DOOM - 01

12-16th Feb 2001

Skotel, Whakapapa Village

ABSTRACTS

IMBS - Massey University

BRC - University of Canterbury

supported by the New Zealand Marsden Fund

Monday

Reconstructing phylogenies from characters on a large state space (40min)

Mike Steel, Biomathematics Research Centre, University of Canterbury,
Christchurch, New Zealand
tel: 64 3 364 2987, fax: 64 3 3642587

One can regard certain types of genomic data as qualitative characters on a very large (but finite) state space. We will describe conditions under which simple methods will correctly reconstruct an underlying tree (assuming some conditions concerning the evolution of the characters), and some new results concerning how many characters are required to uniquely determine an underlying tree. We will see that some results contrast sharply with the binary character (or 4-state DNA character) setting.

The geometry of the space of phylogenetic trees (40min)

Susan Holmes, INRA, Montpellier, France and Department of Statistics,
Stanford University, Stanford, CA 94305,USA
address: Statistics Dept, Sequoia Hall, Stanford University, CA 94305 USA
tel: 1 650 725 1925,fax: 1 650 725 8977

{A Geometrical Perspective for Phylogenetic Trees} Joint work with Louis J. Billera and Karen Vogtmann (both of Cornell)

I will present a geometric model of the space of all phylogenetic trees. The geometry of the space gives a way of measuring a distance between phylogenetic trees as well as a way of `averaging' or `combining' several trees whose leaves are identicial. The convex hull of a set of trees can also be defined, thus making possible a "peeling" procedure which finds the most central trees in a set.

Up to now many confidence statements have had to be reduced to the existence or nonexistence of particular clades, this geometric model of tree space provides a setting in which questions of more general multidimensional aspects of the trees may be explored.

For example, this geometry provides a justification for disregarding portions of a collection of trees that agree, thus simplifying the space in which comparisons are to be made, in some sense a way of projecting a collection of trees into a lower dimensional space is provided.

It turns out that the space has a non positive curvature, this just means that triangles are thinner than in Euclidean space, this strengthens the power of averaging.

Calculating Invariants for the General Model (40min)

Thomas Hagedorn, Centre de recherches mathématiques, Université de Montréal, CP 6128 succursale Centre-Ville, Montréal, Québec H3C 3J7.

The phylogenetic invariants for the Jukes-Cantor, Kimura 2-parameter, and Kimura 3-parameter model of sequence evolution have been determined by [1,2,4]. In this paper, we begin the study of phylogenetic invariants for the general model, in which no assumption is made upon the structure of the mutation matrices (in other words, they can be any matrix) and the general stochastic model (in which the columns, alternatively rows, sum to one).

Let $I_M(T,k)$ be the ideal of invariants for a tree *T* for a sequence composed of *k* different bases and a class of Markov models *M*. We use a combinatorial approach to calculate the low-degree invariants in $I_M(T,k)$ for trees *T* with *n* taxa for small *n.* We can show that the homogeneous invariants for the general, stochastic Markov model are the same as those for the general model. We show that homogeneous invariants of degree *d* can only arise when d >= k+1. We also calculate the invariants of low-degree when *k*=2, 3, 4 for the general model. When k=2, we provide a theoretical explanation for the cubic invariants found by [3] for the four-leaf tree. We can also show that these cubic invariants generate the quartic and quintic invariants when *k*=2 for the four-leaf tree for the general model. When *k*=4, we determine all the quintic invariants when *n*=3 and can construct a large number, (19,079,424), of quintic invariants when *n*=4 for the general model.

[1] Evans, S. and Speed, T., Invariants of Some Probability Models Used in Phylogenetic Inference, *The Annals of Statistics* (1993), 21(1), 355--377.

[2] Evans, S. and Zhou, X., Constructing and Counting Phylogenetic Invariants. *J. Comp. Biol*., (1998), **5**, 713--724.

[3] Ferretti, V. and Sankoff, D., Phylogenetic Invariants for More General Evolutionary Models. *J. Theor. Biol.* (1995), **173** (2) 147--162.

[4] Steel, M.A., Szekely, L.A., Erdos, P.L., and Waddell, P., A Complete Family of Phylogenetic Invariants for Any Number of Taxa under Kimura's 3ST Model, *N.Z. J. of Botany* (1993), **31**, 289--296.

Appropriate Likelihood Ratio Tests and Marginal Distributions for Evolutionary Tree Models with Constraints on Parameters (20min)

Rissa Ota*, Peter J. Waddell,, Masami Hasegawa, Hidetoshi Shimodaira and Hirohisa Kishino¤
*The Graduate University for Advanced Studies and The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo, Japan; Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand; and ¤Department of Agriculture and Life Sciences, University of Tokyo, Tokyo, Japan

We show how to make appropriate likelihood ratio tests for evolutionary tree models when parameters such as edge (internodes or branches) lengths have nonnegativity constraints. In such cases, under the null model of an edge length being zero, the marginal distribution of this parameter is proven to be a "half-normal", that is, 50% zero values and 50% the positive half of a normal distribution. Other constrained parameters, such as the proportion of invariant sites, give similar results. To make likelihood ratio tests between nested models, e.g., H0: homogeneous site rates, and H1: site rates follow a gamma distribution with variance 1/k, then asymptotically as sequence length increases, the distribution under H0 becomes a mixture of distributions, in this case 50% 0, and 50% 1 (where the subscript denotes degrees of freedom, i.e., not the usually assumed 100% 1; which leads to a conservative test). Such mixtures are sometimes called distributions. Simulations show that even with sequences as short as 125 sites, some parameters, including the proportion of invariant sites, fit asymptotic distributions closely.

