From trees to processes

Peter Lockhart
Allan Wilson Centre for Molecular Ecology and Evolution, and Institute of Molecular BioSciences,
Massey University, Private Bag 11-222, Palmerston North

One of the first things David Penny ever told me was, '*Only follow a computer to the edge of an intellectual cliff* .' He shocked me by saying, '*Do you really think that tree building is scientific?'* We were discussing, by email, efforts at the time to reconstruct the earliest origins of photosynthesis. I was trying to write up my PhD in Sydney and was a bit miserable because I doubted what I and everybody else was doing – trying to reconstruct the evolutionary history of anciently diverged organisms by comparing amino acid sequences.

The problem was that tree-building methods make such simple assumptions about how DNA and protein sequences evolve, and the patterns that I noticed in my data didn't really make that much sense under these assumptions – but did that really matter? Were the reconstructed histories from sequences accurate? I wondered if my evolutionary trees could be wrong and misleading. David emailed me that he and Mike Hendy had recently described some theoretical findings that would soon be published (Penny *et al.* 1990) showing that they had similar concerns. This thinking drew me to New Zealand; I thought I was coming for a few weeks but that was 17 years ago.

I really liked the way David and the Mikes did Science – it was all about testing ideas and using mathematical approaches to do that. Best of all, the problem solving was often done in out-of-the-way places, such as in the Kaikoura mountains, on the Volcanic Plateau, or in Arthurs Pass. David and Mike formalised tree thinking in terms of evolutionary models (Penny *et al.* 1990). This meant thinking clearly about your hypothesis (usually a tree with specified branch lengths) and your assumptions about how sequences evolve.

Assumptions about how sequences evolve

In the early days we published a paper in *Nature* (Steel *et al.* 1993) about the problem that occupied and worried me during my PhD studies at the University of Sydney. In this and a subsequent paper (Lockhart *et al.* 1994) we showed how tree building could be misled when the process of evolution differed in evolutionary lineages. We later showed how mis-specification of substitution models worked against you in two ways: it could (a) induce topological distortion that could wrongly favour the recovery of an incorrect tree (Lockhart *et al.* 1996, 2000, 2006; Lockhart & Steel 2005) and (b) reduce the reconstruction accuracy of tree building even if it did not favour any particular tree (Lockhart *et al.* 1996). This way of thinking followed from Mike and David's earlier work (Hendy & Penny 1989) that showed that the long branch attraction problem in phylogenetic reconstruction, first identified by Joe Felsenstein (Felsenstein 1978), could result from any property of data that causes irregular lengths of branches in reconstructed trees.

Two of David's current students, Klaus Schliep and Liat Shavit Grievink, have projects continuing some of this work. Liat has recently developed a simulator that allows the user to simulate properties of data that are more realistic. She is using this program to study the performance of different tree-building methods and to investigate the nature of biological controversies concerning the evolution of parasites, and the early origins of plant evolution. The idea being tested in both cases is whether lineage-specific relaxation of structural constraints explains the difficulty reconstructing evolutionary history. Klaus has been involved in a study to help understand why molecular clock analyses of different proteins give very different age estimates for the origin of oxygenic photosynthesis. He is looking at whether co-evolution of protein-protein interactions misleads our efforts to reconstruct the earliest events of evolutionary history. Many others are also pursuing similar questions and the problem of model mis-specification is currently recognised as one of the most important concerns for evaluating the reliability of tree building.

Is a bifurcating tree a sufficient descriptor for evolutionary relationships?

Those trips to Arthurs Pass, skirmishes into the alpine zone, and the very many discussions with David and his PhD student at the time, Robert Hickson, opened my eyes to the New Zealand flora. It was also David's thinking about whether a bifurcating tree was a good descriptor for evolutionary relationships that started

Peter Lockhart is a Principal Investigator at the Allen Wilson Centre for Molecular Ecology and Evolution and Professor of Molecular Evolution at the Institute of Molecular BioSciences (IMBS) at Massey University.

Professor Lockhart's research interests include plant species radiation, Pacific biodiversity, endosymbiosis and organelle evolution and evolutionary properties of DNA and protein sequences. He may be contacted at p.j.lockhart@massey.ac.nz me thinking about the sufficiency of bifurcating evolutionary models to explain certain evolutionary relationships. This led to studies with David ranging from questions about sexual selection in swordtail fish (Lockhart *et al.* 1995) to the studies on the nature of basal angiosperms (Lockhart & Penny 2005).

