How useful are lexicostatistical and phylogenetic methods, in plotting the migration of Polynesian peoples across Oceania?

Noah Atkin[1]

[1]Imperial College London

April 30, 2018

Abstract

The importance of linguistics in the early study of Polynesian migration cannot be overstated. The work of many early linguists traced Polynesian languages back to their origins in South East Asia, and further linguistic analysis showed the Austronesian language family stretched from Madagascar to Easter Island. Historically, Polynesia has been an area where new fields of linguistics are applied, including glottochronology. In the modern day, the historical isolation of Polynesian sea-faring peoples provides a unique opportunity to test theories on early Austronesian migration by applying Bayesian and parsimony based phylogenetic techniques to the Oceanic languages. In this paper I hope to explain how languages were classically examined, and the extent these methods and modern day techniques, have played a role in elucidating the path of migration of these Oceanic sea-faring peoples.

Classical linguistic research methods

The earliest classical linguistic techniques for studying languages were versions of the comparative method. The comparative method attempts to discover the linguistic relatedness of languages by assembling word lists of the languages you wish to compare, and trying to establish which words are cognates of each other (words of common descent). From the changes observed between cognates, you try to construct a proto-language, and from that, a family-tree of the language family's evolution and relationships.

1

This comparative method was first successfully used to construct a useful proto-language in 1819 by Jacob Grimm and his colleague Rasmus Rusk; and they attempted reconstructing Proto-Indo-European (PIE) from the differences in Germanic languages and non-Germanic languages. This research led to the creation of Grimm's Law, a series of 3 statements describing how voiced stops in PIE became voiceless in proto-Germanic (PG); how voiceless stops in PIE became fricatives in PG and how aspirated stops in PIE changed to non-aspirated stops (or fricatives) (Grimm, 1819). While these laws are not without exception, they were the first a major turning point in the field of linguistics, and one of the first examples of a sound law describing how phonemes evolve.

The comparative method is intrinsically genealogical, and many modern day linguists consider the tree model implied as overly simplistic, and express doubts that biological comparison is misleading (François, 2014). A tree model implies a series of distinct nodes, which in turn implies sudden and irrevocable change in a population preventing further contact ie many different proto-languages coexisting without further interaction. This idea is clearly misleading in areas of continuous landmass (Argunova, 1994). Distinct nodes also imply linguistic uniformity within a proto-language, despite the existence of dialects within even small language communities (However the real life implications of this simplification are doubted) (Campbell, 2004). Furthermore critics of the tree model claim it ignores situations where many dialects within a language evolve into distinct languages, over a long period of time where innovations are shared (known as areal diffusion). Populations bordering others borrow their neighbours linguistic features and tend to share similar cultures(P'iankova, 1996) - influencing linguistic change in a shared direction, this is known as linkage. However in island populations, separated by vast distances and only limited contact with each other, these issues are largely mitigated.

While the comparative method can imply relatedness, early comparative work was flawed by an emphasis of *quantity* of words that should be collected from each language. The demand for more words led to poor datasets, incorrect cognate pairings and due to cultural difference, sometimes a lack of appropriate cognates (Ray, 1911; Elbert, 1953). This problem with linguistics was tackled in 1952 courtesy of Morris Swadesh, with his eponymous Swadesh list. The list comprised of 215 basic words that he believed would be common to every language, regardless of culture. Swadesh's belief that these words were non-cultural led him

to the idea that they would evolve at the same rate regardless of language or population, due to their universal importance in communication. Originally, the universality of the words was purely based on his own opinion (Makihara, 2006), however later revisions narrow down the list to just 100 words, and would comprise of 4 main types of words: Basic nouns (including familial terms, objects found in nature and anatomical parts), basic verbs (including basic actions and reactions), basic adjectives (colours & relative temperatures) and pronouns (you, I, they, etc.) (Heggarty, 2012). Over time, the adoption of the Swadesh list marked a change in linguistics, emphasising the importance of a smaller selection of high quality data, which in turn was more likely to be etymologically homologous.

While the comparative method and the Swadesh list can imply relatedness, on their own they don't offer a way to determine absolute branch lengths or the age of a language family. However Swadesh's idea of a constant rate of word evolution of his "basic vocabulary items" led to the development of glottochronolgy - a field of linguistics which attempts to date the most recent common ancestor of a pair of languages. The first attempt to accurately date a language split was by Swadesh himself in the paper *Salish Internal Relationships*, and quantified the elapsed time since a language split with the following equation:

$$i \; = \; \frac{\log\,(c)}{2\cdot\log(r)}$$

Where c is the percentage of shared cognates, i is the indicated period of depth in time (an arbitrary unit) and r is the percentage of basic vocabulary retained after 1 period. r was estimated in (Swadesh, 1950) as 0.85 by comparing contemporary English to old English (a time difference of roughly 1000 years). This constant was formalised and re-estimated in 1953 by Robert Lees in *The Basis of Glottochonrology* (Lees, 1953). Now coined the glottochronological constant, Lees re-estimated the constant by using a 1000 year comparison between 13 languages and generating a mean of the result: $0.80484\pm.0176$ (with a 90% certainty). i was then replaced by t, now indicating a 1000 year time difference, not an arbitrary unit.

$$t \; = \; \frac{\log(c)}{2\log(r)}$$

The field of glottochronology was contentious since its birth, and many anthropologists doubt the effectiveness of the field in absolute dating, however glottochronology has been showed to be effective at establishing chronology of migratory events, especially on isolated populations. (Lees, 1956; Hirsch, 1954)

History of the study of of Polynesian languages

Polynesia is a section of Oceania, consisting of over 1,000 islands, forming a triangle stretching from New Zealand to Hawaii to Easter Island, known to be colonised by humans for hundreds of years. Since their discovery to the western world, the question was posed how they arrived there. Early hypotheses suggested a migration from South America, and while some modern genetic evidence suggests limited peopling of Easter Island by Native Americans (Thorsby, 2016), it's clear today that Polynesian migration started from the west. Historically it was the naturalist Joseph Banks who began to discover the breadth of the language family Polynesian languages belong to. He recorded vocabulary used from Polynesia, Micronesia and South East Asia (Banks, 1790), and was able to draw a linguistic relationship by directly comparing Polynesian languages, which showed no clear difference, and then applying this comparative method to the examination of Polynesian and Micronesian languages, and follow the relationship back to South East Asia. This relationship was discovered to have spread to Madagascar, forming what we now know as the Austronesian language family. Given the discovery of the widest spread language family of all time, linguists tried to extrapolate the path of migration of these people using known data.

