Darwin, language, and two great Pacific voyages

Simon J Greenhill and Russell D Gray
Department of Psychology, University of Auckland, Private Bag 92-019, Auckland 1142. New Zealand

On 21 December 1835 Charles Darwin arrived in New Zealand on the *HMS Beagle.* The *Beagle* had just visited the Galapagos Islands, where Darwin had made some of the critical observations that he would later incorporate into his theory of evolution. Darwin did not like New Zealand:

> *I believe we were all glad to leave New Zealand. It is not a pleasant place. Amongst the natives there is absent that charming simplicity which is found in Tahiti; and the greater part of the English are the very refuse of society. Neither is the country itself attractive.* (Darwin 1860, p. 430)

Around 700 years earlier another set of travellers arrived in New Zealand – the ancestors of the Maori. Unlike Darwin, the Maori liked New Zealand and decided to stay.

On returning to England, Darwin went on to develop his theory of biological evolution. However, he was also fascinated by language. In *The Descent of Man*, he noted that languages also evolve:

> *The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel... We find in distinct languages striking homologies due to continuity of descent, and analogies due to a similar process of formation.* (Darwin 1871, p. 59–60)

There are numerous parallels between the evolution of languages and biological evolution. Languages contain units – phonemes, morphemes, words, and grammatical structures – that have been inherited from ancestral languages. Like species, languages mutate, combine and recombine, borrow or hybridise with each other, gain new functions, lose old ones, and become extinct. Like genes, lexicon carries historical signal about the people who spoke these languages and their cultures, and can be used to explore prehistory.

As early as the 16th and 17th centuries, European explorers were struck by the obvious similarities in the languages of the Pacific (Greenhill & Gray 2009). Table 1 shows some items of basic vocabulary from nine Austronesian languages. The similarities in these items are so great that it is extremely unlikely that they could have arisen by chance. Instead, these words must have sprung from some common source, and have changed through a process of descent with modification. Linguists have developed a rigorous methodology for identifying historical relationships between languages that requires more than just similarities in form and meaning (Durie & Ross 1996). Homologous words – termed cognates – are identified by linguists looking for systematic sound correspondences between languages. For example, in Table 1 there is a regular correspondence between Tahitian 'v' and Maori 'w' reflected in the words for woman (Tahitian ***v****ahine*, Maori ***w****ahine*) and eight (Tahitian ***v****aru*, Maori ***w****aru)*. Another correspondence shown in Table 1 reflects the common 'l' to 'r' sound change, with Navosa three (*to**l**u*) related to Maori and Tahitian (*to**r**u*). This sound change is also reflected in the words for five (Navosa ***l****ima*, Tahitian ***r****ima*, Maori ***r****ima*).

The cognate words shown in Table 1 show affiliations stretching from Taiwan to Polynesia and thus suggest a common origin of Pacific peoples. The settlement of the Pacific is one of the greatest human migrations and has been a topic of enduring fascination for decades. In the 1940s, Thor Heyerdahl suggested (like Darwin did) that the Polynesians arrived from South America on balsa-log rafts. However, modern research has shown this to be incorrect – genetically and linguistically the Polynesian peoples are more closely related to those in South East Asia. Polynesian languages are a subset of the large Austronesian language family.

Simon Greenhill is a Postdoctoral Research Fellow in the Computational Evolution Group at the University of Auckland. His PhD thesis applied computational Bayesian phylogenetic methods to language data to test hypotheses about the human settlement of the Pacific. Simon is currently interested in applying these methods to broader questions about the evolution of languages and cultures. He may be contacted at: s.greenhill@auckland.ac.nz

Russell Gray is a Professor in the Department of Psychology at the University of Auckland. His research interest spans four areas: language evolution, animal cognition, avian evolution and philosophy of biology. He may be contacted at rd.gray@auckland.ac.nz

See also their website http://language.psy.auckland.ac.nz/austronesian

Table 1. There are striking similarities in the basic vocabulary of Austronesian languages stretching from Taiwan to Polynesia.

