**Pacific Linguistics 601**

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.



# Austronesian historical linguistics and culture history: a festschrift for Robert Blust

Edited by
Alexander Adelaar and Andrew Pawley

Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University

# 22 *Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods*

SIMON J. GREENHILL and RUSSELL D. GRAY [1]

Historical linguistics has never been particularly intimate with computers. The first wave of computational historical linguistics—lexicostatistics—was developed in the 1950s (Swadesh 1952; Lees 1953) and quickly applied to language groups around the world from Indo-European to Austronesian (Lees 1953; Hymes 1960; Embleton 1986). However, critics were quick to point out the problems caused by assuming a single constant rate of lexical replacement and repeatedly noted the erroneous results that this produced (Hoijer 1956; Bergsland and Vogt 1962; Blust 1981; McMahon and McMahon 2006). As a consequence of these critiques lexicostatistics has been widely rejected by mainstream historical linguists (Campbell 2004).

The last few years have seen a second wave of computational approaches entering historical linguistics: phylogenetic methods. These techniques, drawn from evolutionary biology, have been used to investigate some provocative and controversial claims about human prehistory. For example, we have applied phylogenetic methods to lexical data compiled by Bob Blust to test hypotheses about the settlement of the Pacific (Gray and Jordan 2000; Greenhill and Gray 2005; Gray et al. 2009). Our results reflected a settlement pattern through Island South-East Asia, New Guinea and then into Oceania, consistent with the 'Out of Taiwan' scenario (e.g. Blust 1999; Pawley 2002; Diamond and Bellwood 2003). We have also used these methods to investigate the origins of the Indo-European (Gray and Atkinson 2003) and Bantu languages (Holden 2002; Holden and Gray 2006). Other groups have applied phylogenetic methods to investigate the internal subgrouping of these families (Ringe et al. 1998; Rexovà et al. 2003, 2006). The application of

[1] We would like to thank Andreea Calude, Andy Pawley, and Malcolm Ross for comments on this paper. We would like to note that some of the analyses reported in this paper have been superseded by those conducted using a better fitting model of cognate evolution reported in Gray et al. (2009).

Alexander Adelaar and Andrew Pawley, eds.
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 375–397.
Pacific Linguistics, 2009.

computational phylogenetic methods has not been restricted to just lexical data. Phylogenetic analyses of structural features have revealed historical signals in Papuan languages that may stretch back around 10,000 years (Dunn et al. 2005). Nor have phylogenetic methods been restricted to just building trees. Phylogenetic network methods have been used to investigate conflicting signals in Indo-European (Bryant et al. 2005), Bantu (Holden and Gray 2002), Chinese dialects (Hamed and Wang 2004), and Polynesian (Bryant 2006; Gray 2007). Finally, phylogenetic methods have recently been used to investigate general claims about the factors that affect the rate of language change. Pagel et al. (2007) used phylogenetic methods to estimate the rates of lexical replacement in Indo-European languages and showed an almost hundred-fold difference between the rates of rapidly evolving words (e.g. 'dirty') and the slowly evolving words (e.g. 'tongue'). They then calculated the frequency at which these words were currently used in four large language corpora. Their results showed a strong correlation between the frequency with which words are used today and their stability over time: the more a word is used, the slower it evolves. This striking result suggests that over the 9000 years of Indo-European language history, there have been consistent underlying mechanisms controlling lexical replacement. A second study (Atkinson et al. 2008) used phylogenetic methods to test claims that speakers often use their language as a social tool for increasing group cohesion and demarcating groups (Labov 1994). The results showed a strong relationship between the total amount of lexical change and the number of language splitting events along the tree: between 10% to 33% of the total lexical change in the Bantu, Indo-European, and Austronesian languages occurred as a rapid burst of change shortly after languages diverged. This punctuational change (e.g. Bowern 2006) is consistent with rapid language change in small founder populations and differentiation as a cultural marker.

Given the combination of strong claims, new techniques, and the high-profile reporting of results, it is not surprising that these studies are often controversial. Responses have ranged from the positive: 'Computational methodologies of this kind can only be helpful for historical linguistics' (April McMahon in Balter 2003:1491), to the skeptical: 'There is no reason whatsoever to assume that vocabulary would behave the same way that organisms do.' (Alexander Lehrman in Balter 2004:1326), to the negative: '... have ignored the fatal shortcomings of glottochronology ...' (Eska and Ringe 2004:569), and the painfully incorrect; 'sledg(ing) the dead horse of the Swadesh algorithm' (Holm 2007:201).

Sadly many of these criticisms are mired in misunderstanding. Computational phylogenetic methods are not just lexicostatistics redux, but a powerful supplement to the comparative method used in historical linguistics. On several occasions Bob Blust has challenged us to specify exactly how phylogenetic methods differ from lexicostatistics and explain why they are superior. Here we respond to his challenge. To do this, we will focus on one of the great battlegrounds between lexicostatistics and the traditional comparative method: the Austronesian language family. First, we will describe how Bayesian phylogenetic methods work, and then give a step-by-step explanation of an analysis of a large lexical dataset for 400 Austronesian languages (Gray et al. 2009; Greenhill et al. 2008).

## 1 The Austronesian language family

The Austronesian language family is one of the two largest in the world, containing around 1000 to 1200 languages (Gordon 2005). Before Columbus, these Austronesian languages were also the most widely dispersed with speakers in Mainland and Island South-East Asia, Madagascar, Micronesia, Melanesia, and Polynesia (Bellwood et al. 1995). The groundwork that identified this family began in the 16th and 17th centuries as European scholars began to compare word lists that trickled back from early explorers and missionaries (e.g. Houtman 1603; Reland 1708; Forster 1778; Brandes 1884; Kern 1886). Dempwolff (1934, 1938) systematically reconstructed early Austronesian phonology and lexicon, and identified a large subgroup, Oceanic, to which he assigned the languages of Melanesia, Polynesia and (most of) Micronesia (Dempwolff 1937). The evidence that all these Oceanic languages formed a subgroup of Austronesian implied that they stem from a single Austronesian settlement of this region from the west (Grace 1961, 1964a; Pawley and Green 1973; Pawley and Ross 1995).



A major challenge to this hypothesis came from Dyen's lexicostatistical analyses of vocabulary from 352 Austronesian languages (1962, 1965). Lexicostatistics had previously been applied to subgroups within Austronesian (an early paper by Elbert (1953) explored Polynesia), but Dyen's was by far the largest in scale. At the time, Dyen's analysis was an impressive computational feat; his program compared 7,000,000 pairs of words. The lexicostatistical results suggested a tree with 40 first-order branches, no fewer than 30 of which were located in Melanesia. Dyen took this to indicate that the most probable area of origin of the Austronesian languages was in Melanesia, possibly in the Bismarck Archipelago north of New Guinea, with subsequent expansions east into Polynesia, and west into Indonesia then to the Philippines and Taiwan. This study was hailed by Murdock (1964:117) as '... a significant work—one which may conceivably be as revolutionary for Oceanic linguistics and culture history as was the work of Greenberg (1949–54) for the interpretation of African languages and cultures'.

This enthusiasm was short-lived. Grace (1964b, 1966) was quick to to suggest that the difference between the lexicostatistical view of Austronesian relationships and that of the traditional view may be a consequence of faster rates of lexical replacement in Melanesia. Blust (1981, 2000) quantitatively demonstrated that the Austronesian languages varied markedly in their retention rates across a 200-item basic vocabulary word-list. Retention rates in Malayo-Polynesian languages ranged from 5% to 60% in the interval between Proto Malayo-Polynesian and the present, a time period of around 4000 years. Moreover, Blust (2000) argued that the inability of lexicostatistics to discriminate between shared retentions and innovations—a distinction that had been critical in historical linguistics since Brugmann (1884)—exacerbated the effect of different rates. These differences in retention rates, especially in regions such as Melanesia where there have been high levels of language contact and borrowing (Ross 1996) rendered the lexicostatistical conclusions invalid.

In contrast to a Melanesian origin for Austronesian languages suggested by lexicostatistics, the comparative method has provided strong evidence that all languages outside Taiwan belong to a single sub-group (Dahl 1973; Blust 1977), which Blust (1977) named Malayo-Polynesian. In a series of publications Blust (e.g. 1977, 1978, 1982, 1999) marshalled a large array of evidence for the claim that the Proto Austronesian (PAn) homeland lay in Formosa (Taiwan). First, Blust (1999) concluded there are at least nine primary subgroups of Austronesian within Taiwan, whereas all Austronesian languages spoken outside of Taiwan fall into a single first order subgroup. There are a number of phonological and morphological innovations that are shared by the Malayo-Polynesian subgroup but are not found in the Formosan languages. If we assume that the region with the most primary subgroups is likely to be the primary dispersal centre Taiwan is thus strongly favoured as the Austronesian homeland. Blust (1982) also used the distribution of flora and fauna lexicon to delimit the range of possible Austronesian homelands. The

distribution of cognate words for placental and marsupial mammals in Austronesian languages suggests that ancestral Austronesian society was located in the Asiatic faunal zone to the west of the Wallace line. Archaeological evidence indicates that the spread of Neolithic cultures from Taiwan parallels the directions and dates of the Austronesian linguistic expansion. This conjunction of different lines of evidence has convinced most specialists in Austronesian historical linguistics that the Austronesian-speaking people were present in Taiwan around 5500 years ago, before spreading into the Philippines, Indonesia and through the Pacific (e.g. Shutler and Marck 1975; Bellwood 1997; Blust 1995; Kirch 2000; Kirch and Green 2001; Pawley 2002).

The failure of lexicostatistics to get Austronesian 'right' is not surprising—computing Austronesian language relationships is a very difficult problem. First, the rapid expansion of the Austronesian family means that it is likely to be difficult to resolve the fine branching structure of the Austronesian language tree as there is little time for the internal branches on the tree to develop numerous shared innovations (Pawley 1999). Second, as these languages moved across the Pacific they encountered new environments and the consequent need for new terminology may have increased the rates of language replacement. This acceleration in rates is likely to be exacerbated by the effects of language contact—particularly within Near Oceania (Ross 1996). Additionally, many Austronesian languages have small speech communities, which are also likely to speed up the rates of language evolution (Nettle 1999). The effects of these factors can be seen in the substantial variation in cognate retention rates in Austronesian languages (Blust 1981; 2000; Pawley this volume). Finally, the sheer scale of the Austronesian language family is daunting—with around 1000 to 1200 languages there are more than $10^{2864}$ possible rooted family trees. In the following section we will outline a Bayesian phylogenetic analysis on the Austronesian languages.

## 2 A phylogenetic approach

Much of biology and linguistics is historical. That is, to understand these systems properly we need to know their history. Where did particular languages or species come from? When did they arise and diverge? What sequence of changes took place? Are two characteristics similar because they share common ancestry or are they similar because they've evolved to fill the same function? To investigate these questions biologists have developed a large collection of tools collectively known as phylogenetics. Biologists initially constructed phylogenetic trees with clustering algorithms such as UPGMA ('Unweighted Pair-Group Method using Arithmetic averages', Sneath and Sokal 1963), that analysed pairwise similarity matrices (just like the lexicostatistical percentage shared cognacy matrices). Not surprisingly, this approach also produced inaccurate results when there were substantial differences in the rates of genetic change between lineages (Felsenstein 1978). However, rather than abandon a computational approach when confronted with this difficulty, biologists improved the computational methods. In the last few decades phylogenetic methods have revolutionised biology and have become the dominant way of testing historical evolutionary hypotheses (Huelsenbeck and Rannala 1997; Pagel 1999). Currently, the Bayesian phylogenetic approach is seen as the most powerful and robust approach available (Lewis 2001; Huelsenbeck et al. 2001, 2002). In the section below we will outline the major components of Bayesian phylogenetic analysis: dataset construction, maximum likelihood modeling, and the search for the most probable evolutionary trees.



### 2.1 Data

For successful phylogenetic analysis we need a large amount of well-sampled data with sufficient historical information to resolve the aspects of the phylogeny we are interested in. The comparative method commonly used in historical linguistics takes a sample of lexicon and proceeds to reconstruct systematic sound correspondences between the languages in order to uncover historically related 'cognate' forms (Durie and Ross 1996). This information about cognate sets can easily be coded as binary characters. An example of this is shown in Table 1. The data, in this case the words meaning 'bone' in a number of Austronesian languages (Column A) are divided into cognate sets on the basis of systematic sound correspondences (Column B). Once the cognate sets have been determined and any known loan words removed, then the data can be coded into a binary matrix showing the presence or absence of each cognate set for every language (Column C). In the 400-language dataset used in this paper, the cognate sets in a 210-item word-list produced 34,440 binary characters.