Understanding processes

We have learnt in our DNA studies on New Zealand plants that much of the species diversity in the NZ flora results from Late Tertiary diversification processes (Winkworth *et al.* 2002). That is, from a single transoceanic introduction, numerous plant groups have radiated within the last few million years to produce a diversity of ecological and morphological forms. Reconstructing the evolutionary relationships of these forms is difficult not only because of the shape of the underlying phylogeny but also because hybridisation is an important feature of these plant groups. It was realisation of this complexity that largely stimulated my interest in phylogenetic networks (Lockhart *et al.* 2001; Huber *et al.* 2001; Winkworth *et al.* 2005) and most recently in the network methods that can make use of the genome-wide markers that will soon be readily available as a consequence of next-generation sequencing technologies (McBreen & Lockhart 2006). Through the Allan Wilson Centre and the recent New Zealand Genome Consortium Initiative, David has also been a strong advocate for New Zealand obtaining next-generation sequencing technologies.

Many automatically think of the sequencing of complete genomes when this technology is discussed, but its greatest value in Ecology and Evolution is most likely to be in developing genome-wide molecular markers and for use in studies of differential gene expression. Not only can these tools help us to understand better the nature of New Zealand and Pacific biota; the technology provides a means for monitoring the environment and investigating the adaptive response of our biota to future environmental changes (Huson *et al.* 2007; Hoffman & Willi 2008).

There is so much excitement ahead for Science in New Zealand, and I have David to thank very much for the opportunity to be involved.

References

Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology 27*: 401–410.

Hendy, M.; Penny, D. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology 38*: 297–309.

Hoffman, A.A.; Willi, Y. 2008. Detecting genetic responses to environmental change. *Nature Reviews Genetics 9*: 421–432

Huber, K.T.; Moulton, V.; Lockhart, P.J; Dress A. 2001. Pruned Median Networks: A technique for studying plant speciations. *Molecular Phylogenetics and Evolution 19*: 302–310.

Huson, D.; Auch, A.F.; Qui, J.; Schuster.; S.C. 2007. MEGAN analysis of metagenomic data. *Genome Research 17*: 377–386

Lockhart, P.J.; Huson, D.; Maier, U.G.; Faunholz, M.J.; Van de Peer, Y.; Barbrook, A.C.; Howe, C.J.; Steel, M.A. 2000. How molecules evolve in eubacteria. *Molecular Biology and Evolution 17*: 835–838.

Lockhart, P.J.; Larkum, A.W.D.; Steel, M.A.; Waddell, P.J.; Penny D. 1996. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. *Proceedings National Academy of Science USA 93*: 1930–1934.

Lockhart, P.J.; McLenachan, P.A.; Havell, D.; Glenny, D.; Huson, D; Jensen, U. 2001. Phylogeny, dispersal and radiation of New Zealand alpine buttercups: molecular evidence under split decomposition. *Annals of the Missouri Botanical Gardens 88*: 458–477.

Lockhart, P.J.; Novis, P.; Milligan, B.G.; Riden, J.; Rambaut, A; Larkum, A.W.D.; 2006. Heterotachy and Tree Building: A case study with plastids and eubacteria. *Molecular Biology and Evolution 2*3: 40–45.

Lockhart, P.J.; Penny, D. 2005. The place of *Amborella* in the radiation of angiosperms. *Trends in Plant Science 10*: 201–202.

Lockhart, P.J.; Steel, M.A. 2005. A tale of two processes. *Systematic Biology 54*: 948–951.

Lockhart, P.J.; Steel, M.A.; Hendy, M.D.; Penny D. 1994. Recovering the correct evolutionary tree under a more realistic model of sequence evolution. *Molecular Biology and Evolution 11*: 605–612.

Lockhart; P.J.; Meyer, A.E.; Penny, D. 1995. Testing the phylogeny of swordtail fishes using split decomposition and spectral analysis. *Journal of Molecular Evolution 41*: 666–674.

McBreen, K.; Lockhart, P.J. 2006. Reconstructing the evolutionary history of plants. *Trends in Plant Science 11*: 398-404.

Penny, D.; Hendy, M.; Zimmer, E.A.: Hamby, R.K. 1990. Trees from sequences, panacea or Pandora's box? *Australian Systematic Botany 3*: 21–28.

Steel, M.A.; Lockhart, P.J.; Penny D. 1993. Confidence in evolutionary trees from biological sequence data. *Nature 364*: 440–442.

Winkworth, R.C.; Bryant, D.; Lockhart, P.J.; Havell D.; Moulton V. 2005. Biogeographic interpretation of split graphs: least squares optimization of branch lengths. *Systematic Biology 54*: 56–65.

Winkworth, R.C.; Wagstaff, S.; Glenny, D.; Lockhart, P.J. 2002. Plant dispersal N.E.W.S. from New Zealand. *Trends in Ecology and Evolution 17*: 514–520.