One of the earliest attempts at this migratory plot was by William Churchill, who compared Melanesian and Polynesian languages to test the two prevailing theories of Polynesian origin at the time, and to attempt to plot migratory routes through East Asia (Churchill, 1911) (See Figure 1.). The sieve theory posited that the islands of Melanesia and west Polynesia acted like meshes in a sieve, catching seafarers originating from central Polynesia who were blown westwards by the strong prevailing winds (known today as the "roaring forties") . This theory also posited that the initial peopling of central Polynesia occurred by seafarers from East Malaysia, who travelled with the current north of New Guinea, through the Marshall

Islands to Samoa and Fiji. The migration theory on the other hand suggested that the migration began in India, and they journeyed eastwards through the Malay Archipelago, travelling slowly (and suggests of many generations between migrations), they eventually reached Fiji, and settled there for generations before voyaging again. Churchill tested these theories by comparing cognate lists of 18 languages, and comparing word differences and language differences, he discovered that the linguistic features of Melanesian islands share many deep commonalities with the Polynesian languages, while Marshallese (which according to the sieve theory will have been one of the oldest Polynesian languages) lacks many of these features, and has many more unique features. This was the first major linguistic work showing the migration to Polynesia occurred through Melanesia, not Micronesia. He also suggested the migration occurred in two ”*swarms*“, the first travelled north of Papa New Guinea, between the Bismarck Archipelago and Solomon Islands until settling in Samoa; while the second ”swarm”travelled south of New Guinea, through the Torres Straits and settled in Fiji (used interchangably with Viti, as Viti is the largest island of what we know call Fiji). His work however, even by the standards of the time was very flawed, as pointed out in a review of his book in Nature: Churchill compared languages with vastly different lengths of word lists (of varying quality), leading to false conclusions (Ray, 1911). Later linguists took issue with Churchill's work as there was no attempt was at proto-language reconstructions and Churchill was notoriously poor at identifying cognates and was often fooled by analogous terminology and performed no phoneme-grapheme correspondences (letter-sound correspondences) (Elbert, 1953). All that being said, Churchill's work was an enormous step forward for linguistic based migratory study into Polynesia.
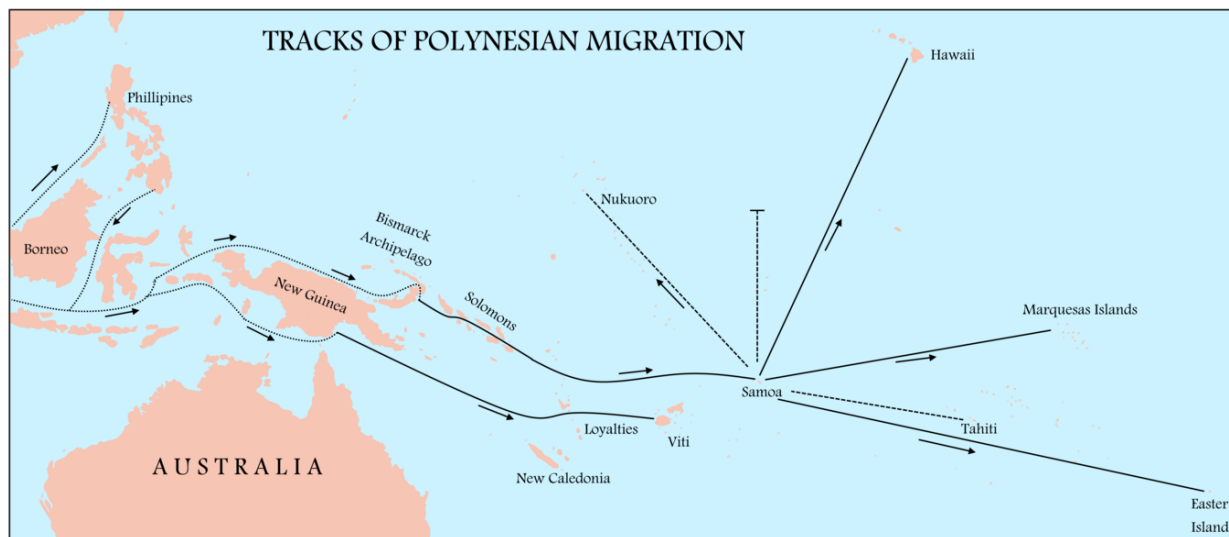
Figure 1: A recreation of William Churchill's predicted migratory routes through Melanesia, created from the data published in The Polynesian Wanderings (1911).

While work concerning the migration of Polynesian peoples slowed following Churchill's work, linguistic research into the grouping of the Oceanic languages flourished. Otto Dempwolff, a German linguist, was the first to hypothesise that all Oceanic languages may have originated from the same proto-language, which he named *Urmelanesisch*, and stressed its place as a distinct subgroup within the Austronesian language family (Ross, 2001). In his work *Comparative Phonology of Austronesian Word Lists*, Dempwolff outlined a series of novel features present in the Oceanic sub-group *not* present in the larger Austronesian family, including: A series of phonemic merges (a loss of distinction between phonemes over time) and final-consonant loss in a large number of Proto-Oceanic words (that are present in Proto-Austronesian). Dempwolff believed these changes, as well as other novel features present in Oceanic languages, were most plausibly due to changes in a shared proto-language, rather than convergent language evolution(Dempwolff, 1934).