It is worth emphasising that whilst most recent work computing language phylogenies has primarily been based on cognate datasets (e.g. Atkinson et al. 2008; Gray and Atkinson 2003; Gray and Jordan 2000; Greenhill and Gray 2005; Gray et al. 2009; Holden 2002, Holden and Gray 2006; Pagel et al. 2007; Rexovà et al. 2003), other linguistic characters could also be used as long as there is sufficient data and an appropriate way of modeling the changes in these characters. Indeed, some studies have used combinations of lexical and grammatical data (Rexovà et al. 2006) and typological information (Dunn et al. 2005).

**Table 1:** Cognate data coding from original lexical data (A), to cognate set information (B), to binary characters (C)

| Language | (A) Item | (B) Cognacy | (C) Binary Coding |
|---|---|---|---|
| Paiwan | tsuqela | 1 | 1000 |
| Itbayaten | tuqgan | 1 | 1000 |
| Bare'e | wuku | 2 | 0100 |
| Mangarrai | toko | 2 | 0100 |
| Numfor | kor | 3 | 0010 |
| Motu | turia | 4 | 0001 |
| Fijian (Bau) | sui-na | 4 | 0001 |
| Tongan | hui | 4 | 0001 |
| Samoan | ivi | 4 | 0001 |
| Maori | iwi | 4 | 0001 |

### 2.2 Maximum likelihood models

The next step is to analyze the data. Bayesian phylogenetic inference builds on an older tradition of Maximum Likelihood methods (Fisher 1922; Edwards 1964; Felsenstein 1981; Pagel 1999). In this framework the data is treated as a fixed and given observation, and the analysis aims to find the values of model parameters that explain this data well (Pagel 1999; Steel and Penny 2000). To do this we need a stochastic model of language evolution that specifies how the changes between the character states should be counted. In modeling

language evolution in this way we make simplifying assumptions about relevant processes and explicitly build these into the model. For example, a very simple model of lexical evolution would require one parameter—the rate of change between the absence of a specific cognate and the presence of that cognate. In this simplest model, this rate would be symmetrical in the sense that the rate at which any cognate was gained would be equal to the rate at which a cognate was lost. Obviously, this is not very realistic. Once a cognate set has arisen it is much more likely to be lost than for another language to independently derive it. A more realistic model would accommodate the differential ease of losing a cognate over gaining it by adding a second parameter, so there is now one rate for cognate gain and one rate for cognate loss (we will refer to this as the two-parameter mode below).

What other important parameters could be added? One of the major problems with lexicostatistics is that it assumed a constant rate of cognate loss of around 19% every thousand years in the 200-item Swadesh list (Lees 1953). This fixed rate did not allow for differences in rates of change between cognate sets, or for differences in rates of change between languages. Both of these types of rate variation are common in Austronesian languages (Blust 1981, 2000). Site-specific rate heterogeneity (different sites in DNA sequences evolving at different rates) was also a problem for early phylogenetic methods (Posada and Crandall 2001). More recent approaches, however, have solved this by enabling a distribution of rates instead of a single rate. One common method is to estimate a gamma distribution of rate changes from the data (Yang 1994). This method gives each character an *inherent* rate of change so that some cognates are gained or lost rapidly, whilst others are more resistant to change. Modeling lexical change in this way allows for the differences between highly persistent characters like reflexes of 'hand—Proto Austronesian **(qa)-lima* (Blust 1999)—and highly unstable characters such as words meaning 'dirty'.

The full model with two rate parameters and gamma-distributed rate heterogeneity can then be used to calculate a numeric value known as the likelihood. The likelihood measures how well the data are explained by the tree under this model. Our aim is to find the set of trees that explain the data well, or in other words, find those trees with the maximum likelihood. The general approach to finding trees here is to take a tree and then permute it in some fashion (e.g. by changing the tree shape, or the amount of change along a branch, or model parameters, etc.) to give a second tree. The likelihood of both those trees under the given model of language evolution can then be compared to find the better tree. Here the search algorithm (usually a Markov Chain Monte Carlo approach—described below), preferentially selects the tree with the better likelihood, and iterates over this procedure many times, to find a set of good trees.

Critics of early language studies using Bayesian phylogenetic methods claimed that the models were 'inappropriate' as they had been designed for biological analyses rather than linguistic change (Eska and Ringe 2005; Naklekh et al. 2005). This criticism demonstrates a misunderstanding of the rationale behind model-based inference. Whilst it is true that language change is complex, and the model employed here and elsewhere (e.g. Gray and Atkinson 2003; Gray et al. 2009) is simple, this simplicity does not necessarily discredit or invalidate the methodology. Developing a model is a trade-off between over- and under-fitting model parameters (Burnham and Anderson 1998). Typically the fit will improve as parameters are added to the model, especially if the new parameters capture an important



aspect of the process. More complex models are not uncommon in biology; one of the most popular models used for genetic data is the General Time-Reversible model (Yang et al. 1994). This model has six parameters: one for each of the rates of change between each combination of the four bases found in DNA. This is often coupled with gamma distributed rate heterogeneity, and an allowance for invariant sites, giving a total of eight parameters.

However, as parameters are added the sampling error also increases and therefore it becomes difficult to reliably estimate the model parameters (Swofford et al. 1996). Therefore, the goal of modeling language evolution is not to build a complex model that captures every aspect of language change, but rather to construct the simplest model that provides reliable estimates of the parameters with finite amounts of data. Choosing the most appropriate model is not an issue for armchair speculation. We can evaluate the performance of the model by analysing the data with a range of models, and then selecting the best model with a standard model comparison test such as the Likelihood Ratio Test (Goldman 1993), or Bayes Factor Comparison (Suchard et al. 2001).

### 2.3 An example of a likelihood calculation

To clarify the way in which likelihood scores are calculated we have outlined a simple example in Figure 1 (adapted from Swofford et al. 1996, and Atkinson and Gray 2006). This figure shows the basic procedure on a set of data coded in a binary matrix as described above (1A). We will follow the process of likelihood calculation for one of these characters: character 'a'. Character 'a' represents a cognate set found in the Oceanic languages Motu, Fijian, Samoan and Maori, and absent from the other languages in our example dataset. To show how the likelihood can measure how well a topology describes the data we will compare two different trees (1B). The tree on the left represents the accepted linguistic history of these languages, whilst the tree on the right does not. First, character 'a' is mapped onto both the trees, and all the possible ancestral states of this character are enumerated. The likelihood of this distribution of character state change on the tree is then calculated using the chosen model of cognate evolution that specifies the probabilities of transitions between cognate presence and absence (1C). The likelihood of the distribution of character 'a' on the tree is the product of all possible ancestral state reconstructions for this character (1D). Finally, the overall likelihood of each tree can be calculated by repeating this process for all the characters in the data, giving rise to a single score for each tree. Note that, in contrast to lexicostatistics, actual character state changes are inferred on the tree. This means that the distinction between retentions and innovations is part of the analysis. The overall likelihood score, generally reported as the log of the likelihood (lnL), represents how well the data is explained by the tree given the model. Better trees are characterized by less negative log likelihoods (1E). In figure 1, the tree on the left has a log likelihood of -4178, whilst the second tree scores -4627. Thus, the former tree is a better explanation of the data.

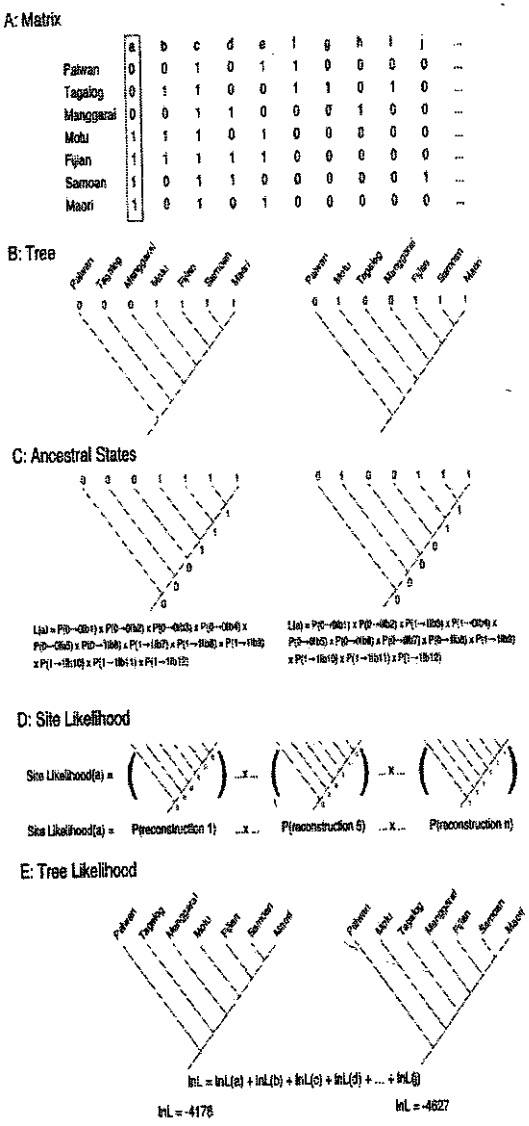



**Figure 1:** The calculation of the log likelihood of a tree. (A) A hypothetical cognate presence/absence matrix for seven Austronesian languages. (B) Two different trees for these languages with character 'a' mapped onto them. (C) An example of one possible ancestral state reconstruction of character 'a' on these topologies. (D) The site likelihood for character 'a' on the tree is calculated as the product of the probability of all possible ancestral state combinations for that character. (E) The overall tree likelihood is calculated as the sum of the likelihoods of each site, where the tree with the lower (less negative) log likelihood fits the data better. Here, under a two-parameter model of cognate gain/loss (with no gamma distribution), the tree on the left is a better fit to the data with a log likelihood of -4178, whilst the other tree fits the data less well with a likelihood of -4627.



## 2.4 Finding the most probable trees

Once we've chosen an appropriate model we then have an explicit *optimality criterion* with which to measure how good a tree is. This means that we can search through the range of possible trees until we find the one(s) with the highest likelihood under this optimality criterion. However, as the number of languages analysed increases so does the number of possible trees. If a tree is strictly bifurcating (i.e. each node can only have two daughter languages), then the number of trees can be calculated as shown in (1) where $n$ is the number of languages (Graham and Folds 1972).

$$(1) \qquad \frac{(2n-3)!}{2^{n-1}(n-1)!}$$

Thus, when there are four languages there are 15 possible trees. Adding one more language increases the number of possible trees to 105. When the data contains more than 50 languages, there are more possible trees than there are atoms in the universe. If Austronesian has around 1000 languages, then there are an intimidating $3.8 \times 10^{2864}$ possible combinations. This unfortunately means that it is not possible to search through all the trees in any non-trivial dataset. A systematic technique for finding a subset of the good trees from this huge space of possible trees is therefore required. Moreover, as with any statistical estimate, we need some way of evaluating how robust our inferences are.

To do this we use a Bayesian inferential approach that combines the likelihood with our prior knowledge of the trees to give the *posterior probability distribution* of trees. This can be calculated using Bayes's theorem as in (2) (Huelsenbeck et al. 2001).

$$(2) \qquad P[\text{Tree} \mid \text{Data}] = \frac{P[\text{Data} \mid \text{Tree}] \times P[\text{Tree}]}{P[\text{Data}]}$$

The posterior distribution contains the trees that have high likelihoods and fit the data well, given the data and the priors. Priors are the initial values of the model parameters. Often the prior distribution of the parameters is 'flat'; that is all values are considered equally probable. However, if there is strong external evidence supporting some hypothesis, then this can be taken into account explicitly (Lewis 2001). For example, if one wanted to assume that new languages were born at a constant rate across the tree, then a 'Yule' prior on branching rate could be implemented. The ability to incorporate extra information using priors is very powerful—but must be justified. Calculating the posterior probability distribution is hard as it involves the integration of all model parameters, across all branch length combinations, over every single tree (Huelsenbeck et al. 2001). However, using Markov Chain Monte Carlo methods (MCMC, Metropolis et al. 1953; Huelsenbeck et al. 2001), we can sample from the posterior probability distribution. The phrase 'Monte Carlo' refers to a random sampling method, and a 'Markov Chain' is a process which draws each sample from the probability distribution of the previous state (Larget 2005). To find trees this method starts with a tree (usually randomly generated) and permutes it in some fashion (e.g. changing the topology, branch lengths or model parameters)—this is the Markov

Chain process. The chain preferentially samples trees from this distribution according to their likelihood scores—the Monto Carlo process. If run long enough the chain provides a representative sample of the most probable trees. There are two further considerations in the use of Bayesian MCMC methods. First, the initial trees sampled are heavily contingent on the model's starting parameters (i.e. the priors). To avoid this early samples in an MCMC run are usually discarded as 'burn-in'. Second, each successive tree in an MCMC run is a permutation of the previous one due to the nature of the Markov Chain process (i.e. tree 2 is tree 1 with a branch moved or a change in branch length, etc). This means that each tree is highly correlated with its neighbors. To avoid this auto-correlation, and thus make each sample statistically independent, it is common to only keep every 1,000th or 10,000th tree from the post-burn-in set of trees.