Some recent results in phylogenetic combinatorics (20min)

Vincent Moulton, FMI (Physics and Mathematics Department), Mid Sweden University,
S 851-70 Sundsvall, Sweden.
FAX: +46 60 148875, Tel: +46 60 148741

Phylogenetic combinatorics is the term recently coined for topics of a combinatorial nature that are motivated to a large extent by problems arising in phylogenetic analysis. We will discuss some recent results with this flavor and also present some open problems. If we are lucky, we will also point out some connections with the motivating biological problems (no promises!).

Assembly of the Human Genome (40min)

Daniel H. Huson, Informatics Research
Celera Genomics Corp, Rockville MD USA

Current sequencing technology can only determine about 600 consecutive base pairs of DNA in a single read. Hence, to obtain the sequence of a larger stretch of DNA, one must randomly chop it up into small pieces, which are then sequenced and "assembled" (so-called shotgun sequencing). One strategy to sequence the human genome has been to sequence and assemble 150 Kps at a time, in a slow-but-sure approach. Celera Genomics favors "Whole Genome Shotgun Assembly" (using a "double-barreled" shotgun), relying heavily on bioinformatics to assemble the whole genome in-silico in a short time-frame.

Endogenous retroviruses: data mining, phylogenetics and the risk of cross-species transmission (20min)

Lindell Bromham, Dept. of Zoology, University of Queensland, Brisbane 4072, Australia.
Fax: 61-7-3365 1655,

Analysis of genomic databases can lead to the discovery and characterization of novel retroviruses and transposable elements, providing insight into the role of retroelements in host genome evolution, and allowing exploration of cross-species transmission of retroviruses.

Searching genomes for RNA world relics (20min)

Anthony Poole, Institute of Molecular BioSciences, Massey University, New Zealand

The entire DNA sequence, or genome sequence, is now known for numerous organisms. These consist of strings of the letters A, G, C & T, and these strings can be billions of letters long. The task that the biological community faces is how to deal with this data, and how to extract meaningful information (i.e. the location and nature of genes) from it. Broadly speaking, genes come in two flavours: protein and RNA. The information for making a protein is translated from this four letter code and turned into the language of proteins: amino acids, of which there are 20. Knowing this code helps immensely in identification of protein-coding genes. RNA genes on the other hand are a lot harder to find since they are encoded only in the four letter code. Furthermore, RNAs related in evolution tend not to be similar in their sequence, rather, it is their active biological structure that appears to be most conserved. I shall describe the nature of the problem of RNA gene finding, focusing on our group's work on one particular class, the snoRNAs.

ncRNA Gene Location Based on RNA Secondary Structure (20min)

Paul Gardner, Institute of Molecular BioSciences, Massey University
New Zealand

Non-coding RNAs (ncRNAs) are believed by many to be relics of ancient RNA based life-forms that preceded the modern day DNA/protein based life that currently infest Earth. The genes coding for ncRNAs do not code for proteins as most genes do. Instead they produce functional RNA. The most commonly known ncRNAs are ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and small nuclear RNAs (snRNAs). However new and exciting ncRNAs are still being discovered such as Rnase P RNA, telomerase RNA, meiRNA, bacterial tmRNA and bacteriophage f 29 RNA.

The usual methods for locating protein genes fail when applied to ncRNAs. The statistically based similarity searching protein methods (e.g. BLAST and FASTA) rely heavily upon sequence conservation between sequences from the same family. These methods fail because many ncRNAs conserve a consensus secondary structure more than they conserve a primary sequence. Therefore sequence alignment methods are ineffective when applied to ncRNAs.

This talk will include a general introduction to RNA secondary structure and its prediction, and possible methods for locating ncRNAs within a newly sequenced genome.

Overview of the Cebl application environment underlying the Pebble (20min)

Matthew Goode, University of Auckland
Phone: 373-7599 x 7466, 021-255-2981
Web: http://www.cs.auckland.ac.nz/~mgoo003

PEBBLE (Phylogenetics, Evolutionary Biology and Bioinformatics in a modular Environment) is an integrated application for evolutionary analysis. The application is built around the vCebl application framework, which through the use of a custom programming language allows users to build custom applications, simulations and analyses. vCebl handles GUI components implicitly, so users are free to concentrate on the development of customized modules. In this talk, I will describe the basic architecture, the standard GUI front end, the ease of extensibility, and the vCebl langauge. We present specific examples using PEBBLE, and illustrate the advantages of such the modular approach for computational evolutionary biology. Finally, I discuss plans for future versions of vCebl/PEBBLE.

Tuesday

Pattern of nucleotide substitution and rate heterogeneity in the HVR-I of the mitochondrial control region in large white-headed gulls (*Larus argentatus - fuscus - cachinnans* complex) (40min)

Dorit Liebers, Max-Planck-Institute for Evolutionary Anthropology, Inselstr. 22,
D-04103 Leipzig, Germany

This study on circumpolar distributed large white-headed gulls provides a comprehensive survey of the complex pattern of nucleotide substitution in the hypervariable domain I (HVR-I) of the mitochondrial control region. Among a total of 1217 sequences of 430 bp length, 196 different lineages were analysed using a complex substitution model. We used a maximum-likelihood method to assign relatively rates to each site of the sequence. Estimation was based on randomly selected subsets of sequences.As in other birds studies, HVR-I contains putative TAS elements and highly conserved sequence blocks. Cloverleaf-like structures associated with the TAS element may be involved with the stop of the control region synthesis. Understanding the evolution of control region sequences is of fundamental importance for the study of speciation.