Language	Amis	Paiwan	Cebuano	Manggarai	Marau	Kwai	Western Fijian (Navosa)	Tahitian	Maori
	Taiwan	*Taiwan*	*Philippines*	*Indonesia*	*West New Guinea*	*Solomons*	*Polynesia*	*Polynesia*	*Polynesia*
eye	mata	mata	matsa	mata	mā-na	mata-na	mata	mata	mata
louse	kuto	kutsu	kúto	utu	ʔū	ʔuu	kutu	'utu	kutu
woman	fafahian	vavaian	babáye	wina	*keni*	*noni*	*lewa*	vahine	wahine
one	cecaj	ita	usa	ca	eta	eta	kia	tahi	tahi
two	tusa	ḍusa	duhá	sua	rua	rua	rua	*piti*, rua (archaic)	rua
three	tulu	tjelu	tulo	telu	ɔru	olu	tolu	toru	toru
four	sepat	sepatj	upát	pat	hai	fai	vā	maha	whā
five	lima	rimá?	lima	lima	nima	nima	lima	rima	rima
six	'enem	unum	unu'm	enem	ono	ono	ono	ono	ono
seven	pitu	piṭu?	pito	pitu	hiu	fiu	vitu	hitu	whitu
eight	falu	'áru?	walo'	alo	waru	kwalu	walu	varu	waru
nine	siwa	siva?	siam	ciok	siwa	sikwa	ciwa	iva	iwa
ten	polo	púruq	napulo	cempulu	tanahuru	taŋafulu	*jini*	'ahuru	*tekau* ŋahuru (archaic)

Note: Items in italics are non-cognate.

There are two major competing hypotheses for the settlement of the Pacific. The first hypothesis is the 'pulse–pause' scenario (e.g. Blust 1999, Kirch & Green 2001, Pawley 2002, Gray *et al.* 2009). This scenario argues that the ancestral Austronesian society developed in Taiwan around 5500 years ago. Around 4000–4500 years ago, there was a rapid expansion pulse across the Bashii channel into the Philippines, into Island South-East Asia, along the coast of New Guinea, reaching Near Oceania by around 3200–3600 years ago. As the Austronesians travelled this route, they integrated with existing populations in the area (particularly in New Guinea), and innovated new technologies. After reaching Western Polynesia (Fiji, Tonga, and Samoa), the Austronesian expansion paused for around 1000 years, before a second rapid expansion pulse into Remote Oceania – eventually reaching as far afield as New Zealand, Hawaii, and Rapanui. During this long pause in Near Oceania, the distinctive features of Polynesian society were developed.

The second hypothesis of Pacific settlement – the 'slow boat' scenario – argues for a much older origin in Island South-East Asia (Oppenheimer & Richards 2001, Hill *et al.* 2007, Soares *et al.* 2008). According to this scenario, date estimates from mitochondrial DNA lineages suggest that Austronesian society developed around 13 000–17 000 years ago in an extensive network of sociocultural exchange in the Wallacean region around Sulawesi and the Moluccas. Proponents of this scenario propose that the submerging of the Sunda shelf at the end of the last ice-age triggered the Austronesian expansion (Soares *et al.* 2008). This 'flood' led to a two-pronged movement of people, north into the Philippines and Taiwan, and east into the Pacific.

These two scenarios of Pacific settlement make quite different predictions about the origin, age, and sequence of the Austronesian expansion. How might one test these predictions? Darwin argued that evolution is best represented as a tree of species. This insight has led to the development of explicit computational methods for building evolutionary trees (phylogenies) and testing evolutionary hypotheses. Our work has applied these biological phylogenetic methods to analyses of languages and cultures (Gray & Jordan 2000; Gray & Atkinson 2003; Gray *et al.* 2007, 2009; Jordan *et al.* 2009).