## 3 Using phylogenetic trees

Using this procedure we will be left with a collection of trees sampled from the posterior probability distribution that should explain the data well. The results we present here are drawn from an analysis using the two-parameter model of cognate gain/loss and gamma-distributed rate variation (Pagel and Meade 2004). This was run for 100,000,000 generations on a cluster of over 150 processors (over 21 years of computer time). The trees were sampled every 10,000 generations after a burn-in of 20,000,000 generations. This gave us a final sample of 8000 trees. However, the endpoint of a phylogenetic analysis is not finding the trees: trees by themselves are boring. Instead, the rationale is to use them to test hypotheses and to investigate the process of evolution. There are many things one can do with trees (Gray et al. 2007). Here we will describe how this set of 8000 most probable trees from the MCMC run can be used to test hypotheses about subgrouping, to date events on the trees, and to trace character change.

### 3.1 Subgrouping

In historical linguistics it is common to use a family tree to depict the groupings (families, groups, clades, etc) once the groups have been identified using the comparative method. However, there is no formal way of quantifying the support for subgroups. The phylogenetic trees provide a statistical *estimate* of the sub-groupings in the data, and provide a measure of the uncertainty in this estimate. A common way of doing this is to use a Majority Rule 'consensus' tree. This combines the groupings present in all trees in the posterior tree sample. The percentage of trees containing a certain group can be taken as a measure of the support for that grouping in the data. Figure 2 shows an example majority rule consensus tree from our Austronesian data. Subgroups with posterior probability values close to 1.0 are well-supported. For example, the grouping of the Philippine languages is strongly supported by the data (0.99). More surprisingly, the branch grouping the languages of Vanuatu and New Caledonia is also well-supported (0.98). These values mean that 99% and 98% of the 8000 trees in the posterior tree distribution contain those respective groupings. In contrast, other regions of the tree are more poorly supported (e.g. the branch placing the Admiralties languages inside Oceanic after the New North Guinea/Papuan Tip languages has only 0.58 support). Groups with very weak support (<0.50) are not shown. Weakly supported groups could either be the consequence of little signal in the data due to rapid population expansions, or conflicting stronger signals (perhaps produced by borrowing), or non-tree-like descent processes such as dialect chains and linkages.



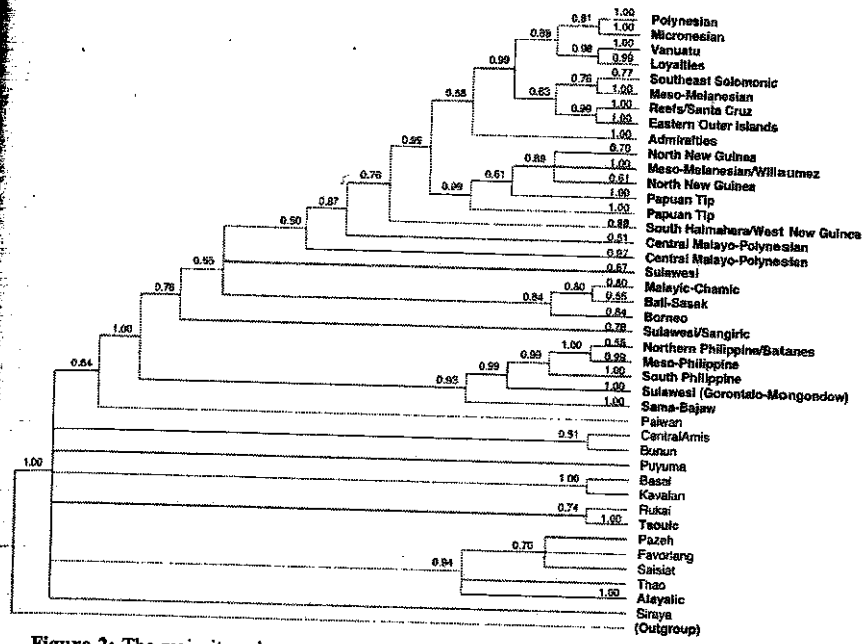



**Figure 2:** The majority-rule consensus tree of all post burn-in Austronesian trees. Labels in bold represent subgroups of languages, normally-weighted labels denote languages. Where subgroups appear twice in the tree this indicates that they are not monophyletic (e.g. Central Malayo-Polynesian). The numbers on the branches denote the posterior probability of each node. For example, the split between the Northern- and Meso-Philippine languages is strongly supported (1.00). Posterior probability values below 0.50 are considered weak and are not included.

Recall that the lexicostatistical analyses incorrectly 'rooted' the Austronesian languages in East New Guinea. Our phylogenetic analyses, however, support Blust's (1999) rooting of the Austronesian languages in Taiwan. The Formosan languages are placed at the base of the tree after the outgroup languages. There is no unified Formosan subgroup but at least seven higher-order branches of Formosan derived from Proto Austronesian. Moreover, whilst the Tsouic and Atayalic subgroups of Formosan languages are robust, there is little support for other higher-order sub-groupings within Formosan. These results are all concordant with Blust (1999).

Not only do the phylogenetic trees support a Formosan origin of the Austronesian languages, the sequence of the higher-order subgroups closely conforms to the 'Out of Taiwan' scenario of Austronesian settlement. Moving down the tree, after Formosan languages we find the languages of Island South-East Asia, with strong support for the Philippine and Malayic-Chamic language groups. This is followed by two weakly supported groups of Central Malayo-Polynesian languages, and then the well-supported South Halmahera/West New Guinea group. Finally, there is a well-supported Oceanic subgroup, with strong support for the recognized subgroups within Oceanic (Polynesian, Micronesian, Southeast Solomonic, Eastern Outer Islands, Admiralties). Our results split

Oceanic into two major groups, both strongly supported (0.99). The first of these Oceanic subgroups is comprised of the Papuan Tip, North New Guinea and Meso-Melanesian languages. This represents the Western Oceanic group identified by Ross (1988). However, only the Willaumez languages of Meso-Melanesian are in this subgroup, the remainder is located in our second Oceanic grouping. This second Oceanic group contains the Remote Oceanic language subgroups and the majority of the Meso-Melanesian languages. Interestingly, we show strong support (0.99) for the recently identified subgroup Temotu containing the languages from the Eastern Outer Islands and the Reefs–Santa Cruz region (Ross and Næss 2007). In contrast to Blust (1998), the Admiralties subgroup is not at the base of the entire Oceanic subgroup, but is situated—albeit very weakly (0.58)—between Western and Remote Oceanic. Some of the higher-order nodes within our two Oceanic groupings are only weakly supported, such as the cluster grouping Temotu to Southeast Solomonic (0.63). These low values may reflect the rapid dispersal of languages through this region (Pawley 1999), or the large amounts of contact induced change in large-scale dialect networks found in this region (Ross 1996).

### 3.2 Dating

One of the great attractions of lexicostatistics was its apparent ability to calculate absolute dates of language divergence times through a method known as *glottochronology* (Lees 1953). This technique calculated absolute ages by assuming that as languages split they lost vocabulary at a constant rate. Accordingly, a simple decay curve of cognate loss could be used to calculate divergence times by solving the equation in (3) where $C$ is the percentage of shared cognates between the two languages, $r$ is the retention rate, and $t$ is the estimated time depth.

$$t = \frac{\log C}{2 \log r} \tag{3}$$

Over 1000 years the retention rate $r$ was often assumed to be 81% for the 200 item Swadesh list (Lees 1953). Therefore, if two languages shared 90% of their basic vocabulary, they should have diverged 250 years ago, whilst languages that were 75% similar should have diverged around 680 years ago. However, these glottochronological calculations magnified all the shortcomings of lexicostatistics. Languages vary substantially in their retention rates, and this rate variation produced some obviously inaccurate dates (Bergsland and Vogt 1962; Blust 2000). For example, Icelandic shares over 95% of its core vocabulary with Old Norse. According to glottochronology Old Norse and Icelandic would have diverged less than 200 years ago. This is incorrect—Old Norse was spoken around 1000 years ago (Bergsland and Vogt 1962). Problems such as this led to such a strong rejection of glottochronology that over fifty years later we are still being cautioned about its inaccuracy (McMahon and McMahon 2006).

The age of the Indo-European language family has been a topic of considerable interest and much debate. There are two main theories. The first proposes that Proto Indo-European broke up 5000–6000 years ago when Indo-European languages spread with the expansion of the archaeological culture known as Kurgan (Gimbutas 1973). The main alternative account suggests that Indo-European spread with the advent of farming technology around 8000–9000 years ago (Renfrew 1987). Naturally, one of the first uses we put phylogenetic methods to was dating the divergence of particular branches of Indo-European (Gray and Atkinson 2003, Atkinson and Gray 2006). Our results showed strong support for an initial breakup of the Indo-European family around 8000–9000 years ago, with a subsequent breakup of 'Nuclear Indo-European' (Indo-European minus Anatolian and Tocharian) around 6000 years ago. The results were robust to different calibrations, cognate coding, and likelihood models (Atkinson et al. 2005). However, we were promptly criticized for merely, 'reintroducing glottochronology by the back door' (Gamble et al. 2005:208), and 'ignor(ing) the fatal shortcomings of glottochronology' (Eska and Ringe 2004:569). These are unfortunate misunderstandings. Phylogenetic dating methods, such as the Penalized Likelihood rate smoothing approach (Sanderson 1997, 2002) used by Gray and Atkinson (2003), as well as newer methods which can 'relax the clock' (Drummond et al. 2006), do not have the fatal shortcomings of glottochronology. These approaches need not assume that there is a single 'clock-like' rate of lexical change (Atkinson and Gray 2006).

To demonstrate how divergence date estimation can be obtained without a strict 'glottoclock' we will estimate the age of Proto Austronesian on the (expanded) tree from Figure 2. The branches on the trees in our posterior sample are proportional to the amount of change along that lineage. This is usually expressed as the rate of substitutions (in this case the gain or loss of cognates in a language). These branch lengths can be converted to time by adding historically attested calibration points. For example, the Eastern Polynesian subgroup can be constrained to around 1200 to 1300 years ago on the basis of initial settlement times (Green and Weisler 2002). Similarly, the Chamic subgroup can also be calibrated based on the fact that Chamic speakers were mentioned in Chinese records around 1800 years ago, and probably entered Vietnam around 2600 years ago (Thurgood 1999). This calibration of nodes on the tree within a historical time range allows the method to estimate how fast the changes measured by the branch lengths are occurring. The Penalized Likelihood rate-smoothing approach can then convert branch lengths into time estimates by smoothing the rates of change across the tree. Instead of assuming a constant retention rate, this allows certain parts of the tree to change faster or slower than others. We applied this approach to one tree from the posterior distribution of trees for our analysis. The resulting dated tree (Figure 3) shows an age of around 5310 years for Proto Austronesian, and an age of 4240 years for Proto Malayo-Polynesian. We must emphasise at this point that the date estimates should be done on all trees in the posterior sample and not just a single one. Calculating divergence dates on all the trees would produce a distribution of the most probable age of Proto Austronesian. This distribution can then be used to provide a confidence interval on any date estimate. As our aim in this paper is to illustrate the overall approach rather than to test specific hypotheses, we have just dated one tree for illustrative purposes. However, dates from this tree support the emergence of Proto Austronesian in Taiwan around 5500 years ago (e.g. Blust 1995; Pawley 2002). Note also the presence of pauses and rapid pulses of expansion as has been argued by Blust (1999), Green (1999) and Pawley (1999, 2002). In this tree, we see a pause of around 1000 years before Proto Malayo-Polynesian arises, and a subsequent rapid pulse of expansion through to Proto Oceanic. Another pause then expansion pulse occurs after the initial settlement of the Central Pacific region.

