.40
	1		**2.00**	2.10	2.18	2.22	2.25	2.30	**3.00**	3.35
	1	D	**3.00**	3.10	3.18	3.22	3.25	3.30	**4.00**	4.40
	1		**4.00**	4.10	4.18	4.22	4.25	4.30	**5.00**	5.35
	1	D	**5.00**	5.10	5.18	5.22	5.25	5.30	**6.00**	6.40
	1		**6.00**	6.10	6.18	6.22	6.25	6.30	**6.45**	7.20
	3		**7.35**	7.42	7.57	8.00	8.02	8.05	**8.15**	8.50
	3		**9.05**	9.12	9.27	9.30	9.32	9.35	**9.45**	10.20
	1		**10.40**	10.50	10.56	11.01	11.04	11.06	**11.15**	11.50

Explanation of notes used

JW = these services travel via Jellicoe Parade and Wellington Rd
D = ferry travels via Devonport
# = route 1a starts from Waiheke Road by Belle Terrace, then Onetangi Rd to Sea View Rd. Does not travel to Onetangi Bay

Bus service provided by Waiheke Bus Company
Ferry service provided by Fullers

Matiatia Ferry Terminal to Oneroa • Blackpool • Surfdale • Ostend • Onetangi

	Route	Notes	Ferry from Auckland	Arrives Waiheke	Matiatia	Oneroa	Blackpool	Surfdale	Ostend	Onetangi
Monday to Friday									1, 1a 3	4
AM	3		-	-	**6.02**	6.05	6.07	6.10	6.20	6.30
	1		**6.00**	6.35	**6.35**	6.38	6.40	6.45	6.55	7.05
	4		**6.00**	6.35	**6.35**	6.38	6.40	6.45	6.55	7.05
	3	DO	**6.00**	6.35	**6.45**	6.48	6.50	6.55	7.05	7.15
	1		-	-	**7.10**	7.13	7.15	7.20	7.30	7.40
	1		**7.20**	7.55	**7.55**	7.58	8.00	8.05	8.15	8.25
	1		**8.15**	8.50	**8.55**	8.58	9.00	9.05	9.15	9.25
	1	D	**9.00**	9.40	**9.45**	9.48	9.50	9.55	10.05	10.15
	1		**10.00**	10.35	**10.45**	10.48	10.50	10.55	11.05	11.15
	1	D	**11.00**	11.40	**11.45**	11.48	11.50	11.55	12.05	12.15
PM	1		**12.00**	12.35	**12.45**	12.48	12.50	12.55	1.05	1.15
	1	D	**1.00**	1.40	**1.45**	1.48	1.50	1.55	2.05	2.15
	1		**2.00**	2.35	**2.45**	2.48	2.50	2.55	3.05	3.15
	1	D	**3.00**	3.40	**3.45**	3.48	3.50	3.55	4.05	4.15
	4		**4.00**	4.35	**4.40**	4.43	-	4.48	4.55	5.05
	1a	#	**4.00**	4.35	**4.40**	4.43	4.45	4.50	5.00	5.15
	4		**5.00**	5.35	**5.40**	5.43	-	5.48	5.55	6.05
	1a	#	**5.00**	5.35	**5.40**	5.43	5.45	5.50	6.00	6.15
	4		**5.30**	6.05	**6.10**	6.13	-	6.18	6.25	6.35
	1a	#	**5.30**	6.05	**6.10**	6.13	6.15	6.20	6.30	6.40
	4		**6.30**	7.05	**7.10**	7.13	-	7.18	7.25	7.35
	1a	#	**6.30**	7.05	**7.10**	7.13	7.15	7.20	7.30	7.40
	1		**7.15**	7.50	**7.55**	7.58	8.00	8.05	8.15	8.25
	1		**8.45**	9.20	**9.25**	9.28	9.30	9.35	9.45	9.55
	1		**10.15**	10.50	**10.55**	10.58	11.00	11.05	11.15	11.25
	3		**11.45**	12.20	**12.25**	12.28	12.30	12.33	12.48	12.58
Saturday									1 3	
AM			**6.15**	6.50	-	-	-	-	-	-
	1	D	**7.00**	7.40	**7.45**	7.48	7.50	7.55	8.02	8.12
	1		**8.15**	8.50	**8.55**	8.58	9.00	9.05	9.15	9.25