To make accurate inferences, computational phylogenetic methods require a large amount of data. Basic comparative data on the languages of the world are often widely dispersed in hard-to-obtain sources. In collaboration with a world-wide team of linguists, we have collected wordlists from over 600 languages around the Pacific, and built a web-accessible database to store them (Austronesian Basic Vocabulary Database, http://language.psy.auckland.ac.nz/austronesian). This database contains items of basic vocabulary such as words for body parts, kinship terms, colours, numbers, and simple verbs. Basic vocabulary is ideal for phylogenetic analysis, as it is highly stable over time and resistant to being borrowed between languages (Greenhill *et al.* 2008).

To test between the hypotheses of Pacific settlement we applied Bayesian phylogenetic methods to the cognate data from 400 of the languages from the Austronesian Basic Vocabulary Database (Gray *et al.* 2009). These methods allow us to model how languages change over time (i.e. how cognate sets are gained and lost), and allow different cognate sets to change at different rates over time. However, finding the most probable trees linking 400 languages is not an easy task – there are over 5.8×10^{984} possibilities (Felsenstein 1977). Therefore, the tactic used by modern tree-building methods is to 'walk' through the space of these possible trees and sample a subset of the best-fitting trees (Greenhill & Gray 2009). This process is extremely computationally intensive – the analyses took over 100 years of computer time on a large cluster of computers.