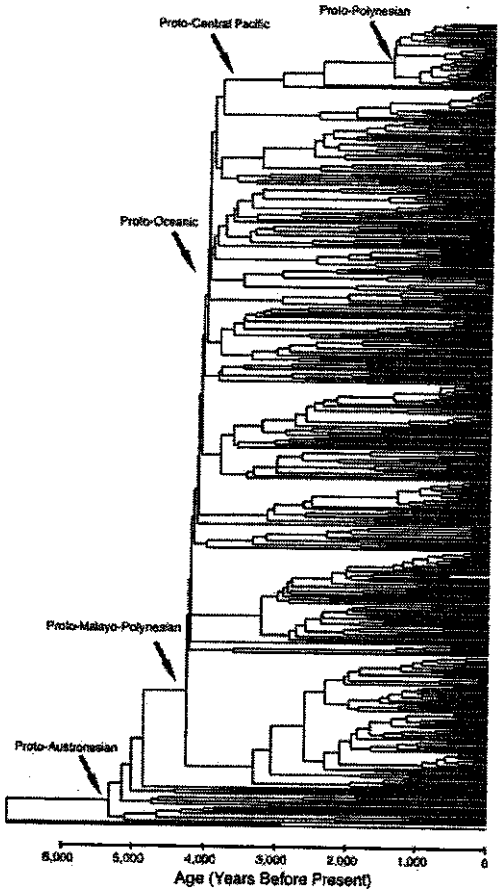



**Figure 3:** A dated tree of 400 Austronesian languages showing the age of a number of protolanguages estimated using Penalized Likelihood rate smoothing. On this tree Proto Austronesian is estimated to be 5310 years old and Proto Malayo-Polynesian 4240 years.

### 3.3 Tracing character history

Much of historical linguistics is concerned with the reconstruction of protoforms. This is done both as a means to subgrouping and as a way of making inferences about society and culture of ancestral speech communities. Biologists have also developed phylogenetic methods to reconstruct ancestral states. These methods have been used to tackle problems such as identifying the origin of ancestral genes in the eukaryote genome (Lester et al. 2006). One common phylogenetic approach essentially 'maps' a character of interest onto the posterior tree sample using a continuous-time Markov model of trait evolution (Pagel et al. 2004). Under this model a character can change between a finite number of states over infinitesimally small time periods. The rates of change between these states along the



branches can be estimated directly from the posterior tree sample. These model parameters can then be used to calculate the probability of a certain state at any given node. For example, one might want to evaluate how the words for 'earth, soil' had evolved in the Polynesian languages, and infer what variant was spoken by Proto Polynesian. Figure 4 shows three cognate sets for words meaning 'earth/soil' mapped onto a tree of the Central Pacific subfamily (the expanded form of Figure 2). Cognate set A (colored white) reflects forms like Tongan *kelekele*, Samoan *'ele'ele* and Fijian (Bau) *qele*. Cognate set B (colored gray) reflects forms like the Tahitian *repo* and Hawaiian *lepo*. Cognate set C (colored black) reflects forms like Vaeakau-Taumako's *pela*. Using the Bayesian ancestral state reconstruction method (Pagel et al. 2004) we can estimate that, on this tree, the probability that Proto East Polynesian and Proto Tahitic had cognate set B was 0.99. This is concordant with the comparative method, where the reconstructed Proto East Polynesian form is **repo* (Biggs and Clark 2000). Deeper in the tree, the Proto Polynesian and Proto Central Pacific nodes reflect cognate set A with a probability very close to 1. Again, this matches the reconstructed Proto Central Pacific form **g(w)ele* (Ross et al. 1998). Cognate set C presumably reflects Prot Oceanic **pela* 'muddy' (Biggs and Clark 2000) with semantic change. We emphasise again that ideally this estimation should be integrated over the set of trees in the posterior sample, not just a single tree.

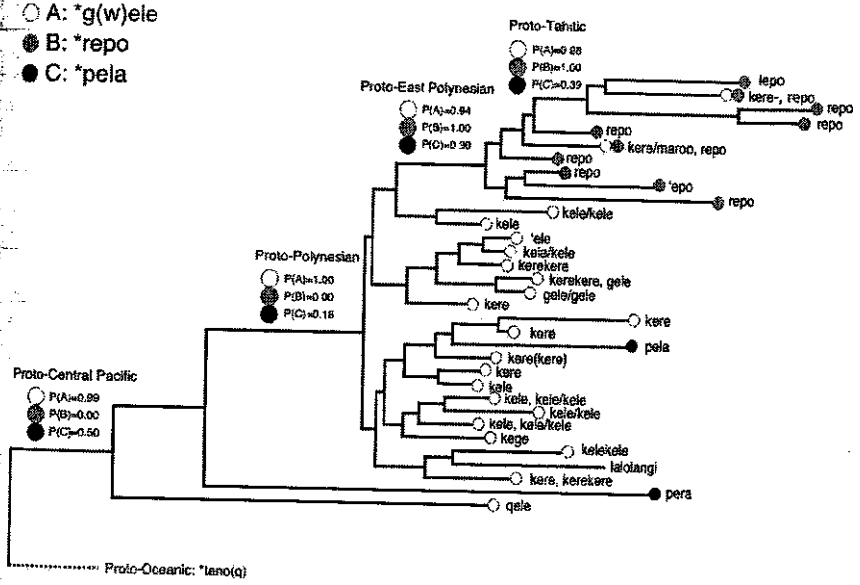



**Figure 4:** Tree of Central Pacific languages showing the distribution of three cognate sets A (in white), B (in gray), and C (in black) containing words for 'earth/soil'. Branch lengths are proportional to amount of change along the lineage. The probability of the ancestral states are marked for a number of protolanguages. The probability of Proto Tahitic and Proto East Polynesian belonging to cognate set B (*repo) is >0.99. Proto Polynesian and Proto Central Pacific instead contain cognate set A (*g(w)ele) also with a probability of >0.99.

## 4 Conclusion

We hope that this chapter has corrected most of the persistent myths and misconceptions about the application of computational phylogenetic methods to historical linguistics. Let us be very clear. Phylogenetic methods do not make the flawed assumptions of lexicostatistics or glottochronology. They do not count cognates to calculate pairwise similarity measures. Instead, the likelihood calculations are based on each cognate set and how it fits onto the tree. Phylogenetic methods do not require a single 'one size fits all' rate of lexical replacement. These methods can allow for different rates of change both between cognate sets and between different lineages. Moreover, this framework can explicitly take into account external evidence such as archaeological dates and known historical events to make robust inferences about divergence dates. In marked contrast to lexicostatistics, the phylogenetic methods we have detailed here perform exceptionally well on the very difficult problem of the Austronesian subgrouping and dating. First, the trees are rooted in Taiwan, in line with the results of the comparative method. Second, the sequence and subgrouping of these phylogenetic trees strongly reflect the structure of the family tree suggested by the comparative method, at least in those cases where there is a consensus among comparative linguists. Third, the timing of events on these trees again corresponds extremely well to the 'Out-of-Taiwan' scenario.

We also hope to have laid to rest a final vexing misconception about phylogenetic linguistics: 'this method is not giving anything new' (Jasanoff in Wade 2004:1). Not only do phylogenetic methods work well and outperform lexicostatistics, they also provide a range of new tools that can be of great benefit to linguistics. First, phylogenetic methods provide an explicit optimality criterion for evaluating how well different trees (i.e. historical scenarios) are supported by the data. Second, they provide an empirical way of assessing the statistical robustness of any subgroup in those trees. We have shown here a number of Austronesian examples where the support values on our trees coincide well with linguistic intuitions about the strength of support for these groupings. Third, despite the failure of glottochronology to provide robust date estimates, the attraction of absolute dating is strong. Dates are critically important for inferences about human prehistory. They provide a powerful way of linking linguistic, archaeological, cultural, and genetic evidence. It is not uncommon to still see glottochronological age estimates cited in publications, along with the standard disclaimer that this method cannot be trusted (e.g. Campbell 1997; Comrie 2002; Pawley 2002). Phylogenetic dating methods can, when used carefully and appropriately, help integrate our inferences about human prehistory without these glaring disclaimers. Fourth, these methods enable us to investigate how linguistic and cultural traits have evolved in families by tracing their history. These tools can infer ancestral states and can even be used to infer functional dependency between linguistic characters (Gray et al. 2007). Far from being lexicostatistics-redux, Bayesian phylogenetic methods provide exciting new tools for historical linguistics.

## References

Atkinson, Q.D., G. Nicholls, D. Welch and R.D. Gray. 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103:193–219.



Atkinson, Q.D. and R.D. Gray. 2006. Are accurate dates an intractable problem for historical linguistics? In C.P. Lipo, M.J. O'Brien, M. Collard and S.J. Shennan, eds *Mapping our ancestors: phylogenetic approaches in anthropology and prehistory*, 269–298. New Brunswick: Aldine.

Atkinson, Q.D., A. Meade, C. Venditti, S.J. Greenhill and M. Pagel. 2008. Languages evolve in punctuational bursts. *Science* 319:588.

Balter, M. 2003. Early date for the birth of Indo-European languages. *Science* 302:1490–1491.

Balter, M. 2004. Search for the Indo-Europeans. *Science* 303:1323–1326.

Bellwood, P. 1997. *Prehistory of the Indo-Malaysian Archipelago*. Honolulu: University of Hawai'i Press.

Bellwood, P., J.F. Fox and D. Tryon. 1995. *The Austronesians: historical and comparative perspectives*. Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Bergsland, K. and H. Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3:115–153.

Blust, R.A. 1977. The proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai'i Working Papers in Linguistics* 9:1–15.

—— 1981. *Variation in retention rate among Austronesian languages*. Talk given to the Third International Conference on Austronesian Linguistics. Bali.

—— 1982. The linguistic value of the Wallace line. *Bijdragen tot de taal-, land- en. volkenkunde* 138:231–250.

—— 1995. The prehistory of the Austronesian-speaking peoples: the view from language. *Journal of World Prehistory* 9:453–510.

—— 1998. A note on higher-order subgroups in Oceanic. *Oceanic Linguistics* 37:182–188.

—— 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In E. Zeitoun and P. Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics* vol. 1, 31–94. Taipei, Taiwan: Symposium Series of the Institute of Linguistics, Academia Sinica.

—— 2000. Why lexicostatistics doesn't work: the 'universal' constant hypothesis and the Austronesian languages. In C. Renfrew, A. McMahon and L. Trask, eds *Time depth in historical linguistics*, 311–331. Cambridge: The McDonald Institute for Archaeological Research.

Bowern, C. 2006. Punctuated equilibrium and language change. In K. Brown, ed. *Encyclopedia of language and linguistics*, 286–289. Oxford: Elsevier.

Brandes, J.L.A. 1884. *Bijdrage tot de vergelijkende klankleer der westersche afdeeling van de Maleisch-Polynesische taalfamilie*. Utrecht.

Brugmann, K. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der Indogermanischen Sprachen. *Internationale Zeitschrift für allgemeine Sprachewissenschaft* 1:226–256.

Bryant, D. 2006. Radiation and Network Breaking in Polynesian Language Evolution. In P. Forster and C. Renfrew, eds Phylogenetic Methods and the Prehistory of Languages, 111–118. Cambridge: McDonald Institute Press, University of Cambridge.

Bryant, D., F. Filimon and R.D. Gray. 2005. Untangling our past: languages, trees, splits and networks. In R. Mace, C.J. Holden and S. Shennan, eds *The evolution of cultural diversity: phylogenetic approaches*, 67–84. London: UCL Press.

Burnham, K.P. and D.R. Anderson. 1998. *Model selection and inference — a practical information-theoretic approach*. New York: Springer.

Campbell, L. 1997. *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.

—— 2004. *Historical linguistics: an introduction*. 2nd edition. Edinburgh: Edinburgh University Press.

Comrie, B. 2002. Farming dispersal in Europe and the spread of the Indo-European language family. In P. Bellwood and C. Renfrew, eds *Examining the farming/ language dispersal hypothesis*, 409–419. Cambridge: The McDonald Institute for Archaeological Research.