	1	D	**9.00**	9.40	**9.45**	9.48	9.50	9.55	10.05	10.15
	1		**10.00**	10.35	**10.45**	10.48	10.50	10.55	11.05	11.15
	1	D	**11.00**	11.40	**11.45**	11.48	11.50	11.55	12.05	12.15
PM	1		**12.00**	12.35	**12.45**	12.48	12.50	12.55	1.05	1.15
	1	D	**1.00**	1.40	**1.45**	1.48	1.50	1.55	2.05	2.15
	1		**2.00**	2.35	**2.45**	2.48	2.50	2.55	3.05	3.15
	1	D	**3.00**	3.40	**3.45**	3.48	3.50	3.55	4.05	4.15
	1		**4.00**	4.35	**4.45**	4.48	4.50	4.55	5.05	5.15
	1	D	**5.00**	5.40	**5.40**	5.43	5.45	5.50	6.00	6.10
	1		**6.00**	6.35	**6.40**	6.43	6.45	6.50	7.00	7.10
	1		**7.30**	8.05	**8.10**	8.13	8.16	8.21	8.27	8.37
	1		**9.00**	9.35	**9.40**	9.43	9.46	9.51	9.57	10.07
	1		**10.30**	11.05	**11.10**	11.13	11.16	11.21	11.27	11.37
Sunday and Public Holidays										1
AM	1	D	**7.00**	7.40	**7.45**	7.48	7.50	7.55	8.02	8.12
	1		**8.15**	8.50	**8.55**	8.58	9.00	9.05	9.15	9.25
	1	D	**9.00**	9.40	**9.45**	9.48	9.50	9.55	10.05	10.15
	1		**10.00**	10.35	**10.45**	10.48	10.50	10.55	11.05	11.15
	1	D	**11.00**	11.40	**11.45**	11.48	11.50	11.55	12.05	12.15
PM	1		**12.00**	12.35	**12.45**	12.48	12.50	12.55	1.05	1.15
	1	D	**1.00**	1.40	**1.45**	1.48	1.50	1.55	2.05	2.15
	1		**2.00**	2.35	**2.45**	2.48	2.50	2.55	3.05	3.15
	1	D	**3.00**	3.40	**3.45**	3.48	3.50	3.55	4.05	4.15
	1		**4.00**	4.35	**4.45**	4.48	4.50	4.55	5.05	5.15
	1	D	**5.00**	5.40	**5.40**	5.43	5.45	5.50	6.00	6.10
	1		**6.00**	6.35	**6.40**	6.43	6.45	6.50	7.00	7.10
	1		**7.30**	8.05	**8.10**	8.13	8.16	8.21	8.27	8.37
	1		**9.00**	9.35	**9.40**	9.43	9.46	9.51	9.57	10.07
	1		**10.30**	11.05	**11.10**	11.13	11.16	11.21	11.27	11.37

Explanation of notes used

DO = bus drops off only, no pickups
D = ferry travels via Devonport
# = Route 1a terminates at the corner of Waiheke Road and Belle Terrace, not Onetangi Bay

Route 1 from Matiatia to Onetangi will travel via Wellington Rd and Jellicoe Parade on request only

Bus service provided by Waiheke Bus Company
Ferry service provided by Fullers

DEPARTING AUCKLAND
Ferry Terminal, 99 Quay Street - To Matiatia Wharf

MON TO FRI	SATURDAY	SUN & PUBLIC HOLIDAYS
5.20 am‡	–	–
6.00 am	6.15 am	–
7.20 am	7.00 am*	7.00 am*
8.15 am	8.15 am	8.15 am
9.00 am*	9.00 am*	9.00 am*
10.00 am	10.00 am	10.00 am
11.00 am*	11.00 am*	11.00 am*
12.00 noon	12.00 noon	12.00 noon
1.00 pm*	1.00 pm*	1.00 pm*
2.00 pm	2.00 pm	2.00 pm
3.00 pm*	3.00 pm*	3.00 pm*
4.00 pm	4.00 pm	4.00 pm
5.00 pm	5.00 pm*	5.00 pm*
5.30 pm	–	–
6.30 pm**	6.00 pm	6.00 pm
7.15 pm	7.15 pm	7.30 pm
8.45 pm	8.45 pm	9.00 pm+
10.15 pm	10.15 pm	10.30pm+
11.45 pm	11.45 pm	-