In large gulls the rate pattern vary extensively among sites and sequences of HVR-I are extremely useful in recovering intraspecific phylogeographic splits within this recently diverged group of birds. Contrary to current classification, a basic split was revealed between an Atlantic-Mediterranean clade (*atlantis, michahellis, armenicus*) and a NW Palearctic/Central Asian clade (*cachinnans, barabensis, mongolicus, argentatus-fuscus*-group). There was almost no mitochondrial gene flow between these groups, although they are in geographic contact in two areas (eastern North Atlantic, Black Sea). Also within each of the two major groups, there was strong phylogeographic structure with clear gene flow barriers between some neighbouring taxa (e.g. *cachinanns* vs. *armenicus*), but also a case of poor genetic differentiation between phenotypically distinct forms (*barabensis* vs. *heuglini*). At the subspecies level, current taxonomy corresponded well to molecular genetic structure: over 80% of the molecular genetic variance was partitioned among six groups of taxa. Within-taxon haplotype phylogenies and mismatch distributions revealed quite different demographic histories: the Ponto-Aralo-Caspian form (*cachinnans*) and the NE Atlantic form (*atlantis*) represent ancient lineages with large long-term population sizes, inland forms like *barabensis* (C Kazakhstan) and *mongolicus* (Mongolia) apparently result from very recent colonization events, while *armenicus* (Asia Minor) passed through a population bottleneck.

Tempo and mode of molecular evolution in island radiations (20min)

Meg Woolfit, Department of Zoology and Entomology, University of Queensland
St Lucia, 4072, Australia
tel: 61 7 3365 3753, fax: 61 7 3365 1655

The discrepancy between molecular and fossil dates for a number of explosive radiations (e.g. the Cambrian explosion of metazoan taxa, and the Tertiary radiations of birds and mammals) has led to suggestions that the molecular clock runs fast during adaptive radiations. This must be investigated before molecular dates for rapid adaptive radiations, from ancient divergences to recent epidemics, can be trusted. I am using island radiations to explore the effect of rate of adaptation and diversification on the patterns and rates of molecular evolution. Islands are ideal locations to witness explosive radiations in progress, demonstrating many evolutionary processes that could affect rate of molecular evolution: genetic bottlenecks as populations are initiated from a small number of colonists, rapid rate of phenotypic evolution as species are released from the constraints of the mainland ecosystem, and novel adaptations as they evolve into a range of new niches. Furthermore, many island radiations have been well-studied, with close mainland relatives identified and molecular sequence data available for both island and mainland taxa. Island radiations can act as model systems, aiding the interpretation of adaptive radiations ranging from ancient evolutionary events to recent epidemics.

Molecular Clocks, Geology, and the origins of New Zealand Cicadas (20min)

CHRIS SIMON[*1,2], THOMAS BUCKLEY[2,3], KENT HOLSINGER[1], AND PETER ARENSBURGER[1]. Ecology and Evolutionary Biology, U-3043, University of Connecticut, Storrs, CT 06269[1], School of Biological Sciences, PO Box 600, Victoria University of Wellington, Wellington, NZ[2], Current address: Department of Biology, Box 90338, Duke University, Durham, NC 27708[3].

New Bayesian dating techniques show promise for overcoming one of the major problems with molecular clock-estimates: DNA often does not evolve in a clock-like manner. New Zealand cicadas provide an unusually good system for the application of these methods because well-supported molecular trees exist for the genera and species, the Cenozoic level of divergence is relatively shallow making the estimation of genetic distances easier, landscape and climatic changes in NZ are dramatic and well-studied, two closely related and diverse NZ cicada genera can be calibrated independently to give estimates of the date of arrival of their ancestors, and finally a third, independent, geological calibration can be used to lend strong support to the result. Our analyses suggest that *Kikihia* and *Maoricicada* diverged from each other approximately 9-10 +/- 3 MYA, that they arrived in NZ shortly before this, and that the closest relatives of these NZ cicada genera are New Caledonian. This fits well with recent data suggesting an island archipelago connection between NZ and NC that was broken in the Mid-Miocene (14-20 MYA). This talk adds data for the genus *Kikihia* to complement the *Maoricicada* data presented at Ruahine 2000 last year, when the project was in its early stages.

A 'total-evidence' approach to knawel (*Scleranthus*) phylogeny and biogeography (20min)

Rob Smissen, Manaaki Whenua - LandCare Research, PO Box 69, Lincoln 8152, New Zealand
Telephone +64 3 325 6701 , Fax +64 3 325 2418
www.landcare.cri.nz

Scleranthus is a genus of about 12 species of plants in the family Caryophyllaceae. The native distribution of these species shows a striking disjunction between a temperate Eurasia-Mediterranean group and an Australia-PNG-New Zealand group. European centered species have comparatively large flowers and high numbers of flowers in each inflorescence. Australasian species include similar large-flowered species, as well as species with very reduced flowers and only one or two flowers in each inflorescence. Morphological characters and ITS sequences both suggest that Scleranthus is monophyletic and provide weak support for northern and southern hemisphere clades. However, the two data sets are incongruent and support alternative relationships among southern hemisphere species. Despite incongruence, combined parsimony analysis produces a well-resolved tree in which species with reduced inflorescences and flowers have evolved only once. Sequence divergence between species is low, suggesting a relatively recent radiation of extant species. Therefore, current distribution of Scleranthus seems to be the result of long distance dispersal.