Dahl, O.C. 1973. *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series No.15. Studentlitteratur: Lund, Sweden.

Dempwolff, O. 1934. *Vergleichende Lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*. 1 Induktiver Aufbau einer indonesischen Ursprache, 15, Berlin: Deitrich Reimer.

—— 1937. *Vergleichende lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*. 2, Deductive Anwendung des urindonesischen auf austronesische Einzelsprachen, 17, Berlin: Deitrich Reimer.

—— 1938. *Vergleichende lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*. 3 Austronesisches Wörterverzeichnis, 19, Berlin: Deitrich Reimer.

Diamond, J. and P. Bellwood. 2003. Farmers and their languages: the first expansions. *Science* 300:597–603.

Dixon, R.M.W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dunn, M., A. Terrill, G. Reesink, R.A. Foley and S.C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309:2072–2075.

Durie, M. and M. Ross. 1996. *The comparative method reviewed: regularity and irregularity in language change*. Oxford: Oxford University Press.

Drummond, A.J., S.Y.W. Ho, M.J. Phillips and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4:e88.

Dyen, I. 1962. The lexicostatistical classification of the Malayopolynesian languages. *Language*, 38:38–46.

—— 1965. A lexicostatistical classification of the Austronesian languages. In *Indiana University Publications in Anthropology and Linguistics: Memoir 19 of the International Journal of American linguistics*. Indiana: Indiana University.

Edwards, A.W.F. and L.L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. In J. McNeill, ed. *Phenetic and phylogenetic classification*, 67–76. Cambridge: Cambridge University Press.



Elbert, S. H. 1953. Internal relationships of Polynesian languages and dialects. *Southwest Journal of Anthropology* 9:147–173.

Eska, J.F. and D. Ringe. 2004. Recent work in computational linguistic phylogeny. *Language* 80:569–582.

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

—— 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368–376.

Fisher, R.A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London* Series A, 222:309–368.

Forster, J.R. 1778. *Observations made during a voyage round the world*. London.

Gamble, C., W. Davies, P. Pettitt, L. Hazelwood and M. Richards. 2005. The archaeological and genetic foundations of the European population during the Late Glacial: implications for 'agricultural thinking'. *Cambridge Archaeological Journal* 15:193–223.

Gimbutas, M. 1973. Old Europe c. 7000–3500 B.C., the earliest European cultures before the infiltration of the Indo-European peoples. *Journal of Indo-European Studies* 1:1–20.

Goldman, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36:182–198.

Gordon, Raymond G., Jr. 2005. *Ethnologue: languages of the World*, Fifteenth edition. Dallas, Texas: SIL International. Online version: http://www. ethnologue. com/.

Grace, G.W. 1961. Austronesian linguistics and culture history. *American Anthropologist* 57:359–368.

—— 1964a. Movement of the Malayo-Polynesians 1500 BC to AD 500: the linguistic evidence. *Current Anthropology* 5:361–368.

—— 1964b. The linguistic evidence. *Current Anthropology*, 5:361–368.

—— 1966. Austronesian lexicostatistical classification: a review article. *Oceanic Linguistics* 5:13–31.

Graham, R.L. and L.R. Foulds. 1982. Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. *Mathematical Biosciences* 60:133–142.

Gray, R.D. 2007. *Tangled trees: what do phylogenetic networks reveal about Oceania linguistic history?* Talk given to the Seventh International Conference on Oceanic Linguistics. Noumea, New Caledonia.

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

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

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

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

Green, R.C. 1999. Integrating historical linguistics with archaeology: insights from research in remote Oceania. *Indo-Pacific Prehistory Association Bulletin* 18 (Melaka Papers), vol. 2:3–16.

Green, R.C. and M.I. Weisler. 2002. The Mangarevan sequence and dating of the geographic expansion into Southeast Polynesia. *Asian Perspectives* 41:213–241.

Greenberg, J.H. 1949–54. Studies in African linguistic classification. *South-western Journal of Anthropology* 5:79–100, 190–198, 309–317; 6:47–63, 143–160, 223–237, 388–398; 10:405–415.

Greenhill, S.J. and R.D. Gray. 2005. Testing population dispersal hypotheses: Pacific settlement, phylogenetic trees and Austronesian languages. In R. Mace, C.J. Holden and S. Shennan, eds *The evolution of cultural diversity: phylogenetic approaches*, 31–52. London: UCL Press.

Greenhill, S.J., R. Blust and R.D. Gray. 2008. *The Austronesian basic vocabulary database*: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics*, 4:271–283. http://language.psy.auckland.ac.nz

Hamed, M.B. and F. Wang. 2006. Stuck in the forest: Trees, networks and Chinese dialects. *Diachronica* 23:29–60.

Hoijer, H. 1956. Lexicostatistics: a critique. *Language* 32:49–60.

Holden, C.J. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony approach. *Proceedings of the Royal Society of London, B Biological Sciences* 269:793–799.

Holden, C.J. and R.D. Gray. 2006. Rapid radiation, borrowing and dialect continua in the Bantu languages. In P. Forster and C. Renfrew, eds *Phylogenetic methods and the prehistory of languages*, 19–31. Cambridge: McDonald Institute for Archaeological Research.

Holm, H.J. 2007. The new arboretum of Indo-European 'trees'. Can new algorithms reveal the phylogeny and even prehistory of Indo-European? *Journal of Quantitative Linguistics* 14:167–214.

Houtman, F. de. 1603. *Spraeck ende Woord-Boeck, Inde Maleysche ende Madagaskarsche Taln met vele Arabische ende Turcsche Woorden*. Amsterdam.

Huelsenbeck, J.P. and B. Rannala. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276:227–232.

Huelsenbeck, J.P., F. Ronquist, R. Nielsen and J.P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.

Huelsenbeck, J.P., B. Larget, R.E. Miller and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51:673–688.

Hymes, D.H. 1960. Lexicostatistics so far. *Current Anthropology* 1:3–44.

Kern, H. 1886. *De Fidji-taal vergeleken met hare verwanten in Indonesië en Polynesië*. Verhandelingen der Koninklijk Akademie van Wetenschappen, 16, Amsterdam.

Kirch, P.V. 2000. *On the road of the winds: an archaeological history of the Pacific Islands before European contact*. Berkeley: University of California Press.



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

Labov, W. 1994. *Principles of linguistic change: internal factors*. Oxford: Blackwell.

Larget, B. 2005. Introduction to Markov Chain Monte Carlo methods in molecular evolution. In R. Nielsen, ed. *Statistical Methods in Molecular Evolution*, 45–62. New York: Springer.

Lees, R.B. 1953. The basis of glottochronology. *Language* 29:113–127.

Lester, L., A. Meade and M. Pagel. 2006. The slow road to the eukaryotic genome. *BioEssays* 28:57–64.

Lewis, P.O. 2001. Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution* 16:30–37.

Lynch, J., M.D. Ross and T. Crowley. 2002. *The Oceanic languages*. Richmond: Curzon Press.

McMahon, A. and R. McMahon. 2006. Why linguists don't do dates: evidence from Indo-European and Australian languages. In P. Forster and C. Renfrew, eds *Phylogenetic methods and the prehistory of languages*, 153–160. Cambridge: McDonald Institute for Archaeological Research.

Murdock, G.P. 1964. Genetic classification of the Austronesian languages: a key to Oceanic culture history. *Ethnology* 3:117–126.

Nakhleh, L., T. Warnow, D. Ringe and S.N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Transactions of the Philological Society* 103:171–192.

Nettle, D. 1999. Is the rate of linguistic change constant? *Lingua* 108:119–136.

Nichols, J. 1997. Modeling ancient population structures and movement in linguistics. *Annual Review of Anthropology* 26:359–384.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.

Pagel, M. and A. Meade. 2004 A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology* 53:571–581.

Pagel, M., Q.D. Atkinson and A. Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449:717–720.

Pawley, A. 1999. Chasing rainbows: implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In E. Zeitoun and P. Jen-kuei Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics* vol. 1, 95–138. Taipei, Taiwan: Symposium Series of the Institute of Linguistics, Academia Sinica.

—— 2002. The Austronesian dispersal: languages, technologies and people. In P. Bellwood and C. Renfrew, eds *Examining the farming/language dispersal hypothesis*, 251–274. Cambridge: McDonald Institute for Archaeological Research.

Pawley, A. and R.C. Green. 1973. Dating the dispersal of the Oceanic languages. *Oceanic Linguistics* 121–2:1–67.

Pawley, A. and M. Ross. 1995. The prehistory of the Oceanic languages: a current view. In P. Bellwood, J.J. Fox and D.T. Tryon, eds *The Austronesians: historical and comparative perspectives*, 39–74. Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Penny, D., B.J. McComish, M.A. Charleston and M.D. Hendy. 2001. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *Journal of Molecular Evolution* 53:711–723.

Posada, D. and K.A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50:580–601.

Renfrew, C. 1987. *Archaeology and language: the puzzle of Indo-European origins*. London: Cape.

Reland, H. 1708. *Dissertatio de linguis insularum quarundem orientalium*. Trajecti ad Rhenum.

Rexovà, K., Y. Bastin and D. Frynta. 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften* 93:189–194.

Rexovà, K., D. Frynta and J. Zrzavy. 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19:120–127.

Ringe, D., T. Warnow and A. Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100:59–129.

Ross, M. 1988. *Proto Oceanic and the Austronesian languages of Western Melanesia*. Canberra: Pacific Linguistics.

—— 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In M. Durie and M.D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 180–217. New York: Oxford University Press.

—— 1997. Social networks and kinds of speech-community events. In R. Blench and M. Spriggs, eds *Archaeology and language*, vol. 1, 209–261 London: Routledge.

Ross, M. and A. Næss. 2007. An Oceanic origin for Äiwoo, the language of the reef islands? *Oceanic Linguistics* 46:456–498.

Sanderson, M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14:1218–1231.

—— 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19:101–109.

Shutler, R. and J.C. Marck. 1975. On the dispersal of the Austronesian horticulturalists. *Archaeology and Physical Anthropology in Oceania* 10:81–113.

Sokal, R.R. and P.H.A. Sneath. 1963. *Principles of numerical taxonomy*. San Francisco: W.H. Freeman.

Steel, M. and D. Penny. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molecular Biology and Evolution* 17:839–850.

Suchard, M.A., R.E. Weiss and J.S. Sinsheimer. 2001. Bayesian selection of continuous-time Markov Chain evolutionary models. *Molecular Biology and Evolution* 18:1001–1013.



Swadesh, M. 1952. Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96:453–463.

Swofford, D.L., G.J. Olsen, P.J. Waddell and D.M. Hillis. 1996. Phylogenetic inference. In D.M. Hillis, C. Moritz and B.K. Mable, eds *Molecular Systematics*, 407–514. Sinauer Associates, Sunderland, MA.

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

Wade, N. 2004, 16 March. A biological dig for the roots of language. *The New York Times*.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* 39:306–314.

Yang, Z., N. Goldman and A.E. Friday. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Molecular Biology and Evolution* 11:316–324.

# 23 *Linguistic differences among Siraya, Taivuan and Makatau*

PAUL JEN-KUEI LI

## 1 Introduction

Siraya, Taivuan and Makatau were the Formosan languages or dialects formerly spoken in the southwestern plains of Taiwan.[1] Roughly speaking, Siraya was spoken in the coastal area of Tainan Plain and Taivuan mostly in the inland of Tainan Plain to the north, while Makatao was spoken in Kaohsiung and Pingtung prefectures to the south. The languages or dialects probably became extinct in the first half of the 19th century (Li 2002). Dutch missionaries left behind three main written documents, namely *The Gospel of St. Matthew in Formosan Sinkang Dialect* (Gravius 1661, henceforth *St. Matthew*)), *'t Formulier des Christendoms* (Gravius 1662, henceforth *Formulary*), and the *Utrecht Manuscript* (unknown author, published in Van der Vlis 1842).[2] Ever since then only short wordlists have been recorded in various villages in the southwest plains at different times between 1717 and 1917 by the Chinese, Europeans, and Japanese. Ogawa (1917) assembled those wordlists and classified them into three main groups: Siraya, Makatau and Taivuan. There are altogether 75 villages or sources of language data and 163 lexical entries represented in his comparative wordlist (see Tsuchida et al. 1991). Due to the paucity of language data in that area, his comparative wordlist is extremely valuable; especially for Makatau. Tsuchida (Tsuchida et al. 1991:ix) prepared a map, which shows the location of 39 villages. It uses three different signs to indicate the three different groups of languages or dialects, which gives us an idea about the geographical distribution of the erstwhile linguistic communities in the southwestern plains.

