Papers

Mathematics in evolution

Michael D. Hendy ONZM

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

I have held the chair in Mathematical Biology at Massey University for 15 years, and I can thank (or blame?) no one more than Professor David Penny as the catalyst for this singular honour. In my early education, the subjects of mathematics and biology were regarded as polar opposites in the spectrum of science subjects. At my high school in the 1960s, I was not able to advance a combination of these two subjects beyond the 4th form. Sadly there exists a disconnect between these subjects in schools and universities even today, with so few graduates of biology having an adequate mathematical background, although there is a significant drift of mathematics graduates who choose to apply their skills into biological applications.

On the surface, evolution may not seem to have much mathematical content, yet like many scholars of his generation, Darwin was sufficiently numerate for mathematical ideas to have a significant impact on his thinking. Indeed he acknowledged the Malthusian thesis of the exponential population growth quickly outstripping its nutritional support and leading to massive pruning – and so Darwin was able to propose that if there were a mechanism whereby the survivors' attributes could be transmitted to their offspring, then those attributes could become more prevalent in the succeeding generations, paraphrased as 'survival of the fittest'.

Charles Darwin had learned from his grandfather Erasmus that the fossils uncovered in canal tunnels were layered, with progressions in form from layer to layer, implying their ordering in time. Charles also had built a significant background in geological surveying before his *Beagle* voyage. Hence it was fortuitous for his theory that he witnessed at first hand a massive uplift of the Chilean coastline during a major earthquake in 1835. Combining this with his observation of a series of uplifted beaches, and the abundance of marine deposits at high altitude in the Andes, he came to appreciate the dynamics of geology and the long timeframe necessary, time available for significant evolutionary change.

In his 'Origin' Darwin had but one sketch, that of a tree whose tips represented species, with its branches the ancestral groupings reaching back to a trunk of common origin. This metaphorical object has been a major source of research of many of those assembled here today, and in particular of David Penny, who has led us in making significant contributions to this field. At our first contact 35 years ago, David introduced me to evolutionary trees (phylogenies) and we discussed how they were estimated from comparing biological sequence data (then protein sequences, printed in book form (Dayhoff *et al*. 1972)). We soon confronted the challenge of the super-exponential growth of the number of potential phylogenies with the increase in the number of sequences.

This challenge arose when David proposed a plan to investigate Karl Popper's statement 'Darwinism is not a testable scientific theory, but a metaphysical research program', which is often misquoted by evolution deniers to support their creationist views. David proposed a number of potential predictions. We chose to investigate one of these by examining the relationships between phylogenies built from five different proteins, for a common set of 11 species (Penny *et al*. 1982). Evolution would predict that these trees should be closely related, each reflecting the evolutionary history of the species. We soon discovered that the computing power available to us at that time was inadequate to consider the nearly 35 million (34 459 425) potential phylogenies, forcing us to develop the branch and bound algorithm (Hendy & Penny 1982), to focus this search. This branch and bound algorithm is now in common usage in many phylogenetic packages, and each implementation is in itself a test of evolution, as its efficiency is a consequence of the evolutionary nature of the data.

We also had to interpret the similarities between phylogenies by interpreting the distribution of the tree partition metric (Robinson & Foulds 1981). Our first paper on this distribution (Hendy *et al*. 1984) enticed Mike Steel to join us as a PhD student, and Mike subsequently completely solved the distribution

Mike Hendy is Professor of Mathematical Biology at Massey University and a Principal Investigator (formerly Executive Director) at the Allan Wilson Centre for Molecular Biology and Evolution.

Distinguished Professor Hendy's research interests centre on phylogenetic analysis of biological sequence data. One useful, but unexpected discovery in this research has been the utility of Hadamard matrices for modelling the evolution of nucleotide sequences.