The inference of step-wise changes in substitution rates using serial sequence samples (20min)

Alexei Drummond, Roald Forsberg and Allen Rodrigo, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand
Tel: 64-9-3737 599 Fax: 64-9-3737 414

Genealogies of molecular sequences frequently do not conform to a strict molecular clock even when all the sequences are from the same population. Here we present two new methods for estimating step-wise changes in substitution rates when serially sampled molecular sequences are available. These models are members of a class we name varying clock models. The first model presented in this paper is an extension of the genealogy-based maximum likelihood framework for the estimation of substitution rate that allows for step-wise changes in substitution rate. The second is a new parameterization of the pairwise divergence least-squares procedure used in sUPGMA to afford the same relaxation. The utility of these methods is demonstrated on a genealogy of HIV sequences obtained at five different sampling times from a single patient over a period of 34 months. The analysis demonstrates that the substitution rate of the viral population is significantly reduced following treatment. Thus a strict molecular clock hypothesis of evolution can be rejected.

Median Networks: A visual representation of ancient Adelie Penguin DNA (20min)

Barbara Holland, David Lambert (Massey University) and Peter Ritchie (Massey University)
Institute of Fundamental Sciences, Massey University

Massey ecologists David Lambert and Peter Ritche have gathered a data set, consisting of mitochondrial DNA sequences from over 300 Adelie penguins. In addition to this they succeeded in recovering DNA from over 70 ancient bones preserved beneath the Antarctic permafrost. This provides a unique data set that can be used to directly measure the rate of mtDNA evolution.

Median Networks are a useful tool for visualising the structure in mtDNA data sets. In this talk I will explain how a median network was constructed for the Adelie data set and how how it can be used to measure the rate of sequence evolution.

Alpine buttercups (20min)

Peter Lockhart Institute of Molecular BioSciences,
Massey University, Palmerston North,
tel 6463569099 extn 7053, fax 64 6 350 5688

The New Zealand alpine buttercups appear to have radiated in the South Island mounatins as they formed during the last 5 My. More recently – during the Pleistocene, two species migrated north to the Volcanic Plateau to be sure they be would on time for the Doom meeting. Analyses of chloroplast JSA region and nuclear ITS sequences from taxa in the South Island suggest the birth of novel species into fringe habitats. Their ancestors were more widespread species that may have become restricted into different glacial refugia during the Pleistocene. If so, this might be an example of the secondary contact model – an idea which describes the hybrid origin of species as a consequence of glacial – interglacial cycles. I'll show some graphs, pictures of Mike Steel grinning while buttercup hunting and we can also discuss the coding problem for median graphs when there are more than 2 character states in a column.

Climatic variations over 1 million years before present from an astrophysical viewpoint: ice ages and global temperatures (20min)

Sverker Edvardsson, Dept. of Physics and Mathematics, Mid Sweden University, S-851 70 Sundsvall, Sweden
Phone: +46 60 148892, Cellphone: +46 708 493943, Fax: +46 60 148875

Celestial mechanical simulations of the whole Solar system, including our moon and the Mars moons - Phobos and Deimos -, are carried out at an almost extreme level of accuracy. It is confirmed that the obliquity (rotation axis) is stabilized by our own moon which otherwise would be chaotic (Laskar). On the other hand, the Mars obliquity exhibits a chaotic behaviour which makes it impossible to simulate further than 2 million years before present. Climatic influences from astrophysical variables for the Earth and Mars are studied and compared, especially for ice volume measurements (sea sediments of the Northern Atlantic) and average temperatures (ice cores at Vostok, Antarctica). Instead of using an average theory such as that of Milankowitch, our approach focuses on the instantaneous Solar radiation power at 1 million summers.Comparison is then made by differentiating the ice volume and plotting it together with the radiation power in accordance with the first law of thermodynamics. The results are shown to be in excellent agreement.

Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution (20min)

Alan Cooper, Carlos Lalueza-Fox, Simon Anderson, Jeremy Austin, Andrew Rambaut and Ryk Ward
Henry Wellcome Ancient Biomolecules Centre, Departments of Zoology and Biological Anthropology, University of Oxford, U.K.
Mail: 58 Banbury Rd, Oxford, OX2 6QS
Phone: (44-1865) 271-263/265 Fax: 274-699

The origin of the ratites, large flightless birds from the Southern Hemisphere, along with their flighted sister taxa, the South American tinamous, is central to understanding the role of continental drift in the distributions of modern birds and mammals. Defining the dates of ratite divergences is also critical for determining the age of the modern avian orders. To resolve the ratite phylogeny and provide biogeographical data to examine these issues, we determined the first complete mitochondrial genome sequences of any extinct taxa, two New Zealand moa genera, along with a 1,000 bp sequence from an extinct Madagascan elephant-bird. For comparative data, we also generated 12 Kb of contiguous sequence from the kiwi, cassowary, emu, and two tinamou genera. This large dataset allows statistically precise estimates of molecular divergence dates and these support a Late Cretaceous vicariant speciation of ratite taxa, followed by the subsequent dispersal of the kiwi to New Zealand. This first molecular view of the break-up of Gondwana provides an independent temporal framework for speciation events within other Gondwana biota and can be used to evaluate competing biogeographical hypotheses.