1 An earlier version of this paper in Chinese (Li 2006) appeared in a conference proceedings. In this version I have up-dated the language data, revised the internal relationships of the three groups and added some new findings. I also discuss the affiliation of Dutch missionary documents. In preparing this paper, I benefited from Ogawa's pioneering work on Siraya as well as from Tsuchida's and Adelaar's valuable suggestions and the help of my assistants, Hsiu-min Huang and Amy Minnuan Chen. This work was supported in part with a grant from the National Science Council (NSC95-2411-H-001-010-H).

2 Adelaar (1997:364ff.) also discusses dialect variations between the *Utrecht Manuscript* on the one hand, and *St. Matthew* and the *Formulary* on the other.

Alexander Adelaar and Andrew Pawley, eds
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 399–409.
Pacific Linguistics, 2009.

Based on Ogawa's comparative wordlist, Tsuchida pointed out that the three groups have different reflexes of PAn *l and *N, as shown below.

**Table 1:** Sirayaic Reflexes of PAn *l and *N

| | PAn | Siraya | Taivuan | Makatau | |
|---|---|---|---|---|---|
| (1) | *l | r | Ø~h | r | |
| e.g. | *telu | turu | too, toho | toru | 'three' |
| | *lima | rima | hima | rima | 'five' |
| (2) | *N | l | l | n | |
| e.g. | *(qa)Nuang | luang | lowan | noang | 'cow' |
| | *puNi | mapuli | mapuli | mapuni | 'white' |

But as Tsuchida was aware, there are many exceptions to the rules, perhaps due to the poor or inaccurate transcriptions of the language data from various sources. Assuming that the phonological correspondences given above are correct, each group differs from the other two by only one phonological innovation. These might be regarded as dialectal differences, as commonly found in other Formosan languages, such as Rukai (Li 1977) and Atayal (Li 1981).

## 2 Evidence from Sinkang manuscripts

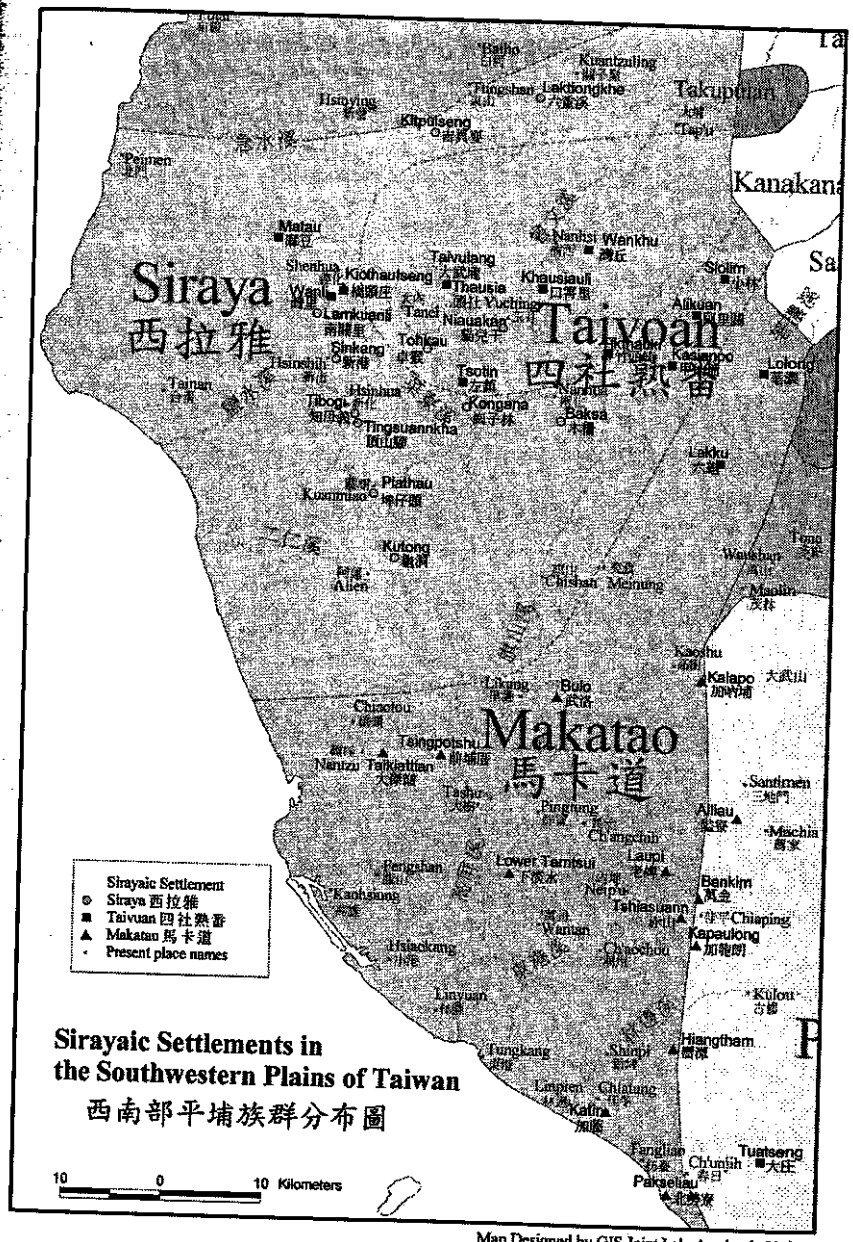
Aside from the Dutch missionary documents and the short wordlists for the language data of the southwestern plains, there is a third type of language data available: the so-called 'Sinkang manuscripts' are contracts written in Romanised script. These manuscripts or texts were found in various villages in the southwestern plains that belong to the three different groups. For example, Sinkang, Tohkau and Kongana belong to Siraya; Wanli, Matau (see below) and Taibulang belong to Taivuan; Lower Tamtsui and Katin belong to Makatau. The earliest text is dated 1663 and the latest 1818. Murakami (1933) collected 101 manuscripts.[3] My colleagues and I have accumulated 170 manuscripts.[4] The great majority came from Siraya villages, only 23 came from Taivuan villages and four from Makatau villages. Among these, 67 are written in both Chinese and a native language, while the remaining ones are monolingual. A careful study of these texts may reveal significant linguistic differences, not found in wordlists.

My assistants Hsiu-min Huang, Chin-wen Chien, and I have worked on Sinkang manuscripts in the past eight years (since 2001). Although they are extremely hard to decipher, we have tried to decipher and transcribe all of them, determine word and sentence boundaries, identify each lexical item, and give interlinear glosses and free translation for each sentence whenever possible. All 170 texts exist in the form of computer files. These texts do reveal some interesting facts about the language or dialects in the southwestern plains.

[3] Of the 101 manuscripts collected by Murakami, 87 are from the village of Sinkang, six from Matau, three from Tohkau, one from Taibulang, one from Lower Tamtsui, and three from Katin. The 87 Sinkang manuscripts are treated as the main body of his monograph, while the other 14 are given in appendices.

[4] Of the 170 manuscripts, one came from Backloan (Siraya, not found in Murakami), seven from Tohkau, ten from Matau, one from Taibulang (Taivuan), eleven from Wanli (Taivuan, not found in Murakami), one from Lower Tamtsui, three from Katin, and two from Gutiaupo 牛稠埔 (Siraya or Makatau, not found in Murakami).





### 2.1 Phonological evidence

In addition to the two phonological innovations observed by Tsuchida (1991), I have found two additional ones based on the language data in the Sinkang manuscripts, as illustrated below.

**Table 2:** Sirayaic Reflexes of PAn *D, *-k- and *-S-/*-R-

| | PAn | Siraya | | Taivuan | | Makatau | |
|---|---|---|---|---|---|---|---|
| | | Sinkang | Tohkau | Wanli | Matau | Lower Tamtsui | |
| (3) | *D | s | s | r~d | r~d | r~d | |
| e.g. | *Daya | saija | saija | raija | – | – | 'east' |
| | *lahuD | raos | raos | raur | – | – | 'west' |
| | *DapaN | sapal | – | rapan | – | – | 'foot' |
| | | sa | sa | ra, da | ra, da | ra, da | 'and' |
| | | hiso | hiso | hairo, ro | haijro, do | – | 'if, as' |
| | | posoh | – | poroh | – | – | 'land' |
| | | maisisang | – | – | – | maeraerang | 'magistrate' |
| (4) | *-k- | -k- | -k- | Ø | Ø | -k~Ø | |
| | *-S-, *-R-[5] | -g~~h- | -g- | Ø | Ø | – | |
| e.g. | | akosaij | akusiuo | ausaij | ausaij | akusai[6] | 'not have' |
| | | tarokaij | – | taroaej | – | tarauwei[7] | 'name' |
| | *DuSa | soso(h)a | – | – | – | – | 'two' |
| | *baqeRu | vaho | – | – | – | – | 'new' |
| | | dagogh | dagogh | daoh | daoh | – | 'price' |
| | | ligig | – | liih | – | – | 'sand' |
| | | matagi-vohak | – | mataij-vohak | – | – | 'regret' |

As shown in the examples above, *s* in Siraya corresponds to *r* or *d* in Taivuan and Makatau in word-initial or final position, derived from PAn *D or *d, as illustrated in (3). As shown in (4), *k* or *g~h* in Siraya is lost in Taivuan in word-medial position. The *k* in Siraya is derived from PAn *k, and the *g~h*, which is interpreted as velar fricative *x* by Adelaar (1999), is derived from PAn *S or *R (Adelaar 1999:334).

Rule (3) shows Siraya in contrast with Taivuan and Makatau, while Rule (1) shows Taivuan in contrast with Siraya and Makatau. Rule (4) shows that the medial velar obstruents *k* and *g* [x] are lost in Taivuan, but retained in both Siraya and Makatau.[8] It is an innovation in Taivuan.

[5] The symbol 'g,' 'gh' or 'h' [x] in Siraya is historically derived from PAn *S or *R; see Table 6 below for examples. There is extremely limited vocabulary in all Sinkang manuscripts, and I can identify few certain PAn cognates for such a derivation in these Sinkang manuscripts , such as *sosoha* or *sosoa* 'two' and *v(ah)o* 'new' found only in Sinkang.

[6] Although the term 'not have' is unavailable for Lower Tamtsui, the form *akusai* is cited for Taikiattian, another village of Makatau in Ogawa's list, and the form *akosaij* appears in a manuscript from Gutiaupo, which might be another village of Makatau or Siraya (see Tsai 2002, Appendix 1, p.3).

[7] The personal name *tarauwei* appears in a Katin text (Murakami 1933:144). This shows that *-k-* is occasionally lost in certain Makatau subdialects.

[8] Ogawa investigated the Piathau dialect of Siraya in 1921. But he (Ogawa 2006:354) cited the form *li'igh* 'sand' for Siraya dialects, probably taken from *St. Matthew*; see Table 6 in §3 below.



Matau was considered to belong to the Siraya group by Chinese and Japanese scholars, as indicated in Ogawa's (1917) grouping and Tsuchida et al.'s map (1991:ix). However, the phonological innovations in Matau generally indicate that it belongs to the Taivuan group rather than the Siraya one: Matau *l > Ø or *h*, as in *telu > *tao* 'three', *lima > *hima* 'five'; and *D > *r*, *d* (see Table 2, above).

Both Matau and Wanli villages of Taivuan are in the coastal and transitional area, geographically close to Sinkang and Tohkau villages of Siraya. The phonological differences, especially (3) and (4), between Siraya, Taivuan, and Makatau, are quite regular.

### 2.2 Morphological evidence

In addition to the phonological differences, a type of morphological difference can be observed, as shown below:

**Table 3:** Sirayaic future markers

| Siraya | | Taivuan | | Makatau | |
|---|---|---|---|---|---|
| Sinkang | Tohkau | Wanli | Matau | Lower Tamtsui | |
| -ali, -ili | -ati, -ili | -ah | -ah | -ani | 'future' |

The verbal suffix indicating future also shows that Taivuan differs from both Siraya and Makatau. Note that *-ani* in Makatau regularly corresponds with *-ali* in Siraya, another bit of morphological evidence indicating that Matau is Taivuan instead of Siraya.