*Sailing time is approximately 10 minutes longer than other sailings.

**Goes via Devonport on Fridays only from 1 November 2013 until 4 April 2014. Sailing time is approximately 10 minutes longer than other sailings.

+ Sailing operates from 1 December 2013 until 21 April 2014.

Auckland to Waiheke Island return tickets do not include a stopover at Devonport. If you disembark the vessel at Devonport, an additional fare will apply for travel back to Auckland.

‡ =This is an unscheduled departure and depending on vessel, may depart earlier than time shown.

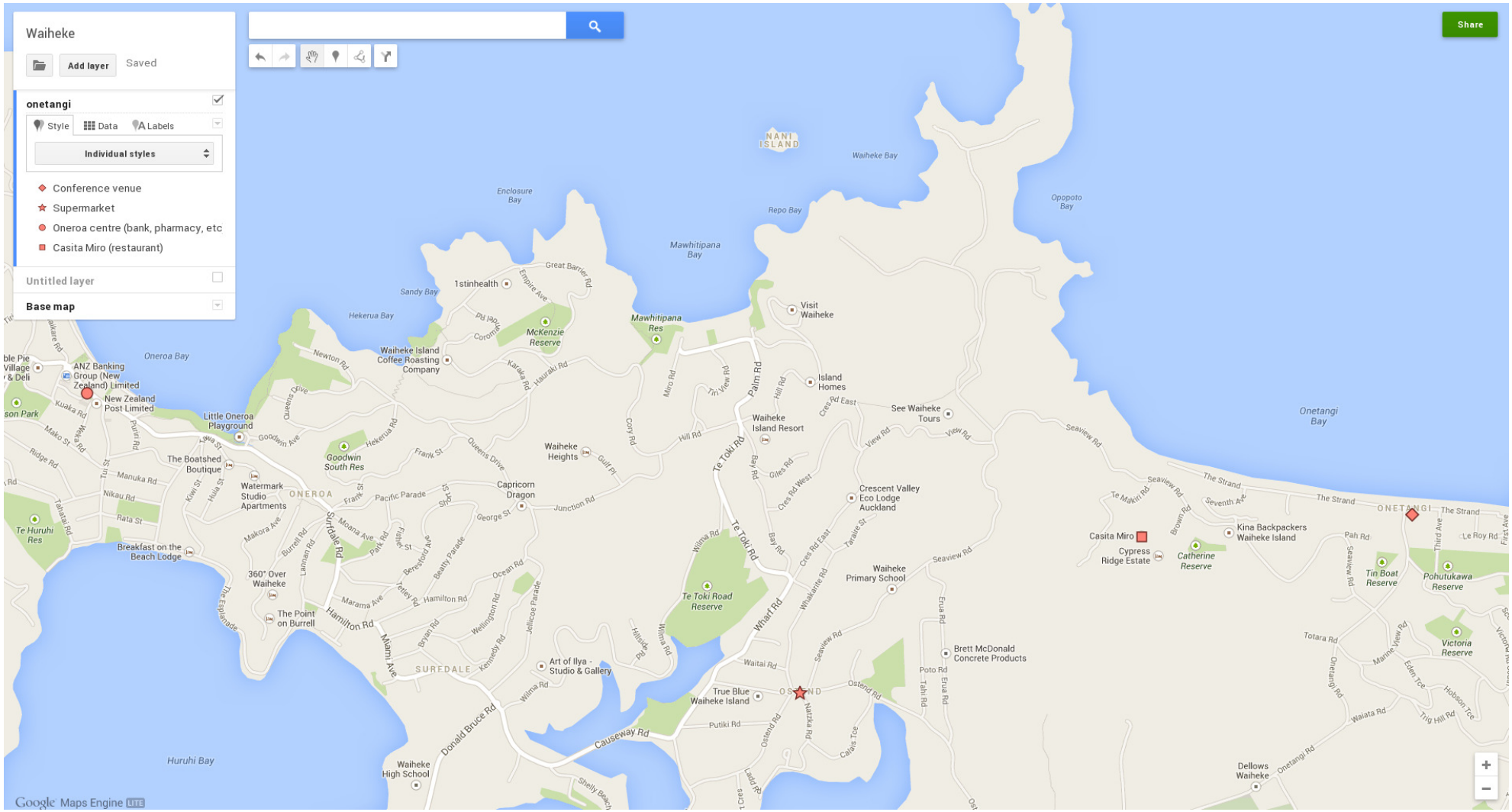
DEPART WAIHEKE ISLAND
Matiatia Wharf - To Auckland