Following Dempwolff's hypothesis of the Oceanic subgroup within the Austronesian language family, later linguists tried to establish the internal relationships of the Oceanic languages within Demowolff's subgroup. The most notable early endeavour at this was by Samuel H. Elbert, who attempted linguistically prove a divide between East and West Polynesia, as well as build a rough family tree of the languages of Polynesia (Elbert, 1953) (Although the existence of a East/West Polynesia divide has been known based on cultural evidence, largely religious differences between Samoa and Tahiti (Burrows, 1938)). To test his the-

ories, Elbert built a 202 word long Swadesh list, modified for cultural differences between Polynesia and Eurasia (removal of Eurasian animals, addition of particular foods and fruit, etc.). A standard glottochronological method was then followed: Establishing cognates and calculating percentage of linguistic homology. These percentages were then used to create a rough family tree for the Polynesian languages, which has been recreated in Figure 2.
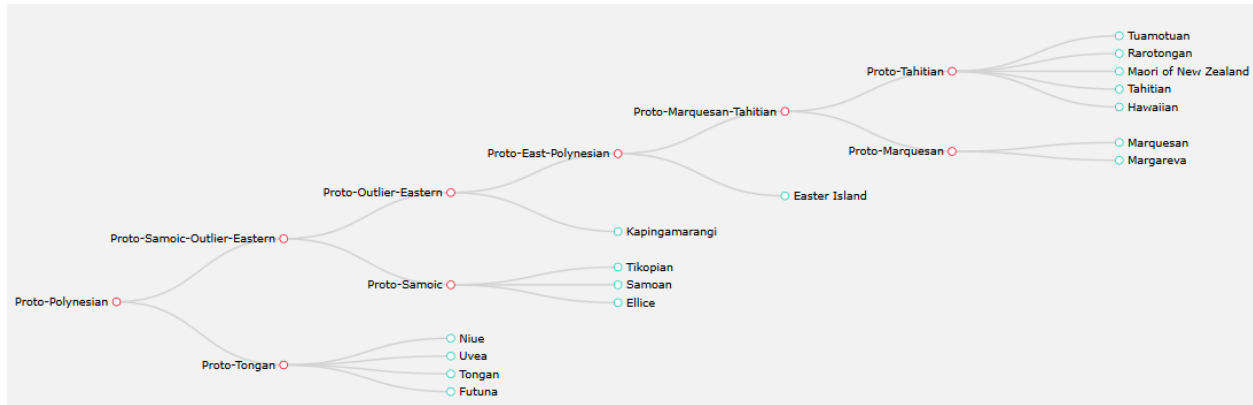
Figure 2: Recreation of Elbert's Polynesian language family, using data from *Internal Relationships of Polynesian Languages and Dialects* . Visualised using multitree.org (MultiTree, 2013)

A lot of Elbert's conclusions were somewhat tentative (due to suspected inaccuracies in his word list (Emory, 1963)), but his main conclusions were: There was a long stay after migrants arrived in West Polynesia, in the order of centuries, before migration continued leading to the linguistic split between East and West Polynesia. This was supported with later work by George W. Grace's *Subgrouping of Malayo-Polynesian: A Report of Tentative Findings.* The report, undertaken by the Tri-Instituional Pacific Program, compiled a list of 427 common English words (including the earliest Swadesh list), and 180 specialised words (to be used to identify culture contact and language diffusion), from these they assembled cognate comparisons for all known languages in the Oceanic region and attempted to find phylogeny. The large number of language families (10) which originated in New Guinea, and a further 4 languages from the Solomon Islands (Grace, 1955) emphasised the split between the migratory Polynesian peoples (who all shared one language family), and the settled Melanesians.
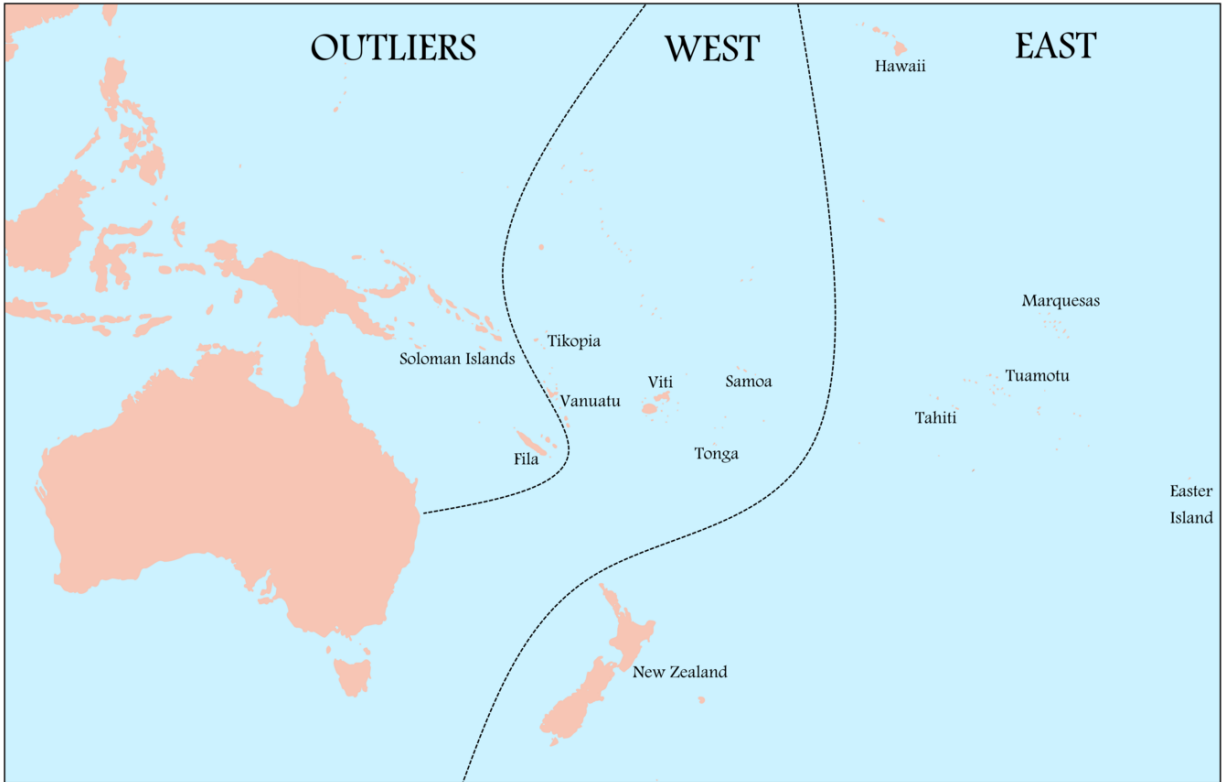
Figure 3: Map displaying the divisions of Polynesia derived by Elbert in his 1953 paper *Internal Relationships of Polynesian Languages and Dialects*