### 2.3 Lexical evidence

Tsuchida et al. (1991:7–8) pointed out the following lexical differences among the three groups, as based on Ogawa's comparative wordlist:

**Table 4:** Lexical differences in Sirayaic

| Siraya | Taivuan | Makatau | |
|---|---|---|---|
| siraya | taivoan | makatao | 'aborigines (self-appelation)' |
| it | tau | lihu | 'liquor, wine' |
| pak | harau | buka | 'rice' |

However, there has also been a fair amount of lexical borrowing between the groups. The term for 'aborigines' in Matau is *siraiya* instead of the anticipated form *taivoan*. The term for 'wine' in Kongana (Siraya) is *diho* instead of the anticipated form *it*; see Texts 5 and 11 published in Li (2002). Similarly the term for 'wine' in *St. Matthew* is also *dihou*. The term for 'rice' in Laupi (Makatau) is *pak* instead of the anticipated form *buka*. In short, unlike phonological innovations, lexical evidence is not very reliable for language subgrouping.

### 2.4 Summary

For the sake of convenience, let's summarise what we have discussed so far in Table 5.

**Table 5:** Summary of Sirayaic Reflexes of PAn *l, *N, *D, *-k- and *-S-/*-R-

| | PAn | Siraya | Taivuan | Makatau |
|---|---|---|---|---|
| (1) | *l | r | Ø~h | r |
| (2) | *N | l | l | n |
| (3) | *D, *d | s | r~d | r~d |
| (4) | *-k- | -k- | Ø | -k~Ø |
| | *-S-, *-R- | -g- | Ø | – |

In addition, Taivuan has a suffix *-ah* indicating 'future', which is different from both Siraya *-ali/-ili* and Makatau *-ani*.

Two phonological innovations, (*l > Ø or *h*, and loss of medial consonants *-k, *-S- and *-R-), plus one piece of morphological evidence, show that Taivuan is in maximal contrast with Siraya and Makatau. Makatau differs from both Siraya and Taivuan only in Rule (2), which is not significant. It shares Rules (1), (4) and the correspondence seen in Sirayaic future markers with Siraya, but it shares only Rule (3) with Taivuan. As a preliminary conclusion, it seems that Taivuan constitutes a first split from the Sirayaic group, while Siraya and Makatau are more closely related, as shown below:[9]

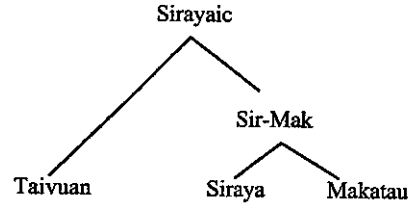



**Figure 1:** Sirayaic tree diagram based on numbers of shared phonological innovations

It seems reasonable to classify Taivuan as a separate language, but it is not clear whether Siraya and Makatau can be treated as separate languages until we find more linguistic difference.

## 3 The Dutch Missionary documents

Is *St. Matthew* based on Siraya or Taivuan? The following vocabulary found in *St. Matthew* seems to indicate that it is Taivuan:

[9] Adelaar's (pers. comm.) interpretation of Table 5 is that there is no clear subgrouping pattern emerging. He also notes that in Rule (2) in Table 5, Makatau *n* would be an innovation, rather than a retention, if PAn is reconstructed as *L.



(5) *D, *Z > r, d, cf. (3) above

*Daya > *reia* 'east', *lahuD > *raour* 'west', *DapaN > *rahpal* 'foot', *likuD > rikour 'back', *ZaNum > *raloum* 'water', *da* 'and' (*sa* in Sinkang), *pourough* 'land' (*posoh* in Sinkang), *haouroung* 'to steal' (*haosong* in Sinkang), *laumari* 'coins' (*lomasi* in Sinkang), *ka-harim-auh* 'will be forgiven' (*ka-hasim-ing* in Sinkang)

(6) Loss of *-k-* and *-g-* [x] < *S, cf. (4) above

- Loss of *-k-*: *(i)aku > *jau, -au* '1sg'; *aousi, akousi* 'not have'

The loss of intervocalic *-k-* seems to be restricted to a few special grammatical categories, namely personal pronoun and negative, and note the free variant of *aousi ~ akousi*. It does not apply to ordinary vocabulary, e.g. *bukeS > *voukugh* 'hair', *likuD > *rikour* 'back' and *takut > *takout* 'to fear'.

- Loss of *-g-*: *daœugh, daœuh* 'price' (*dagogh* in Sinkang); *li'igh, liih* 'sand' (*ligig* in Sinkang)

Exceptions (*-S- is retained as *-h-*): *CaSiq > *t<m>ahy* 'to sew', *DuSa > *rouha, douha* 'two'

Both the phonological innovations, *D, *Z > *r, d* and loss of intervocalic *-k-* and *-g-* [x], and the suffix *-au* or *-auh* 'future' (cf. Table 3 above) indicate that *St. Matthew* was most likely based on some dialect of Taivuan, rather than Siraya proper. However, that there are exceptions to the loss of intervocalic *-k- and *-S- > -g-, -h- [x] seem to indicate there might be mixture of dialects in *St. Matthew*, as suggested by Adelaar (pers. comm.), it is 'most likely that *the Gospel* text [=*St. Matthew*] was not the product of one person only: this is clear from the text itself, and ... that there was a committee deciding over the final edition.' In fact, it is stated in the introduction to *St. Matthew*:

> Hence, too, it follows that the present Translation can be of service to only a few, though populous Villages, such as Soulang, Mattauw, Cinckan (Sinkang), Bacloan, Tavokan, Tevorang, and perhaps also to some of the people in Dorko and Tilocen. These are the places where the work has been carried on for the longest time (p.xiii).

Similarly, it is most likely that *Formulary* (Gravius 1662) was also based on some dialect of Taivuan, whereas *Utrecht Manuscript* was based on a dialect of Siraya, as shown in the following comparison:[10]

[10] Both Adelaar (1997) and Tsuchida (1998) make a comparison of these three Dutch missionary documents. Adelaar (2006) notes two main dialects, the 'Gospel dialect' and the 'Utrecht Manuscript dialect.' He considers *St. Matthew* and *Formulary* to represent the same Gospel dialect.

**Table 6:** A comparison of Utrecht, Matthew and Formulary

| | | Utrecht Ms | St. Matthew | Formulary | |
|---|---|---|---|---|---|
| (7) | *D | s | r, d | r, d | |
| e.g. | *Daya | taga-seia | reia | – | 'east' |
| | *lahuD | taga-raos | raour | – | 'west' |
| | *DapaN | sapal | rahpal | rahpal | 'foot' |
| | *DuSa | so-soa | dou-rouha | rou-rouha | 'two' |
| | *likuD | ricos | rikour | rikour | 'back' |
| | | sama | rama | dama | 'father' |
| | | sa | ra | ra | 'but' |
| | | soo, sou | rou | dou, rou | 'if, as, when' |
| | | isang | irang | irang | 'great, large'[11] |
| | | sasim | rarim | mou-rarim | 'down, below' |
| | | pesanach | paeraenaeh | paeraenaeh | 'tree' |
| | | massou | – | marou | 'corn' |
| | | ka-pousoch-ang | pourough | pourogh | 'land' |
| (8) | *-k- | -k-, Ø | -k-, Ø | Ø | |
| e.g. | | acoussey | akousi, aousi | aoussi | 'not have'[12] |
| | *(i)aku | iau, -au | jau, -au | jau | 'I' |
| (9) | *S/*R | -h-, Ø/-g-, -h- | -h-, Ø/-h-, ', Ø | -h-, Ø/h, ', Ø | |
| e.g. | *kaSu | cau [kaw] | kow | kow | 'thou' |
| | *DuSa | so-soa | dou-rouha | rou-rouha | 'two' |
| | *CaSiq | t<m>ahy | – | – | 'sew' |
| | *waRi | wagi | wæ'i | wæ'i | 'sun, day' |
| | *wiRi | ougi | u-i | ou-i | 'left' |
| | *baqəRu | vacho | vahæu, va'æu | vahæu,va'æu | 'new' |
| | *kiRim | k<m>igim | k<m>i'im | k<m>i'ym | 'seek' |
| | *kaRaC | k<m>agat | – | – | 'bite' |
| | | ligig | li'igh | – | 'sand' |
| | | ma-dagoa | – | dæeua | 'entirely' |
| (10) | | -a, -al, -ale | -ah, -auh | -a, -ah, -al | 'future' |

The examples in (7) show that the language of *Utrecht Manuscript* is similar to that of the Siraya-speaking villages of Sinkang and Tohkau, while the language of *St. Matthew* and *Formulary* is similar to that spoken in the Wanli and Matau villages of Taivuan or Lower Tamtsui village of Makatau, as illustrated in (3) above. However, the pronominal forms, first person and second person singular, and the negative in (8) and (9) do not show much

[11] *isang/irang* is the root of the derivation *maisisang/maeraerang* in Table 2. It is cognate with *ma-'iDang* 'big' in Puyuma, as suggested by Tsuchida (pers. comm.).

[12] As based on Adelaar's (pers. comm.) research, these are two different negatives: *assei, assey* or *assi* means 'no(t),' while *akousi* means 'not have,' which is derived from *akou-* 'to have' + *(a)ssi* 'not.' Still another negative is *inang* 'will not.'



difference between the groups. The medial *-k-* is kept or lost in the negative forms in *Utrecht Manuscript* and *St. Matthew,* but lost in *Formulary*. All the three groups have zero reflex of *S in the form of *kaSu. On one hand, *Utrecht Manuscript* has zero reflex of *S in the form of *DuSa, while *St. Matthew* and *Formulary* have retained *h* as its reflex. On the other hand, while *-g-*, the reflex of *S or *R is retained in the lexical forms *wagi* 'sun,' *ougi* 'left,' *k<m>igim* 'seek,' *ligig* 'sand' and *ma-dagoa* 'entirely' in *Utrecht Manuscript*, it is lost in the forms *wæ'i*, *u-i*, *k<m>i'im*, and *li'igh* in *St. Matthew* and the similar forms in *Formulary*. That is to say, there is some conflicting evidence. But several examples seem to indicate that *Utrecht Manuscript* is based on a Siraya dialect, while *St. Matthew* and *Formulary* are based on a Taivuan dialect. The suffix *-ah* 'future' in (10) also indicates that *St. Matthew* and *Formulary* are based on a Taivuan dialect. It seems clear that there is some dialect mixture in both *St. Matthew* and *Formulary*.

## 4 Relative chronology and subgrouping

The sound change PAn *D, *d, *Z > *s* in Siraya, > *r* ~ *d* in Taivuan and Makatau must have taken place prior to the Dutch occupation of Taiwan (1624–62), as the phonological difference is manifested in the Dutch missionary documents: *s* is found in the *Utrecht Manuscript* vs *r* ~ *d* in *St. Matthew* and the *Formulary*, as discussed in the preceding section.

The change *-S-, *-R- > *x* (written as *g, gh, h*) or Ø may have started at the beginning of the 17th century because the rule applies to some lexical forms, but not to the others containing the consonant even in the same set of language data as recorded by the Dutch missionaries; see Table 6 above.

PAn * l is retained as *r* rather than *h* or Ø (see (1) above) in *St. Matthew*, e.g. *lahuD > *raour* 'west,' *piliq > *peri* 'to choose,' *kalih > *k<m>ari* 'to dig'. That is to say, PAn *l was still retained as *r* in mid 17th century when *St. Matthew* was translated. In fact, it was still retained in a Wanli text dated 1770, as in *lahuD > *raur* 'west,' and as *h* in a Matau text dated 1781, as in *likuD > *mi-likoh* 'to return.' It was not lost until much later when the Japanese started to investigate the languages of the southwestern plains in 1897. So the sound change *l > *h* or Ø was a late innovation limited to Taivuan.

If we take the relative chronology of the sound changes into consideration, then the first split of the Sirayaic group should be Siraya, as shown in Figure 2 below.

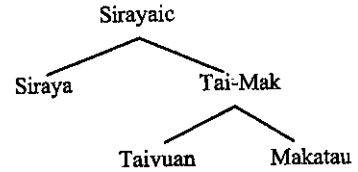



**Figure 2:** Sirayaic tree diagram based on the chronological order of phonological innovations

However, if we compare the number of shared phonological innovations, then the first split would be Taivuan, as shown in Figure 1 above. Which type of evidence should carry more weight: the earliest phonological innovation or the number of shared phonological innovations? It seems to me the former should carry more weight in a subgrouping hypothesis.