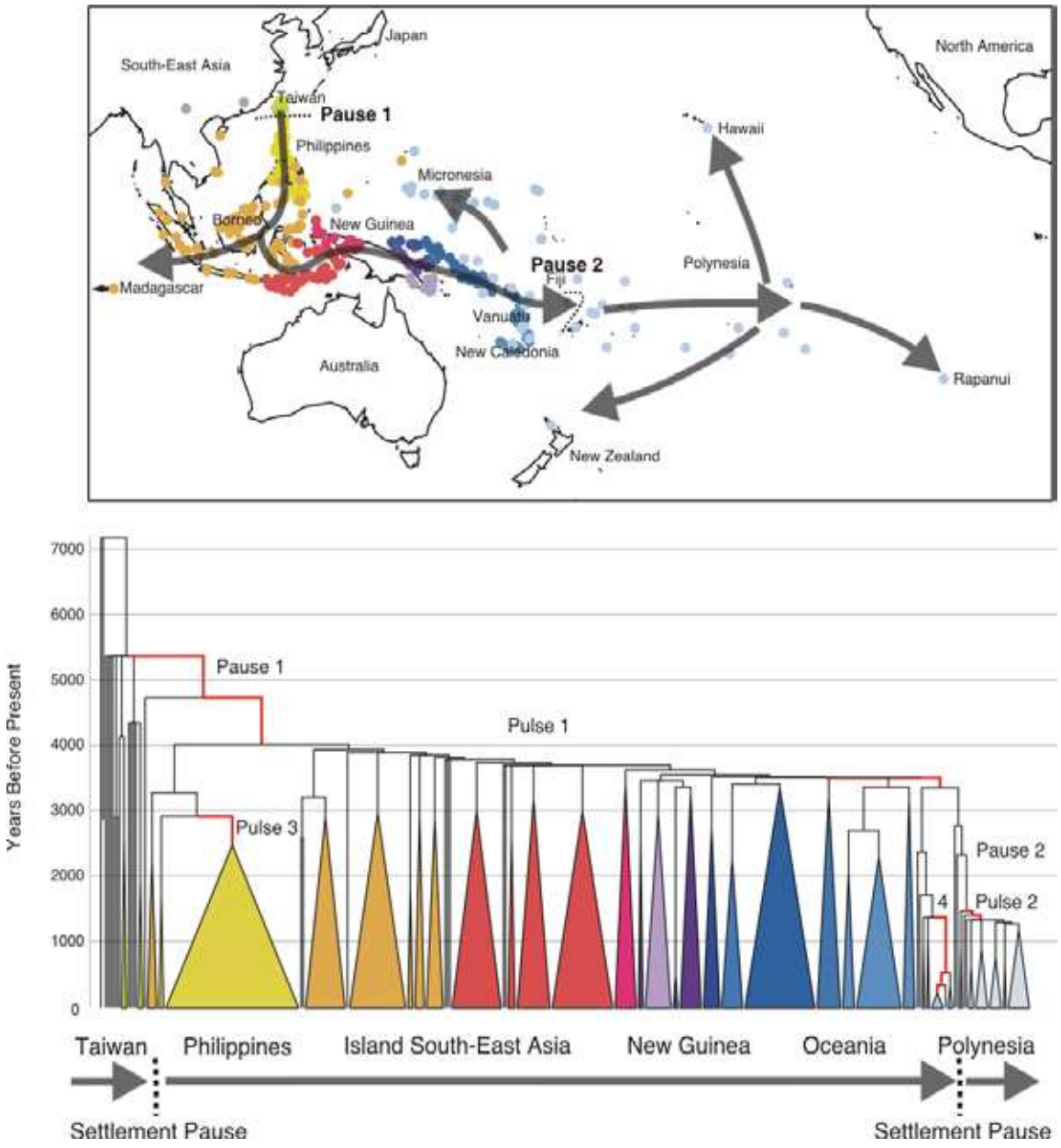

Figure 1. Map and language family tree showing the settlement of the Pacific by Austronesian-speaking peoples. The map shows the settlement sequence and location of expansion pulses and settlement pauses. The tree is rooted with some outgroup languages (Buyang and Old Chinese) at its base. It shows an Austronesian origin in Taiwan around 5200 years ago, followed by a settlement pause (Pause 1) between 5200 and 4000 years ago. After this pause, a rapid expansion pulse (Pulse 1) led to the settlement of Island South-East Asia, New Guinea and Near Oceania in less than 1000 years. A second pause (Pause 2) occurs after the initial settlement of Polynesia. This pause is followed by two pulses further into Polynesia and Micronesia around 1400 years ago (Pulses 2 and 4). A third expansion pulse occurred around 3000–2500 years ago in the Philippines.

The first prediction we tested was the origin and sequence of expansion. Under the pulse–pause scenario, the Austronesians originated in Taiwan and expanded sequentially down through the Philippines, through Wallacea, along New Guinea into Near Oceania and Polynesia. In contrast, the slow-boat scenario posits a two-pronged expansion from a Wallacean origin. Our set of trees placed the root of the trees in Taiwan, and followed it with the sequence predicted by the pulse-pause scenario (Figure 1).

The second key prediction of the two Pacific settlement scenarios concerned the age of the expansion. The pulse–pause scenario argues for an age of around 5500 years ago, whilst the slow-boat argues for an older age around 13 000–17 000 years ago. To test this prediction we estimated the age at the root of our trees. To begin with, we calibrated ten nodes on the trees with archaeological date estimates and known settlement times. For example, speakers of the Chamic language subgroup were described in Chinese records around 1800 years ago and probably entered Vietnam around 2600 years ago (Thurgood 1999). We can therefore calibrate the appearance of the Chamic node on our tree to between 2000 and 3000 years ago. A second calibration, based on archaeological evidence, constrains the age of the hypothesised ancestral language spoken by all the languages of Near Oceania, Proto-Oceanic. The speakers of Proto-Oceanic arrived in Oceania around 3200–3600 years ago and brought with them distinctively Austronesian societal organisation and cultural artefacts. These artefacts have been identified and dated

archaeologically, and include the Lapita adze/axe kits, housing types, fishing equipment (such as the one-piece rotating fish-hooks, and one-piece trolling lure), as well as common food plants and domesticated animals from South-East Asia.

To estimate the age of the Austronesian family we then implemented a phylogenetic dating method that takes the observed amount of change (in cognate sets) and uses the known calibrations to estimate how much change occurs per unit of time. We can then use this estimate to calculate the age of the trees (Gray & Atkinson 2003). The results unequivocally supported the younger age of the pulse–pause scenario, with an origin of the Austronesian family around 5200 years ago.