Following Elbert's attempt at dating Polynesian migration, it was Kenneth P. Emory who was finally able to map the migratory routes from West Polynesia (which he designated as Samoa and Tonga) to East Polynesia (Tahiti, the Marquesas, Easter Island, New Zealand, Mangareva and Hawaii.), as well as establish the rough chronology of these migratory events, and their origins (Emory, 1963). Emory, through his previous work and experience recognised the languages of Polynesia are very stable, and have a very similar vocabulary; because of this, glottochronological dating of Polynesian languages is more utilitous than it is when comparing languages over a continuous landmass (Emory, 1959).

Emory used a 100 word vocabulary list, which was a derivation of the Swadesh list with appropriate cultural substitutions (exclusion of Eurasian mammals & words without direct translations) and created reconstructions of proto-Polynesian and proto-East Polynesian to perform his analysis on. His results were stark - They displayed a very strong linguistic similarity between the languages within East Polynesia, and a strong similarity between

the languages within West Polynesia. As well as showing a comparatively vast difference in similarity between the East Polynesian and West Polynesian language families. The cause of the split, he inferred, was due partly to Elbert's theory of a long Eastern migratory pause (which Emory estimated was at least 500 years long, but admits could be as high as 700 years); however Emory also calculated a migratory pause in West Polynesia before migration to East Polynesia began, and this pause he estimated to be around 500 years as well. Emory's paper also confirmed one of Elbert's theories: That Easter Island was one of the first islands to be inhabited after migration from the East began, contrary to the popular opinion of the time (Parmentier, 1998). His research was able to establish a rough route of migration, which is pictured in Figure 4. with dates of arrival:
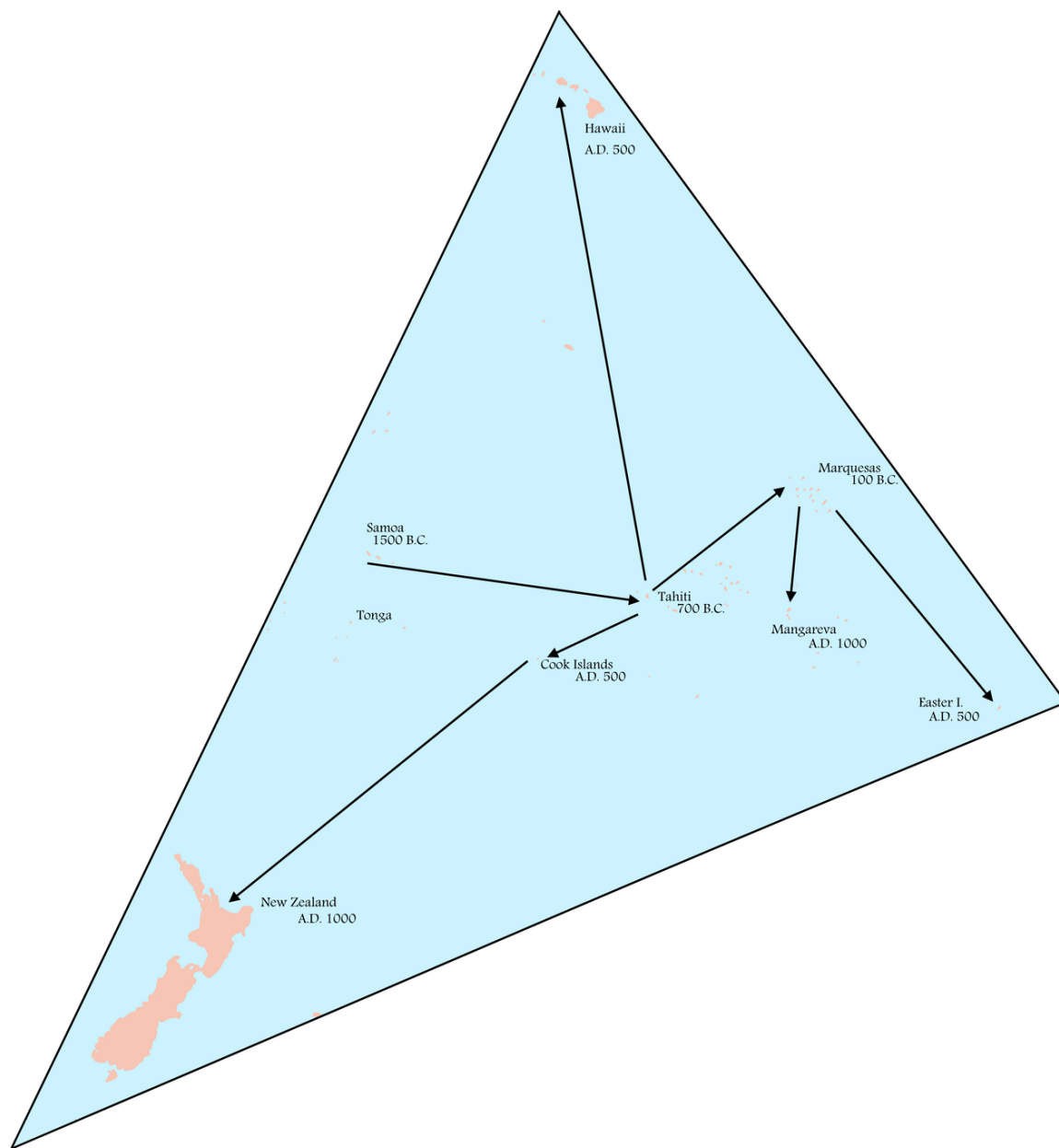
Figure 4: Polynesian settlement pattern, inferred from linguistic data in *East Polynesian Relationships: Settlement Pattern and Time Involved as Indicated by Vocabulary Agreements*

In terms of glottochronological data, this was the pinnacle of classical data that was established on Polynesian migration. Classical lexicostatistics is too inaccurate to provide better estimates on arrival dates, only the order of events. However the Polynesian question, such as their ultimate place of origin was still left unanswered. New methods would be needed to

see where and when the migration began.