MON TO FRI	SATURDAY	SUN & PUBLIC HOLIDAYS
6.05 am	–	–
6.40 am	–	–
7.15 am	7.15 am	–
8.00 am	8.00 am	8.00 am
9.00 am	9.00 am	9.00 am
10.00 am*	10.00 am*	10.00 am*
11.00 am	11.00 am	11.00 am
12.00 noon*	12.00 noon*	12.00 noon*
1.00 pm	1.00 pm	1.00 pm
2.00 pm*	2.00 pm*	2.00 pm*
3.00 pm	3.00 pm	3.00 pm
4.00 pm*	4.00 pm*	4.00 pm*
4.45 pm	–	–
5.40 pm	5.00 pm	5.00 pm
6.15 pm	6.00 pm*	6.00 pm*
7.15 pm	7.15 pm	6.45 pm
8.00 pm	8.00 pm	8.15 pm
9.30 pm	9.30 pm	9.45pm+
11.00 pm	11.00 pm	11.15 pm+
12.30 am	12.30 am	–

* Sailing time is approximately 10 minutes longer than other sailings.

+ Sailing operates from 1 December 2013 until 21 April 2014.

PHONE 09 367 9111 or 0800 FULLERS
WWW.FULLERS.CO.NZ


Waiheke
Add layer
Saved
onetangi
Style
Data
A Labels
Individual styles
Conference venue
Supermarket
Oneroa centre (bank, pharmacy, etc
Casita Miro (restaurant)
Untitled layer
Base map
Share
NANI ISLAND
Waiheke Bay
Enclosure Bay
Repo Bay
Opopoto Bay
Mawhitipana Bay
Onetangi Bay
Sandy Bay
Hekerua Bay
Oneroa Bay
Huruhi Bay
1stinhealth
McKenzie Reserve
Mawhitipana Res
Visit Waiheke
Waiheke Island Coffee Roasting Company
ANZ Banking Group (New Zealand) Limited
New Zealand Post Limited
Little Oneroa Playground
Goodwin South Res
Island Homes
Waiheke Island Resort
See Waiheke Tours
Waiheke Heights
Capricorn Dragon
Crescent Valley Eco Lodge Auckland
The Boatshed Boutique
Watermark Studio Apartments
ONEROA
Breakfast on the Beach Lodge
360° Over Waiheke
The Point on Burrell
Te Huruhi Res
Te Toki Road Reserve
Waiheke Primary School
Casita Miro
Cypress Ridge Estate
Catherine Reserve
Kina Backpackers Waiheke Island
Tin Boat Reserve
Pohutukawa Reserve
Victoria Reserve
Brett McDonald Concrete Products
Art of Ilya - Studio & Gallery
SURFDALE
OSTEND
True Blue Waiheke Island
Waiheke High School
Dellows Waiheke
Te Toki Rd
Wharf Rd
Palm Rd
Hamilton Rd
Causeway Rd
Donald Bruce Rd
Seaview Rd
The Strand
ONETANGI
Google Maps Engine LITE