Hybrid Phylogenies (20min)

Charles Semple, Biomathematics Research Centre, University of Canterbury,
Christchurch, New Zealand
tel: 64 3 364 2987, fax: 64 3 3642587

In this talk, we discuss possible models for hybrid phylogenies.

Extensive mitochondrial introgression in a tropical reef fish (*Acanthochromis polyacanthus*) at a contact zone between colour morphs (20min)

Lynne van Herwerden1, Peter J. Doherty2 & M. Julian Caley1
1School of Marine Biology and Aquaculture, James Cook University, Townsville, Queensland, 4811, Australia2Australian Institute of Marine Science, PMB No3 Townsville, Queensland, 4810, Australia

Unlike nearly all tropical reef fishes, *Acanthochromis polyacanthus* lacks a pelagic larval stage. Larvae complete development in demersal eggs after which they hatch and are cared for by both parents until fledging into surrounding reef habitat. The Great Barrier Reef supports three allopatric colour morphs and a range of intermediate colour morphs at contact zones. Extensive "hybridization" between colour morphs results in large populations of intermediate morphs in contact zones. Two contact zones have been identified, one in the northern and another in the central region of the GBR. The "northern" contact zone is associated with reefs on the continental shelf, which are close (about 15 km) to detached reefs that rise from > 800m. The concordance of mitochondrial sequences (Dloop) with colour in the northern contact zone is investigated. Population and phylogenetic analyses of sequences from 300 individuals sampled from 10 reefs (7 shelf and 3 detached reefs) indicate distinct "stocks". The genetic subdivision (estimated using AMOVA), was more closely related to location (continental shelf vs detached reefs (77% of the variation)) than colour. These findings are discussed in the context of previous allozyme studies of *A. polyacanthus* from the central GBR contact zone, where phylogenetic analyses indicated strong genetic structuring based on colour, irrespective of location.

Some Distance and Balance Properties of a Simple Speciation Model (20min)

Andy McKenzie (joint work with Mike Steel), Biomathematics Research Centre
Department of Mathematics and Statistics, University of Canterbury , Christchurch
New Zealand, Phone (03) 366-7001 extn. 7424

We investigate some properties of rooted trees generated under a simple stochastic model of speciation. Firstly, we find the probability distribution for the distance (number of edges) of a randomly chosen species from the root of the tree. Next we find a formula for the mean distance between two randomly chosen species. Lastly,we describe the probability distribution on the depth (number of edges from the root) of the most recent common ancestor of a random subset of k species. We then look at some modifications of the speciation model in which the speciation rate is non-constant, and investigate the effect this has on the `balance' of the tree shapes produced.

Reticulate evolution in reef corals of the genus *Acropora* (20min)

Madeleine JH van Oppen, Biochemistry and Molecular Biology, James Cook University
Townsville Qld 4811, Australia
Tel: +61 7-47816388/4553, fax: +61 7-47251394

Acropora (Scleractinia, Acroporidae) is one of the world's most widespread genera of scleractinian coral, spanning the Indian and Pacific Oceans and the Caribbean Sea. It is also the largest extant reef-building coral genus. Recent revisions of the genus recognise 113 - 180 *Acropora* species. Up to 70 *Acropora* species can occur sympatrically on Indo-Pacific coral reefs and most of these participate in the annual coral mass-spawning event. This creates an unparalleled opportunity for interspecific hybridisation and introgression. Experimental breeding trials and molecular data are consistent with a reticulate evolutionary history of many Indo-Pacific species in the genus. Only three species occur in the Caribbean: *Acropora cervicornis, A. palmata* and *A. prolifera*. Based on overall coral morphology, abundance and distribution patterns, it has been suggested that *A. prolifera* may be a hybrid between *A. cervicornis* and *A. palmata*. Phylogenetic and limited population genetic analysis of nuclear molecular markers support the hypothesis that *A. prolifera* is a hybrid between *A. cervicornis* and *A. palmata*, which backcrosses with the parental species at low frequency.

Wednesday night

Mitochondrial DNA Sequences in Ancient Australians: Implications for Modern Human Origins

Gregory J. Adcock, Elizabeth S. Dennis, Simon Easteal, Gavin A. Huttley, Lars S. Jermiin, W. James Peacock, and Alan Thorne. *PNAS* Jan 16, 2001 98, 537-542, Australian Genomic Information Centre, Building B19, University of Sydney New South Wales 2006, Australia. School of Biological Sciences, Heydon-Lawrence Building A08, University of Sydney New South Wales 2006, Australia, Phone +61 (02) 9351 3717, Fax +61 (02) 9351 4119

DNA from ancient human remains provides new perspectives on the origin of our species and on the relationship between molecular and morphological variation. We report analysis of mtDNA from the remains of 10 ancient Australians. These include the morphologically *gracile* Lake Mungo 3 (LM3, ~ 60 thousand years (ka) BP), and three other *gracile* individuals from Holocene deposits at Willandra Lakes (< 10 ka), all within the skeletal range of living Australians; and six Pleistocene/Early Holocene individuals (15 to < 8 ka) from Kow Swamp with *robust* morphologies outside the skeletal range of contemporary indigenous Australians. LM3 is the oldest (Pleistocene) "anatomically modern" human from whom DNA has been recovered. His mtDNA belonged to a lineage that only survives as a segment inserted into chromosome 11 of the nuclear genome, which is now widespread among human populations. This lineage probably diverged before the most recent common ancestor (MRCA) of contemporary human mitochondrial genomes. This implies that the deepest known mtDNA lineage from an "anatomically modern" human occurred in Australia; analysis restricted to living humans places the deepest branches in East Africa. The other ancient Australian individuals we examined have mtDNA sequences descended from the MRCA of living humans. Our results indicate that "anatomically modern" humans were present in Australia before the complete fixation of the mtDNA lineage now found in all living people. These results are not consistent with a recent migration of "modern humans" out of Africa generating contemporary patterns of mtDNA variation.