Genetic linguistic research methods

Biologists have leaned into the similarities between linguistics and genetics, even Darwin commented on the similarity between the two fields (Darwin, 1872). Biologically the tree model implied by the comparative method can be approached by maximum parsimony and Bayesian methods. Maximum parsimony is the method of resolving a phylogeny by selecting the tree requiring the fewest number of evolutionary events/changes, and these methods are useful for simplistic methodology such as (Gray & Jordan, 2000), but struggle to cope with the large number of variables complex linguistic questions require, and so complicated lexicogeographic problems are being solved with Bayesian methods.

Bayesian Phylogenetics

Modern day linguistical phylogenetic methods are usually based on Bayesian Markov Chain Monte Carlo (MCMC) methods, the full depth of these methods is beyond the scope of this paper, but we can establish the foundations of the method, and how it can be applied to phylogenetics.

To explain it we will start with Bayes' theorem:

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Figure 5: Bayes' Theorem

Bayes' Theorem is a way of calculating the probability of two competing binary hypotheses, by considering previous data. The various components of Bayes' theorem can be more easily explained with an example. Imagine there's a new disease that's emerged called Polynesian

12

Fatigue (PF), only 1% of a population 10,000 strong has PF, but luckily our new test is 95% effective. How worried should you be if you test positive for PF?

Bayes' theorem can calculate the probability of a hypothesis (that you have PF), given an event (that you tested positive for PF), which we can write as follows.

$$P(PF \mid +)$$

To calculate this we need to take the *prior probability* that the hypothesis is true *before* you took the test, which is the likelihood that you have the disease at all (a good estimate for this is the frequency of PF sufferers in the population, 1%) and multiply that by the probability of a positive test given the hypothesis is true, which we call the *likelihood function*. Our likelihood function is that out of a population of 10,000, we would expect 100 people to have PF, and we would expect 95 of them to test positive on the test, 0.95). We then need to divide all that by the total probability of a positive test (which encompasses both the probability that you have PF, and you've tested positive; and the probability that you *don't* have PF and you have falsely tested positive), we call this the *marginal likelihood*. Which gives us:

$$P(PF \mid +) = \frac{P(+ \mid PF) \cdot P(PF)}{P(PF) \cdot P(+ \mid PF) + P(-PF) \cdot P(+ \mid -PF)}$$

$$P(PF \mid +) = \frac{0.95 \cdot 0.01}{0.01 \cdot 0.95 + 0.99 \cdot 0.05}$$

$$P(PF \mid +) = 0.161$$

Meaning you only have a 16.1% chance of actually having the disease *after* you test positive, this is our *posterior probability.* However if you take another test for PF, and that also comes back positive, what's the odds you have the disease then? This time, however, we update our *prior probability*, it's no longer the level of PF present in the population (0.01), it's 0.161, we've adjusted our prior probability using data from our previous test.

$$P(PF \mid +) = \frac{0.95 \cdot 0.161}{0.95 \cdot 0.161 + 0.05 \cdot 0.839}$$

$$P(PF \mid +) = 0.785$$

From this we can update our likelihood of you actually having PF. However there may be reasons we doubt the certainty of our prior probability, estimates for the levels of PF present in the population vary from hospital to hospital, and so you want to factor that uncertainty in. Instead of a prior probability, we can use a *prior probability distribution,* which we can then use to calculate a *posterior probability distribution.* A prior distribution can be informative, for example a collation of all the data from the hospitals accurately placed into a distribution, it can be weakly informative ie it only loosely contains the data, an example of this would be if we only had data from half the hospitals, and so created a distribution which allows for more uncertainty; or it can be uninformative. Uninformative is sadly a misnomer as it always has some relevant information (e.g an equal probability distribution where $0 < P(PF) < 1$), however in Bayesian statistics the more informative your prior is, the better your result will be (As a less informative prior places more emphasis on the likelihood function, which is more akin to frequentist statistics). However an uninformative prior is not that damaging if you run the test many times, as with our example, as you run more tests, you can be more and more certain of the result.

Markov chain Monte Carlo

Phylogenetic applications of Bayesian inference are very complicated and involve many parameters, and as the number of parameters grows the marginal likelihood becomes more and more complicated to process analytically. Luckily using Markov Chain Monte Carlo (MCMC) we can estimate the posterior probability distribution, without having to calculate the marginal likelihood.

The most commonly used version of an MCMC method in phylogeny is the Metropolis-Hastings algorithm. The mathematics of the method is complicated but it involves taking many samples from the posterior probability distribution over a period of generations until you can approximate the posterior probability distribution, *without* having to calculate the distribution directly. The algorithm begins with a randomly selected tree, which is well defined. If little is initially known about the phylogeny before computation a sensible prior would be uniform probability distribution for all available trees. This tree is compared to one of the other possible trees by the ratio of the new tree over the old (This is how the marginal likelihood is excluded, as it cancels during division). If the new proposed tree has

a greater prior x likelihood function it is selected (ie ratio > 1), it replaces the old tree as the accepted estimate for the posterior probability, and a new tree is chosen for comparison. As the Markov chain usually begins with a randomly chosen value, the first ~10% of generations are discarded as "burn-in" until the algorithm "forgets" its initial tree. But over thousands of generations, and hundreds of samples an accurate tree can be selected with some certainty, due to the prior probability for the most likely tree being updated.

Parsimony, Bayesian phylogenetics & Polynesian migration

Phylogenetic methods are particularly useful in the field of Polynesian language evolution due to the unique migratory lifestyle of Polynesian peoples. This largely eliminates some of comparative method criticism of tree modelling as contact between distal Polynesian islands was sporadic, and in the case of Easter Island, non existent(Drechsel, 2014). As noted before there is a striking similarity of the vocabulary in the Polynesian languages, which also assists in the effectiveness of biological models. As migration through Polynesia has largely been established, and most biological applications of linguistic phylogeny in Polynesia apply to finding the homeland of the Polynesian peoples, and the speed of their migration through Melanesia.