The pulse–pause scenario makes a third key prediction by proposing a sequence of expansion pulses and pauses. Under this scenario, there were two pauses in the great expansion – the first occurred before the Austronesians entered the Philippines around 5000–4000 years ago, and the second after the settlement of Western Polynesia (Fiji, Samoa, Tonga) around 2800 years ago. We tested this prediction in two ways. First, we identified the branches on our trees corresponding to these two pauses (Figure 1). The length of the branches again represents the number of changes in cognate sets. If these pauses did occur, those branches should be much longer than others due to the increased amount of time for linguistic change. Indeed, the length of these branches was significantly longer than the overall branch-length distribution, providing good evidence that pauses did occur in the predicted locations.

The pulse–pause scenario predicts pulses as well as pauses. If there were expansion pulses in language change, we would expect to see increases in language diversification rates after the predicted pauses. To do this, we modelled language diversification rate over our set of language trees. This method identified a number of significant increases in language diversification rates (branches coloured red in Figure 1). Two of these increases occurred as predicted on the branches just after the two pauses. Intriguingly, we identified some unpredicted pulses as well. The third pulse we identified suggested a more recent population expansion in the Philippines around 2000–2500 years ago as one language subgroup expanded at the expense of others. The fourth pulse occurred in the Micronesian languages and occurs almost contemporaneously with the second pulse into Polynesia, suggesting that these pulses may be linked.

What might have caused the great Austronesian expansion? It has been suggested that the first pause might be linked to an inability of the Austronesians to cross the 350 km Bashi channel separating the Philippines from Taiwan (Blust 1999, Pawley 2002). Terms for outrigger canoes and sails can only be reconstructed back to the languages occurring after the first pause (Blust 1999, Pawley 2002). It seems likely therefore that the invention of the outrigger enabled the Austronesians to cross the channel and spread rapidly across the rest of the Pacific. After travelling 7000 kilometres in 1000 years, what might have caused the Austronesians to stop in Western Polynesia? Expanding into Eastern Polynesia presented the Austronesians with a new range of challenges that would have also required technological or social solutions including: the ability to estimate latitude from the stars, the ability to sail across the prevailing easterly tradewinds, double-hulled canoes for greater stability and carrying capacity, and social strategies for handling the greater isolation (Irwin 1998).

Our results reveal the rapidity of cultural spread. The Austronesians travelled – and settled – the 7000 kilometres between the Philippines and Polynesia in around 1000 years. During this relatively short time, the Austronesian culture not only spread, but developed the collection of technologies known as the Lapita cultural complex (Kirch & Green 2001). This complex includes distinctive and elaborately decorated pottery, adzes/axes, tattooing, bark-cloth, and shell ornamentation. Our results suggest that either this complex was generated in a very short time-window (four or five generations), or there was substantial post-settlement contact between Near Oceania and the pre-Polynesian society. One possibility is that there is a more complex history in this region. The languages of New Caledonia and Vanuatu show some strikingly non-Austronesian features such as serial verb constructions, and the cultures there show some unusual similarities with some from highland New Guinea – including nasal septum piercings, penis sheathes, and mop-like headdresses (Blust, in press). It has recently been suggested that one explanation for these similarities might be two waves of settlement into Remote Oceania, with a first wave of Austronesian-speaking settlers being rapidly followed by Papuan peoples who had acquired Austronesian technology (Blust, in press).

The results of our phylogenetic analyses of Austronesian languages show the combined power of computational evolutionary methods and large lexical databases to answer questions about human prehistory. Just as molecular phylogenies form the fundamental framework for studying biological evolution, so language trees open up the exciting possibility of a Darwinian approach to culture based on rigorous phylogenetic methods (Gray *et al.* 2007). As Darwin argued, the tree of life covers the surface of the earth with its 'ever-branching and beautiful ramifications' (Darwin 1859, p. 130). Language phylogenies enable us to answer questions about human origins and explore the beautiful ramifications of human languages and cultures.