A Fellow of the Royal Society of New Zealand, Distinguished Professor Hendy received the New Zealand Mathematical Society's Research Award in 2008 and was made an Officer of the New Zealand Order of Merit in 2009. He may be contacted at M.Hendy@massey.ac.nz

(Steel 1988). David's work also enticed Pete Lockhart, with ink still wet on his PhD certificate, to come to New Zealand to spend a few months with us. This has taken Pete a little longer than he expected, so 20 years on he is still with us. However, it is not that Pete is a slow learner, it reflects the fact that David Penny, in leading our group, was able to provide us with an overwhelming succession of interesting challenges, so that none of us had time to think of going elsewhere. I am sure that others here today will offer similar experiences. Our annual off-site research meetings rapidly grew to encompass many other New Zealand researchers as well as becoming an attractant for overseas visitors.

Hence, when in 2001, our government proposed their Centres of Research Excellence (CoRES) initiative (Thanks Steve Maharey!) we were well placed to propose the Allan Wilson Centre, and we were fortunate to be selected as one of the initial five CoRES in 2002 and for a second round of funding this year. It has been a privilege to work alongside David Penny as co-Director of the Allan Wilson Centre for the last seven years. As you should well know, our Centre commemorates the inspiration of one of New Zealand's most innovative scientists – I shall only mention the concept of 'Molecular Clock' that Allan Wilson (Ochman & Wilson 1987) introduced to molecular biology. This tool, by which we can estimate the timing of evolutionary history, was exploited by David Penny and Alan Cooper (and others) to demonstrate how the major divergences of birds (Cooper & Penny 1997) and mammals (Bromham *et al.* 1998) had preceded, by many tens of millions of years, the bolide collision of 65 million years ago, the event that purportedly caused the extinction of the dinosaurs. Their work overturned the popularly held view that the birds and mammals only developed after the demise of the dinosaurs.

That is just one example of the impact of the Penny bolide on biology. From a biological background, David has recognised and focused the power of mathematical ideas on issues of major evolutionary interest to not just produce some important theoretical and practical results, but to also inspire a whole new generation of students into exploiting mathematical tools to gain a greater understanding of evolutionary biology.

I have mentioned just a few examples; I am sure others here will provide much more evidence of David's inspirational guidance. However, I do want to highlight a particular personal benefit of my association with David. I was fortunate to recognise a beautiful mathematical relationship linking the parameters of evolutionary change on a phylogeny with the patterns of nucleotide differences in the sequences evolving under a simple model. This relationship, which is called Hadamard conjugation (Hendy & Penny 1989), has proved to be a fruitful theoretical tool. While I struggled with a formal mathematical proof, David Penny calmly assured me it must be right, as he had written a computer simulation package, which came up with results in complete agreement with my theoretical calculations.

Finally I would like to express my deep gratitude to David Penny for his leadership, his scholarship, and his friendship. As a scientist who always demands a testable hypothesis at the start of each scientific quest, he has shown us the importance of approaching science with an open mind. As a fierce opponent of dogma, he has shown the importance of mathematical models to measure the applicability of each proposed hypothesis. His legacy will be this new generation of world-leading scientists assembled for this celebration. He has always considered it a most important need to assist in the career development of all students and Fellows he has mentored. His legacy is reflected in the gathering here today.

References

Bromham, L.D.; Rambaut, A.; Fortey, R.; Cooper A.; Penny, D. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proceedings National Academy of Sciences*, USA. 95: 12386–12389.

Cooper, A.; Penny. D 1997. Mass survival of birds across the Cretaceous/Tertiary boundary. *Science* 275: 1109–1113.

Dayhoff, M.O.; Eck, R.V.; Park, C.M. 1972. *Atlas of Protein Sequence and Structure*, vol 5. Washington DC, National Biomedical Research Foundation.

Hendy, M.D.; Little, C.H.C.; Penny, D. 1984. Comparing trees with pendant vertices labelled. *SIAM Journal on Applied Mathematics 44*: 1054–1065.

Hendy M.D.; Penny, D. 1982. Branch and bound algorithms to determine minimal evolutionary trees. *Mathematical Biosciences 59*: 277–290.

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

Ochman, H.; Wilson, A.C. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution 26*: 74–86.

Penny, D.; Foulds, L.R.; Hendy, M.D. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature 297*: 197–200.

Robinson, D.F.; Foulds, L.R. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences 53*: 131–147.

Steel, M.A. 1988. Distribution of the symmetric difference metric on phylogenetic trees. *SIAM Journal on Discrete Mathematics 1(4)*: 541–551.