Current theories of Polynesian people's origin fall into two main camps: Express Train theory and the Entangled Bank theory (Kayser, Brauer, Weiss, *et al.*, 2000). The more popular theory, Express Train, first put forward by Jared Diamond (Diamond, 1988) posits that Polynesian migration originated in Taiwan and started relatively recently 3000-1000BCE. Migration took place via the Philippines and New Guinea, reaching Melanesia by roughly 1400BCE and reaching Samoa (Western Polynesia) by 900BCE. Entangled Bank theory, on the other hand, puts forward that there was no single 'express train' to Polynesia, it emphasises many smaller migration events, as well as the long cultural and genetic interactions between the Polynesians, Melanesians and East Asians. Some newer theories such as Kayser et al.'s Slow Boat theory attempts to marry these two ideas, while supporting a Tawainese origin and rapid migration to Melanesia, Slow Boat theory suggests that upon reaching Melanesia migration slowed, and there was a long delay (leading to cultural and

genetic admixture) before migration to Polynesia.

Biological models were first applied to the Polynesian migration question in the 2001 paper *Language trees support the express-train sequence of Austronesian expansion*. The author, Russell Gray, used a simple system of character mapping based on maximum parsimony to establish the sequence of colonisation events that led to the colonisation of Polynesia. Languages were given a specified character rank based on their geographic position, and that geographic positions relation to the suspected order of migratory events (Gray & Jordan, 2001).
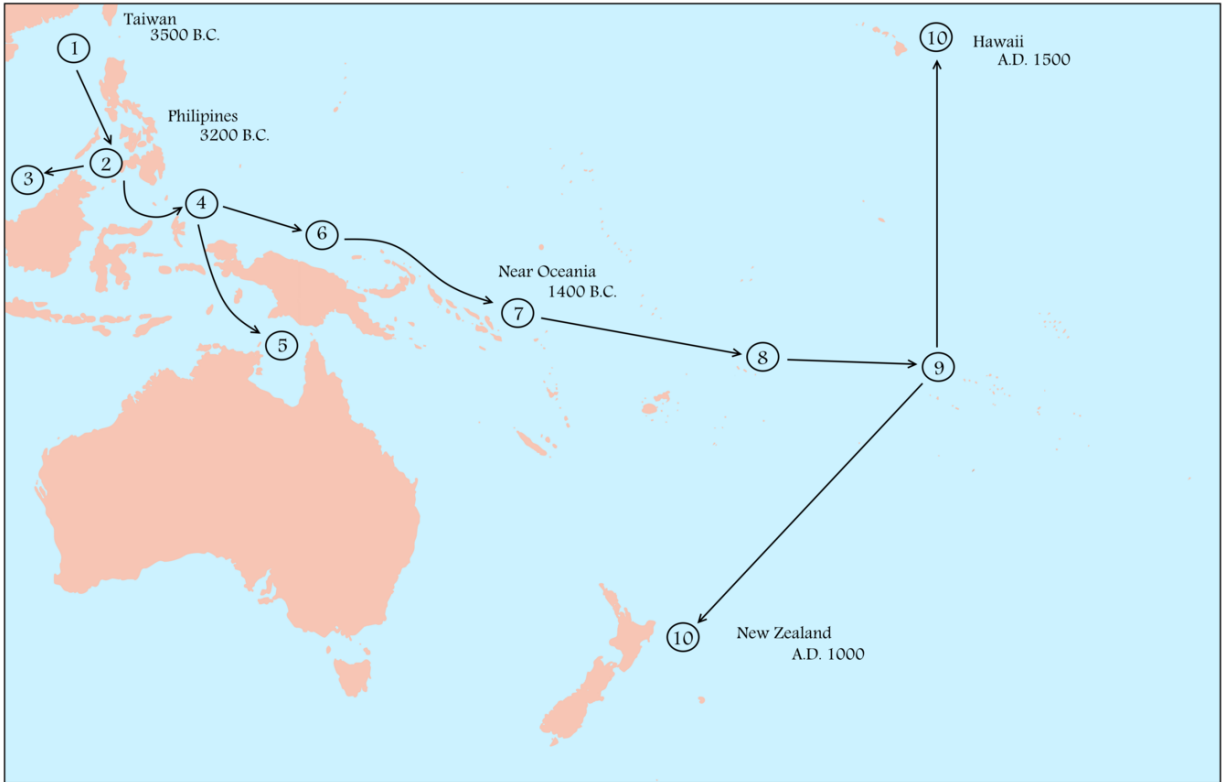
Figure 6: Gray's character map of the Express Train Model. Recreated with permission of the author.

A maximum parsimony tree was then generated based on these characters, and the number of steps required. A second maximum parsimony tree was generated using linguistic data from 77 Austronesian languages comprising of 5,185 morphemes. The character tree was then mapped onto the linguistically produced tree to examine fit; if the tree fit well, it suggests an Express Train model of migration. The tree generated had 13 steps, which in comparison to the maximum fit (9, ie n-1) and the number of steps expected by chance (49) showed significant support for the Express Train version of events. This paper was fairly simple, and due to the nature of the study was unable to model an effective route for the slow train model, however it displayed the effectiveness phylogenetic techniques on linguistic problems.