Another major challenge to the classical Multiregional Model?

David Penny, Institute of Molecular BioSciences, Massey University, Palmerston North Courier: Science Towers, D5.0 New Zealand, Tel +64 6 350 5033, Fax +64 6 350 5688, Cell-phone 021 550 070

The recent report of ancient human DNA from Australia is a major technical achievement, but it has removed one of the main elements on which the classical multiregional model was based. This was that morphological features seen on fossil remains of humans was evidence for geographic continuity of human populations for up to 500,000 years. This model was in itself a major retraction from earlier ideas that regional continuity had existed back to the Pliocene. The current finding that gracile and robust forms in Australia are similar genetically, means that slight differences in morphology cannot be used as evidence for genetic continuity.

This leaves molecular genetic data, and the high genetic similarity of humans throughout the world puts strong limits on hypotheses, and shifts the focus much closer to the present. Despite thousands of samples, Neanderthal sequences have not been found in modern Europeans. With one possible exception, the recent Australian finds fit predictions from an enhanced Out_of_Africa model with a maritime filter around Australia that reduces the rate of continued migration into Australia.

Thursday

Recovering reticulation in human evolution: trees, split graphs, and Pacific languages (40min)

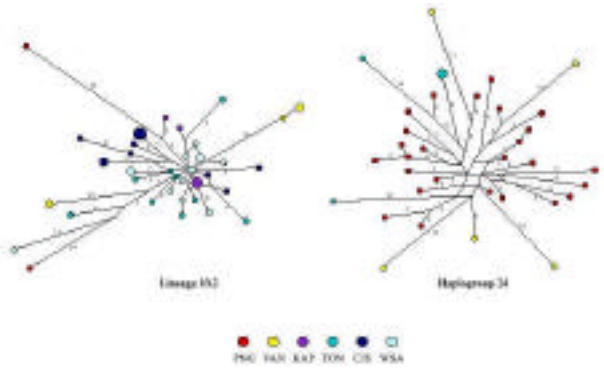
Russell Gray, Department of Psychology, University of Auckland
Auckland , New Zealand
Ph: 64 9 3737 599 ex 8525, Fx: 64 9 3737 450

This talk will discuss the utility of phylogenetic methods for recovering both the tree-like and reticulate signals in linguistic data.

The Y-chromosomal origins of Oceanic-speaking populations (20min)

Matthew E. Hurles, Research Fellow in Population Genetics, McDonald Institute for Archaeological Research , University of Cambridge, Downing Street, Cambridge, CB2 3ER
Tel:01223 339297, Fax:01223 339285

A number of alternative hypotheses seek to explain the origins of the three groups of Pacific populations that speak languages belonging to the Oceanic subfamily of Austronesian languages: Melanesians, Micronesians and Polynesians. In order to test these various hypotheses at the genetic level, we assayed diversity within the non-recombining portion of the Y chromosome which contains within it a simple record of the human past and represents the most informative haplotypic system in the human genome. High-resolution haplotypes combining biallelic, microsatellite and minisatellite markers were generated for 390 Y chromosomes from 17 Austronesian-speaking populations in Southeast Asia and the Pacific. Nineteen paternal lineages were defined and a Bayesian analysis of coalescent simulations performed upon the intra-lineage microsatellite diversity to provide a temporal aspect to their geographical distribution. The ages and distribution of these lineages provide little support for the dominant Archaeo-linguistic model for the origins of Oceanic populations that suggests that these peoples represent the Eastern fringe of an agriculturally-driven expansion initiated in Southeast China and Taiwan. Rather, most Micronesian and Polynesian Y chromosomes originate within different source populations in Melanesia and Eastern Indonesia. The Polynesian outlier, Kapingamarangi, is demonstrated to be an admixed Micronesian/Polynesian population. Furthermore it is demonstrated that a geographical classification of Oceanic islands best accounts for their extant Y-chromosomal diversity.



The complexity of HIV evolution (40min)

Allen Rodrigo, School of Biological Sciences, University of Auckland
Private Bag 92019, Auckland, New Zealand
Tel: 64-9-3737 599 Ext 7296, Fax: 64-9-3737 414
eFax (US): 1-413-826 5970
URL: www.cebl.auckland.ac.nz

HIV accumulates substitutions rapidly -- at the rate of 1% per year in some parts of its genome. This makes it an interesting model system (and a challenge) for evolutionary biologists. In this talk, I will discuss aspects of HIV molecular evolution, including recombination, population subdivision in vivo, and molecular epidemiology. All of these pose problems that we have only begun to address.