References

Blust, R.A. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. Pp. 31–94 in Zeitoun, E.; Li, P.J.K. (Eds) *Selected papers from the eighth international conference on Austronesian linguistics* Vol. 1. Taipei, Taiwan: Symposium Series of the Institute of Linguistics, Academia Sinica.

Blust, R.A. In press. Remote Melanesia: One history or two? An addendum to Donohue & Denham. *Oceanic Linguistics.*

Darwin, C.R. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London, John Murray. (1st edn) 502 p.

Darwin, C.R. 1860. *Journal of Researches Into the Natural History and Geology of the Countries Visited During the Voyage of H.M.S. "Beagle" Round the World, Under the Command of Capt. Fitz Roy, R.N.* London: John Murray. Online: http://darwin-online.org.uk/content/frameset?itemID=F20&viewtype=text&pageseq=1

Darwin, C.R. 1871. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray. Online: http://darwin-online.org.uk/content/frameset?itemID=F937.1&viewtype=text&pageseq=1

Durie, M.; Ross, M. 1996. *The Comparative Method reviewed: Regularity and irregularity in language change*. Oxford, Oxford University Press.

Felsenstein, J. 1977 The number of evolutionary trees. *Systematic Zoology 1*: 27–33.

Gray, R.D.; Atkinson, Q.D. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature 426*: 435–439.

Gray, R.D.; Drummond, A.J.; Greenhill, S.J. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science 323*: 479–483.

Gray, R.D.; Greenhill, S.J.; Ross, R.M. 2007. The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory 2*: 360–375.

Gray, R.D.; Jordan, F.M. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature 405*: 1052–1055.

Greenhill, S.J.; Blust, R.; Gray, R.D. 2008. The Austronesian Basic Vocabulary Database: From bioinformatics to lexomics. *Evolutionary Bioinformatics 4*: 271–283.

Greenhill, S.J.; Gray, R.D. 2009. Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In: Adelaar, A.; Pawley, A. (Eds) *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*. Canberra, Pacific Linguistics.

Hill, C.; Soares, P.; Mormina, M.; Macaulay, V.; Clarke, D.; Blumbach, P.B.; Vizuete-Forster, M.; Forster, P.; Bulbeck, D.; Oppenheimer, S.; Richards, M. 2007. A mitochondrial stratigraphy for island southeast Asia. *American Journal of Human Genetics 80*: 29–43.

Irwin, G. 1998. The colonization of the Pacific: Chronological, navigational and social issues. *Journal of the Polynesian Society 107*: 111–144.

Jordan, F.M.; Gray, R.D.; Greenhill, S.J.; Mace, R. 2009. Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B. 276*: 1957–1964.

Kirch, P.; Green, R. 2001. *Hawaiki, Ancestral Polynesia: An essay in historical anthropology*. Cambridge, Cambridge University Press.

Oppenheimer, S.; Richards, M. 2001. Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Science Progress 84*: 157–181.

Pawley, A. 2002. The Austronesian dispersal: Languages, technologies and people. Pp. 251–274 in Bellwood, P.; Renfrew, C. (Eds) *Examining the farming/language dispersal hypothesis*. Cambridge, McDonald Institute for Archaeological Research.

Soares, P.; Trejaut, J.A.; Loo, J-H.; Hill, C.; Mormina, M.; Lee, C-L.; Chen, Y-M.; Hudjashov, G.; Forster, P.; Macaulay, V.; Bulbeck, D.; Oppenheimer, S.; Lin, M.; Richards, M.B. 2008. Climate change and postglacial human dispersals in Southeast Asia. *Molecular Biology and Evolution 25*: 1209–1218.

Thurgood, G. 1999. *From ancient Cham to modern dialects: Two thousand years of language contact and change*. Hawaii, University of Hawaii Press.