Gray expanded on his evidence for the Express Train model by using phylogenetic models based on Bayesian inference, and application of MCMC algorithms. *Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement* (2009) builds on lessons learnt in
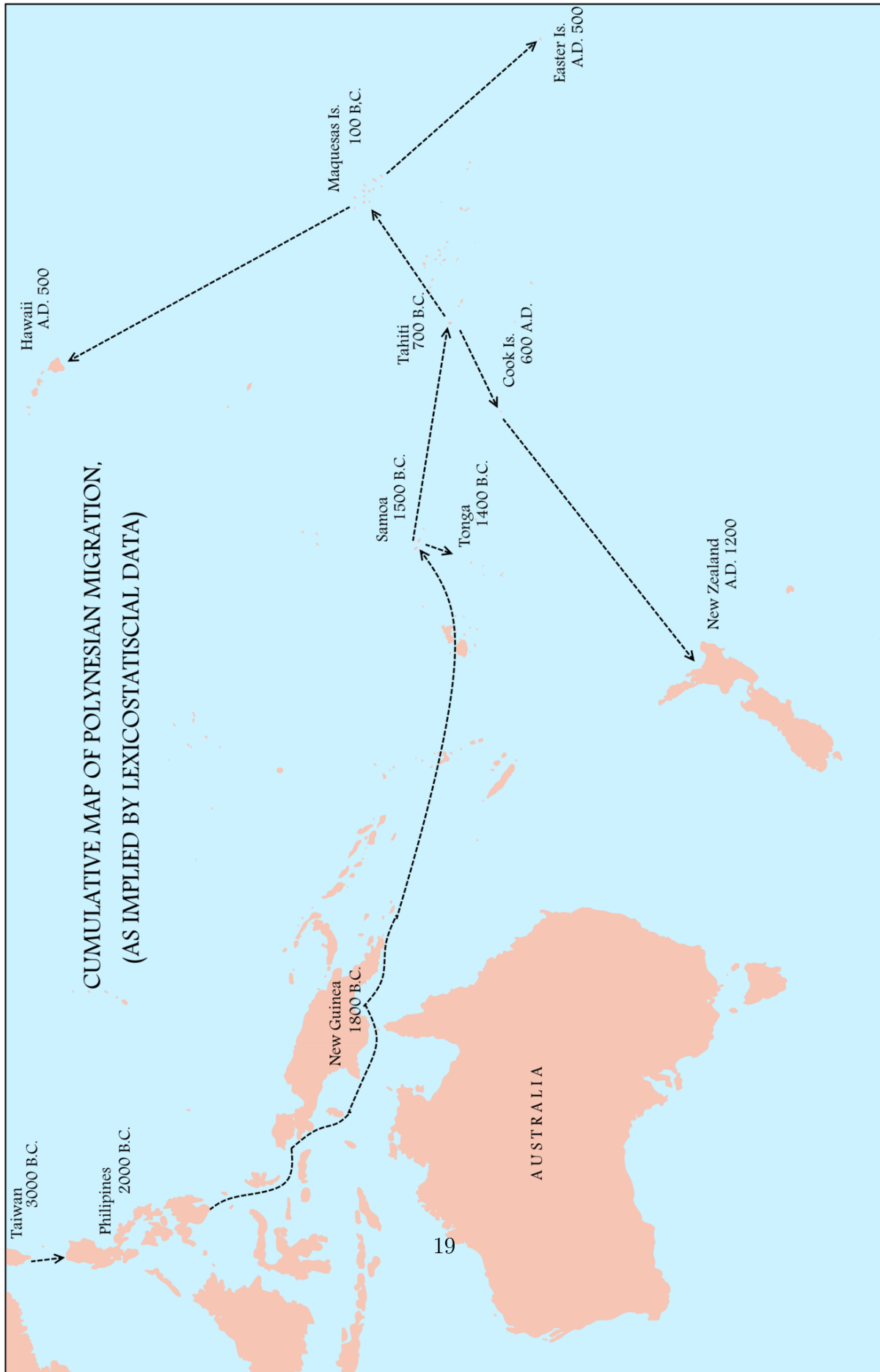
classical lexicostatistics and glottochronology: It uses a relatively short word list (210), and extensively curated it before modelling began. Linguists established cognates between 400 languages, and consulted independent linguists to confirm the accuracy of their list (90-93% accuracy, IRR=91%).

A consensus tree was selected using a Metropolis-Hastings based algorithm, from a posterior probability distribution comprised of 4,200 trees (language evolution was modelled as single cognate gain/loss). The consensus tree provided more evidence of Dempwolff's original theory of the singular oceanic language group. The consensus tree also agreed with Grace et al's positioning of the Polynesian languages nested within the Central-Pacific subgroup. The research provides evidence for a Taiwanese origin of migration through Melanesia, and pause/pulse theory of migration, with an initial refractory period before rapid expansion out of Taiwan through Melanesia, with a second pause occurring once the Polynesians arrived at Samoa. Gray estimated the age of the Austronesian language family to be 5230 years old, nearly 8000 younger than would be predicted with the Slow Boat model, adding more evidence to the Express Train theory of Polynesian expansion.

With the exception of these papers, little to no phylogenetic based work has been applied to Polynesian migration, however given the promising conclusions the papers have delivered, its possible that phylogenetic methods may be able to answer more and more mysteries about Polynesian expansion, migration and settlement.

Summary diagram

From all of these pieces of data, a cumulative map of Polynesian migration has been created:

CUMULATIVE MAP OF POLYNESIAN MIGRATION,

(AS IMPLIED BY LEXICOSTATISCIAL DATA)

Taiwan
3000 B.C.

Philipines
2000 B.C.

New Guinea
1800 B.C.

AUSTRALIA

Hawaii
A.D. 500

Maquesas Is.
100 B.C.

Easter Is.
A.D. 500

Tahiti
700 B.C.

Cook Is.
600 A.D.

Samoa
1500 B.C.

Tonga
1400 B.C.

New Zealand
A.D. 1200

References

Grimm, J. (1819) *Deutsche Grammatik.* [Online]. Erster Teil. Available from: doi:10.1515/9783110949421.102.

François, A. (2014) Trees waves and linkages. In: *The Routledge Handbook of Historical Linguistics.* [Online]. Routledge. Available from: doi:10.4324/9781315794013.ch6.

Argunova, T.V. (1994) The Interaction of Languages in the [Sakha] Republic. *Anthropology & Archeology of Eurasia.* [Online] 32 (4), 85–90. Available from: doi:10.2753/aae1061-1959320485.

Campbell, L. (2004) *Historical Linguistics: An Introduction.* MIT Press.

P'iankova, L.T. (1996) Bronze Age Cultures of the Steppe and Highlands in Central Asia. *Anthropology & Archeology of Eurasia.* [Online] 34 (4), 13–28. Available from: doi:10.2753/aae1061-1959340413.