Phylogenetic Tree structure and the origins of HIV 1 (20min)

Andrew Rambaut,
Zoology Department, WWW - http://evolve.zoo.ox.ac.uk/
University of Oxford, TEL - +44 1865 271261
South Parks Road, Oxford, UK FAX - +44 1865 271249

We have investigated the phylogenetic structure of HIV-1 group M using samples from around the world. We compared these to 197 HIV-1 isolates sampled in 1997 from the Democratic Republic of Congo (DRC), a likely location for the origin of HIV-1. The strains from the DRC contain as much diversity as those for the global epidemic. Further, it was found that the observed phylogenetic structure of the phylogeny of the DRC sequences was consistent with a simple null model of exponential population growth. In contrast, phylogeny of strains from the global epidemic was not consistent with this model, exhibiting a clustered phylogenetic structure defined as subtypes. This suggests the DRC and global strains result from different epidemic histories. We discuss possible explanations and implications for the "natural transfer" hypothesis of the origins of HIV-1 group M.

Principles of co-phylogeny on HIV and primate co-evolution (20min)

Michael Charleston
Zoology Dept., University of Oxford, South Parks Road,
Oxford, UK

There are several systems in which ecologically related groups of organisms cospeciate, or more generally, codiverge. Farenholz' principle that parasite phylogeny mirrors host phylogeny has led to a growing array of analytical tools for determining the degree to which one phylogeny has mirrored another. The principles of cophylogeny and cophylogenetic maps are outlined here, and then applied to a current problem in virus evolution.

Primate lentiviruses (PLV) from closely related primate species have been observed to be more closely related to each other than to PLVs from more distantly related primate species. The current explanation for this observation is the codivergence hypothesis; the divergence of a viral lineage being the result of the divergence of the host lineage. We show that, alternatively, frequent cross-species transmission of PLV, coupled with a tendency for more closely related primate species to exchange viruses "successfully", can result in apparent codivergence. This host-switching hypothesis reconciles several puzzling observations related to the evolution of PLV.

Mitochondrial Genomes and Avian Phylogeny. (20min)

Kerryn Slack and Abby Harrison, Institute of Molecular BioSciences, Massey University
New Zealand
ph (w): + 646 350 5515 ext. 5404, ph (H): + 646 354 6521, fax + 646 350 5694

Despite many years of research using a variety of different molecular data sets, the times of origin and the relationships between avian orders are still uncertain. Nevertheless, there is increasing confidence from molecular studies using DNA sequences that the problem is now solvable, given sufficiently long sequences. Complete mitochondrial (mt) genomes have been a major source of data for studies on mammalian and piscine evolution. However, only twelve avian mt genomes are currently published (two of which are for the same species). Here we present five new avian mt genomes - a goose, a penguin and a New Zealand parrot (kakapo), owl (morepork) and wren (rifleman). The 12 mt protein-coding genes from these five birds plus ten others were analyzed and gave a relatively stable tree. However, the position of the root varied with the combination of outgroup (reptile) taxa used.

Mammal Phylogeny and Evolutionary Stable Niche Discontinuities (20min)

Matthew Phillips
Institute of Molecular BioSciences, Massey University, New Zealand
ph (w): + 646 350 5515 ext. 5404, ph (H): + 646 354 6521, fax + 646 350 5694

Meteor impacts, ice ages and big volcanoes have captured the imagination of film goers and paleontologists alike. Contrary to the convictions of Darwin, the importance of biological interactions for determining long-term evolutionary patterns has come under serious question.. For example, Benton (1987) suggested that “If competition exists at supraspecific levels, a new theory of competition in macroevolution will have to be developed”.

Recent advances in resolving the phylogenetic tree of mammals, and inference from the phylogenetic distribution of mammalian ecological niches implicates biological interaction in shaping long term patterns of extinction and evolution. I suggest that maintenance of niche distinctiveness between coexisting lineages over millions of years, occurs via evolutionary stable niche discontinuities (ESND). Are these ESND’s inherent in biological systems due to environmental discontinuities and the finite number of trophic interactions supported by ecosystems? How can these hypotheses be tested, and what are their implications?

A Stable Tree for Eutherian mammals? (20min)

Yu-hsin Owen Lin, Institute of Molecular BioSciences, Massey University, New Zealand
ph (w): + 646 350 5515 ext. 5404, ph (H): + 646 354 6521, fax + 646 350 5694

We have sequenced so far, 9 new mitochondrial genomes for placental mammals (shrew, pika, vole, dolphin, fur seal, gymnure and 3 bats) in order to get a reliable tree for this group. Taxa were primarily selected to break up long (unstable) edges on the tree, and secondarily to help improve estimation of evolutionary rates. These sequences have been included with the increasingly detailed tree of mammals. Comparison of trees from mitochondrial sequences with those from nuclear sequences allows an estimate of reliability of the eutherian tree.

Evolutionary individuality of amino acids (40min)

Alex Grossman

LogDet distances can be decomposed into contributions associated to individual amino acids (or nucleotides). For any reversible continuous single-generator Markov model, the relative sizes of the contributions obtained from a multiple alignment are independent of the pair of taxa under consideration. An examination of two sets of data (proteins coded by animal mitochondrial genomes and heat shock proteins) shows systematic deviations from this behaviour. The significance of these deviations is examined by rank correlation methods, and questions are raised about their origin.