Ray, S.H. (1911) "The Polynesian Wanderings". *Nature.* [Online] 87 (2186), 382–382. Available from: doi:10.1038/087382a0.

Elbert, S.H. (1953) Internal Relationships of Polynesian Languages and Dialects. *Southwestern Journal of Anthropology.* [Online] 9 (2), 147–173. Available from: doi:10.1086/soutjanth.9.2.3628573.

Makihara, M. (2006) Swadesh Morris (1909–1967). In: *Encyclopedia of Language & Linguistics.* [Online]. Elsevier. pp. 303–304. Available from: doi:10.1016/b0-08-044854-2/02915-1.

Heggarty, P. (2012) Beyond lexicostatistics. In: *Benjamins Current Topics.* [Online]. John Benjamins Publishing Company. pp. 113–137. Available from: doi:10.1075/bct.46.07heg.

Swadesh, M. (1950) Salish Internal Relationships. *International Journal of American Linguistics.* [Online] 16 (4), 157–167. Available from: doi:10.1086/464084.

Lees, R.B. (1953) The Basis of Glottochronology. *Language.* [Online] 29 (2), 113. Available from: doi:10.2307/410164.

Lees, R.B. (1956) Shiro Hattori on Glottochronology and Proto-Japanese. *American Anthropologist.* [Online] 58 (1), 176–177. Available from: doi:10.1525/aa.1956.58.1.02a00160.

Hirsch, D.I. (1954) Glottochronology and Eskimo and Eskimo-Aleut Prehis-

tory. *American Anthropologist.* [Online] 56 (5), 825–838. Available from: doi:10.1525/aa.1954.56.5.02a00070.

Thorsby, E. (2016) Genetic Evidence for a Contribution of Native Americans to the Early Settlement of Rapa Nui (Easter Island). *Frontiers in Ecology and Evolution.* [Online] 4. Available from: doi:10.3389/fevo.2016.00118.

Banks, J. (1790) *Vocabularies of the languages of Tahiti, Prince's Island, Sulu, Samarang, New Holland (i.e. Australia), Samarang, Savi, etc..*

Churchill, W. (1911) *The Polynesian wanderings; tracks of the migration deduced from an examination of the proto-Samoan content of Efate and other languages of Melanesia.* The Carnegie Institution of Washington.

Elbert, S.H. (1953) Internal Relationships of Polynesian Languages and Dialects. *Southwestern Journal of Anthropology.* [Online] 9 (2), 147–173. Available from: doi:10.1086/soutjanth.9.2.3628573.

Ross, T.C.J.L.M.D. (2001) *The Oceanic Languages.* [Online]. Routledge. Available from: doi:10.4324/9780203820384.

Dempwolff, O. (1934) *Comparative phonology of Austronesian word lists.* D. Reimer.

Burrows, E.G. (1938) *Western Polynesia: A Study of Cultural Differentiation.* University Book Shop.

MultiTree (2013) *Proto-Polynesian: Elbert 1953.* [Online]. Available from: http://new.multitree.org/trees/id/24157 date of access = 14/04/2018.

Emory, K.P. (1963) East Polynesian Relationships: Settlement Pattern and Time Involved as Indicated by Vocabulary Agreements. *The Journal of the Polynesian Society.* 72 (2), 78–100.

Grace, G.W. (1955) Subgrouping of Malayo-Polynesian: A Report of Tentative Findings. *American Anthropologist.* [Online] 57 (2), 337–339. Available from: doi:10.1525/aa.1955.57.2.02a00110.

Emory, K.P. (1959) Origin of the Hawaiians. *The Journal of the Polynesian Society.* [Online] 68 (1), 29–35. Available from: http://www.jstor.org/stable/20703707.

Parmentier, R.J. (1998) Rongorongo The Easter Island Script: History, Traditions,

Texts.:Rongorongo, The Easter Island Script: History, Traditions, Texts.. *Journal of Linguistic Anthropology.* [Online] 8 (2), 253–255. Available from: doi:10.1525/jlin.1998.8.2.253.

Darwin, C. (1872) *The descent of man and selection in relation to sex. By Charles Darwin..* [Online]. D. Appleton and company, Available from: doi:10.5962/bhl.title.2112.

Gray, R.D. & Jordan, F.M. (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature.* [Online] 405 (6790), 1052–1055. Available from: doi:10.1038/35016575.

Drechsel, E.J. (2014) *Language Contact in the Early Colonial Pacific.* [Online]. Cambridge University Press. Available from: doi:10.1017/cbo9781139057561.

Kayser, M., Brauer, S., Weiss, G., Underhill, P.A., et al. (2000) Melanesian origin of Polynesian Y chromosomes. *Current Biology.* [Online] 10 (20), 1237–1246. Available from: doi:10.1016/s0960-9822(00)00734-x.

Diamond, J.M. (1988) Express train to Polynesia. *Nature.* [Online] 336 (6197), 307–308. Available from: doi:10.1038/336307a0.

Gray, R.D. & Jordan, F.M. (2001) correction: Language trees support the express-train sequence of Austronesian expansion. *Nature.* [Online] 409 (6821), 743–743. Available from: doi:10.1038/35055596.

Gray, R.D., Drummond, A.J. & Greenhill, S.J. (2009) Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science.* [Online] 323 (5913), 479–483. Available from: doi:10.1126/science.1166858.

Walker, R.S. & Hamilton, M.J. (2010) Social complexity and linguistic diversity in the Austronesian and Bantu population expansions. *Proceedings of the Royal Society B: Biological Sciences.* [Online] 278 (1710), 1399–1404. Available from: doi:10.1098/rspb.2010.1942.

Diamond, J. (2003) Farmers and Their Languages: The First Expansions. *Science.* [Online] 300 (5619), 597–603. Available from: doi:10.1126/science.1078208.

Grace, G.W. (1955) Subgrouping of Malayo-Polynesian: A Report of Tentative Findings. *American Anthropologist.* [Online] 57 (2), 337–339. Available from: doi:10.1525/aa.1955.57.2.02a00110.