Jumping Genes and Slithering Proteins: A Tale About Genes and Proteins Migrating Between Genomes (20min)

Lars Jermiin (address details given previously)

DNA sequencing and gel blot surveys was used to assess the integrity of the chloroplast gene *inf*A, which codes for translation initiation factor 1, in more than 300 diverse angiosperms. Whereas most angiosperms appear to contain an intact chloroplast *inf*A gene, the gene has repeatedly become defunct, in approximately 24 separate lineages of angiosperms. In four species in which chloroplast *inf*A is defunct, transferred and expressed copies of the gene were found in the nucleus, complete with putative chloroplast transit peptide sequences. The transit peptide sequences of the nuclear *inf*A genes from soybean and Arabidopsis were shown to be functional by their ability to target green fluorescent protein to chloroplasts in vivo. Phylogenetic analysis of *inf*A sequences and assessment of transit peptide homology indicate that the four nuclear infA genes are derived from at least two and probably four independent gene transfers from chloroplast to nuclear DNA during angiosperm evolution. Considering the large number of separate functional losses of *inf*A from chloroplast DNA, the gene may have been transferred many more times, making *inf*A the most mobile chloroplast gene to be documented in angiosperms.

Friday

Will the hypothesis of a single plastid origin be (dis) proven? (40min)

Tony Larkum, Peter Lockhart, School of Biological Sciences (A08),
University of Sydney, NSW 2006, Australia
Tel 61 2 9351 2069, Fax 61 2 9351 4119

In the 1980s a paper entitled “When will the endosymbiont hypothesis be proven?” (Gray and Doolittle, 1982) appeared which layed down the criteria needed to test the hypothesis that eukaryotes possessed genomes of different origin. Molecular Biology has since provided convincing evidence for this hypothesis. However, it has not convincingly resolved the question of the number of primary endosymbiotic events in the case of plastids – which are generally accepted as having a cyanobacterial ancestry. In this talk we examine the recent data suggesting that a single primary endosymbiotic event occurred, and discuss the grounds that prevent this hypothesis from being universally accepted at the present time.

From gene orders to phylogeny (20min)

David Bryant, (LIRMM, Montpellier and CRM, Montreal), LIRMM, 161 rue Ada, Montpellier 34392 France cedex 5, Ph: 33 4 67 41 86 55, Fax: 33 4 67 41 85 00

For some time now, the order of genes along genomes has been known to contain phylogenetic information. There are, however, systematic and computational problems that need to be overcome in order to exploit this source of data. I will discuss recent progress made on developing methods for reconstructing phylogenies from gene order data, in particular methods that can cope with unequal gene content. I will illustrate the methods using data from early eukaryotes, animal mitochondria, and, if it is finished in time, recently sequenced chloroplast data.

Reconstructing ancestral genomes by reversals and multiple gene duplication (20min)

Nadia El-Mabrouk, Departement d'Informatique et recherche operationnelle, Universite de Montreal, CP 6128 succ Centre-Ville, Montreal QC H3C 3J7, Canada
tel: 514 343 7481, fax: 514 343 5834

Traditionally, evolutionary relationships between species have been inferred from comparisons of single gene sequences. An alternative approach, the genome rearrangement approach, is to model the genome as a set of chromosomes, each consisting of a linear order of genes, and to compare these sets of gene orders. This approach infers divergence history in terms of global genomic mutations, involving the movement, inversion and duplication of chromosomal segments of various sizes. Implicit in the rearrangement literature is that both genomes to be compared contain an identical set of genes. While this hypothesis of unique genes may be appropriate for small genomes, it is clearly unwarranted for divergent species containing several copies of homologous genes. One of the most important regional events by which gene duplications can occur has been referred to as duplication transposition. In this model, entire regions of chromosomes are duplicated from one location of the genome to another. Our goal here is to take into account the mechanism of chromosomal segment duplication, or multiple duplications, in the evaluation of genomic distances. More precisely, for a genome G with gene families of different lengths, the implicit hypothesis is that G has an ancestor containing exactly one copy of each gene, and that G evolved from this ancestor through series of multiple duplications separated by series of reversals. The question addressed here is then: how can we reconstruct an ancestral genome giving rise to the minimal number of multiple duplications and reversals?

Reality Check (40min)

David Penny (address details given previously)

Inferring phylogenies, where the events being evaluated cannot be observed directly, is conceptually one of the most difficult problems in modern science. Most groups act as if they bring a limited background to the problem, such as Aristotelian logic, or classical statistics of the pre-Turing generation, or the magic of simulation. Some of the potential limitations of different approaches will be discussed. It is essential that all developments in modern mathematics be available for this conceptually difficult problem. We need more care that the mechanisms we use are at least possible biochemically. Bayesian approaches are not much help if the prior on a model being correct is precisely zero. Or is it okay to be purely instrumentalist - where a technique is used if it gives the correct answer, even if it has no biological interpretation. There are still many areas that have been little explored evaluating real data. Mention will be made of the potential of tree comparisons, visualization, and 3-D matrices (tensors). How could we evaluate if the known mechanisms of micro-evolution are sufficient to account for macro-evolution.

Other Participants

Dianne Gleeson
Landcare Research
Private Bag 92170, Auckland
tel: 64 98154200
fax 64 98497093

Tristan Armstrong
Landcare Research
Private Bag 92170, Auckland
tel: 64 98154200
fax 64 98497093

Nicoleen Cloete
Address: Departmant of Mathematics
University of Auckland
Private Bag 92019
Auckland

Leon Perrie
Institute of Molecular BioSciences, Massey University
New Zealand
ph (w): + 646 350 5515 ext. 5404
ph (H): + 646 354 6521
fax + 646 350 5694