## References


Adelaar, K. Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36.2:362–397.

—— 1999. Retrieving Siraya phonology: a new spelling for a dead language. In Elizabeth Zeitoun and Paul Li, eds *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 313–354. Institute of Linguistics (Preparatory Office). Taipei: Academia Sinica.

—— 2006. Siraya dialogues. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho, eds *Streams converging into an ocean: festschrift in honor of Professor Paul Jen-kuei Li on his 70th Birthday*, 665–686. Language and Linguistics Monograph Series Number W-5. Institute of Linguistics. Taipei: Academia Sinica.

—— 2007. Siraya, Taiwan's oldest written language. In *The margins of becoming: identity and culture in Taiwan*, ed. by Carsten Storm and Mark Harrison, 19–34. Wiesbaden: Harrassowitz.

Chen, Bien-horn. 2001. *Notes on vocabulary of Siraya version of Matthew* (in Chinese). Association of the Plain Tribes in Tainan.

—— 2005. *Notes on formulary of Christianity in Formosan Siraya dialect—with translation in English, Chinese and Dutch* (in Chinese). Private Circulation.

Gravius, Daniel. 1661. *Het Heylige Euangelium Matthei en Johannis Ofte Hagnau Ka D'llig Matiktik, Ka na sasoulat ti Mattheus, ti Johannes appa.* Amsterdam: Michiel Hartoch.

—— 1662. *Patar Ki Tna-'msing-an Ki Christang, ka Taukipapatar-en-ato tmaeu'ug tou Sou KA MAKKA-SIDEIA, 't Formulier des Christendoms Met de Verklaringen van dien, Inde Sideis-Formosaansche Tale.* Amsterdam: Michiel Hartogh.

Li, Paul J. K. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* [BIHP] 48.1:1–92.

—— 1981. Reconstruction of proto-Atayalic phonology. *BIHP* 52.2:235–301.

—— 1993. New data on three extinct Formosan languages. *BIHP* 63.2:301–323.

—— 2002. Preliminary interpretations of the 15 recently uncovered Sinkang manuscripts (in Chinese). *Taiwan Historical Research* 9.2:1–68.

—— 2006. The languages of the plain tribes in southern Taiwan—with remarks on the linguistic position of Matau (in Chinese). In Chuen-rong Yeh, ed. *Constructing Siraya: Selected Conference Papers*, 17–38. Tainan: Government of Tainan.

Murakami, Naojirô. 1933. Sinkan manuscripts. Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University, vol. 2, no.1. Formosa: Taihoku Imperial University.

Ogawa, Naoyoshi. [1917]. Siraia, Makatao, Taivoan. Manuscripts.

—— 2006. *A comparative vocabulary of Formosan languages and dialects*, ed. by Paul Li and Masayuki Tohoshima. Asian and African Lexicon Series No.49. Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Steere, Joseph Beal. 1874. [The aborigines of] Formosa. *Journal of the American Geographical Society of New York* 6:302–334. New York.





—— 2002. *Formosa and its inhabitants*. Edited by Paul Li. Institute of Taiwan History (Preparatory Office), Academia Sinica.

Tsai, Cheng-ui. 2002. Transfer of land ownership of the Sinkang villagers in Tianliao district during the Ching Dynasty (1736–1895) (in Chinese). Unpublished MA thesis, National Tainan Normal University.

Tsuchida, Shigeru. 1996. Personal pronouns of Siraya (Formosa). In Bernd Nothofer, ed. *Reconstruction, classification, description. Festschrift in honor of Isidore Dyen*, 231–247. Hamburg: Abera Verlag.

—— 1998. English index of the Siraya vocabulary by Van der Vlis. *Studies of Taiwan Aborigines* 3:281–310.

Tsuchida, Shigeru, Yukihiro Yamada and Tsunekazu Moriguchi. 1991. *Linguistic materials of the Formosan Sinicized populations I: Siraya and Basai.* University of Tokyo.

Utrecht Manuscript. n.d. Vocabularium Formosanum. MS. Utrecht University Library.

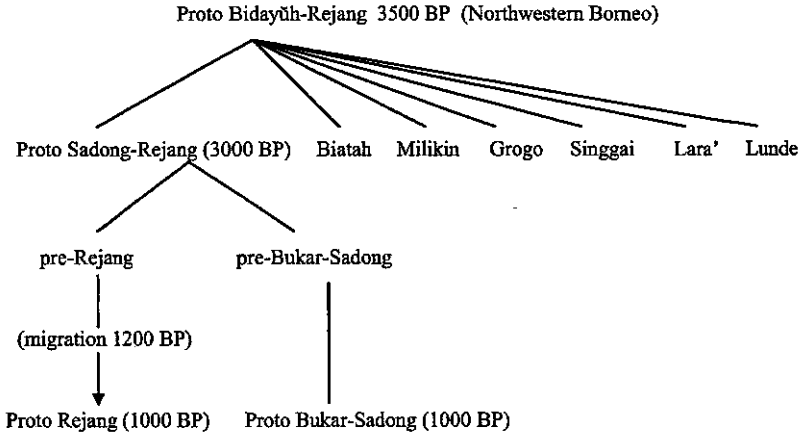
Van der Vlis, C.J. 1842. Formosaansche woorden-lijst, volgens een Utrechtsch Handschrift. Voorafgegaan door eenige korte aanmerkingen betreffende de Formosaansche taal. *Verhandelingen van het Bataviaasch Genootschap* 18:437–452 (Notes), 453–483 (Glossary), 484–488 (Conversation).

# 24 *Out-of-Borneo subgrouping hypothesis for Rejang: re-weighing the evidence*

RICHARD McGINN

This paper revisits a subgrouping hypothesis for the Rejang language of Sumatra presented in McGinn (2003), and offers a new hypothesis based on new evidence and on criticisms received from Austronesianists. The hypothesis to be defended is shown in (1).

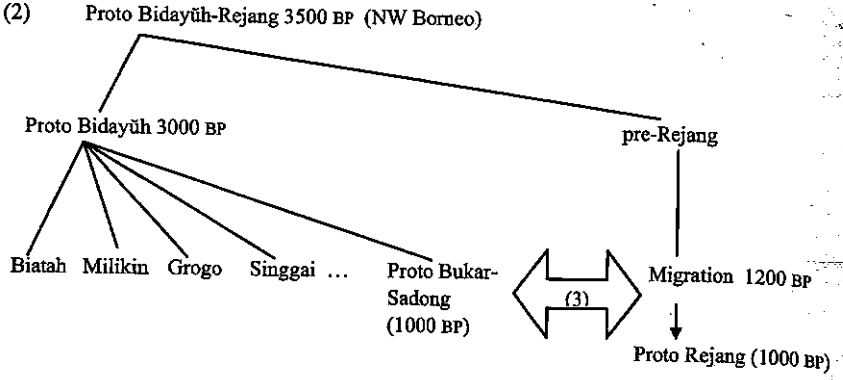
(1)




The subgrouping hypothesis shown in (1) above replaces the proposal by this writer (2003) illustrated in (2).

Alexander Adelaar and Andrew Pawley, eds
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 411–424.
Pacific Linguistics, 2009.

(2)




In earlier publications (McGinn 1999, 2000, 2003), we attempted to construct an 'out-of-Borneo' subgrouping hypothesis for Rejang, and in particular, McGinn (2000) suggested the possibility that Generalised PMP *a Raising, illustrated in (3a-c) below, might constitute a set of innovations shared by Rejang and one or more Land Dayak languages. The following changes were among the first to distinguish pre-Rejang from PMP.

| | | | |
|---|---|---|---|
| (3) | a. | *-ˈVCaC# > ˈVCəC[-velar]# | PMP unstressed *a > /ə/ except before velars. |
| | b. | *-ˈVCa# > ˈVCə# | PMP unstressed *a > /ə/ in word-final position. |
| | c | *-ˈVCaC# > VCˈVC# | PMP stress pattern (symbolised by preceding ˈ) shifted to the word-final syllable. |

These three changes were originally posited by McGinn (1997) solely for the sake of supporting the regularity hypothesis against the challenges posed by Blust's pioneering article (1984), and only later were those same changes used as the basis for subgrouping arguments (McGinn 2000, 2003). The Stress Shift change (3b) was motivated by several factors, the most compelling being that it served to increase the regularity of the other sound changes, thereby explaining many apparent irregularities. The most important change for subgrouping purposes, however, is (3a) because the conditioning factor ('except before velars') is unusual and phonetically unmotivated. (See §4). Moreover, a comparable change had been recorded for the Tapŭ and Mawang (labeled Měntu) dialects of Bukar-Sadong Bidayŭh by Court (1967a) and Topping (1990), and confirmed by this writer during field work on six dialects[1] of Bukar-Sadong conducted in 2000 and 2001.

Exploring the possibility that (3a) represented a shared innovation linking Rejang and Bukar-Sadong, McGinn (2003) provided further evidence to recommend a subgrouping hypothesis, including counterparts of (3a,b,c) (see (7) below). But there was also contradictory evidence in the form of an apparently regular sound change, namely *l > r, which affected all Bukar-Sadong dialects and several other Bidayŭh languages, but not Rejang. Adjusting the hypothesis to accommodate the contrary evidence, McGinn (2003) concluded that *l > r must have preceded *a Raising in Bukar-Sadong, and if so, *a Raising was not a shared innovation after all; therefore, *a Raising must have been the result of rule borrowing after a long period of language contact (presumably in Borneo).

[1] The six dialects surveyed were Tibakang, Ranchan, Mujat, Tapŭ, Mawang and Badŭp. Not included was Kampung Bunun, a Bukar-Sadong language with a six-vowel system as described by Asmah Haji Omar (1983:445), who claims that 'the phoneme /a/ is phonetically [ʌ].'



Problems with the borrowing theory were soon pointed out by Robert Blust (personal communication)[2] and David Zorc (2006:509); and Adelaar (2007) questioned my assumption that sometimes Bidayŭh /l/ directly reflects PMP *l. However, the criticisms were not wholly satisfactory, since taken together, they posed a paradox for the Comparative Method: the basic comparison (3a) seemed to resist explanation by either chance, borrowing or common inheritance. In particular:

(4) a. Chance is ruled out by the unusual conditioning of (3a).
b. Borrowing of (3a) is ruled out because unsupported by independent lexical evidence.
c. Direct inheritance of (3a) is ruled out owing to the contradictory evidence of *l > /r/ in many Bidayŭh languages but not Rejang.[3]

If (4a-c) offers a paradox, then one or more of the statements must be hiding a false assumption. It is theoretically impossible for all three to be valid.

## 1 The status of PMP *l > /r/ in Oceanic

A possible way out of the dilemma posed by (4a-c) follows from a comment by John Lynch which implies that (4c) may be erroneous.[4] The following was his reaction to the treatment of *l > r in McGinn (2003).

(5) At least in Oceanic, the *l > r change is very common, and in many cases not diagnostic of subgrouping. In other words, it seems to be a natural change which could easily occur independently. Looking at Tryon's New Hebrides (Vanuatu) survey, for example, there are a couple of cases in Malakula where groups of languages show *l > r but where, on other criteria, these languages subgroup with other languages which show *l > l. In Tanna, *l and *r (and *R when reflected) have merged, in some languages as l, in others as r. Similar kinds of things have occurred, as far as I am aware, in other areas of Oceanic.
While I would not dismiss *l > r as an innovation, I think it is a weak one — rather like palatalisation of *t before *i: something that doesn't surprise you when you see it, as it happens so often; and therefore something not to be given much weight if there are other, less expected, innovations which would support a different subgrouping theory.
(John Lynch pers. comm.)

[2] Blust (pers. comm.) advised the following experiment: 'Line up the basic vocabularies of Rejang and any two or three L(and)D(ayak) languages and pull out all of the exclusively shared lexical innovations. Do you find any? If not, this is reason to suspect that Rejang is not a recent arrival in Sumatra from the LD area.' In fact, so far we have found vanishingly few shared lexical innovations (listed in McGinn 2003), to which can be added Tibakang laca? 'soaked' and Rejang ləcəa? 'soaked'.

[3] This last point bears repeating for clarity's sake: the Bukar-Sadong version of (3a), namely (7ii), while shared with Rejang, is not shared with other Bidayŭh languages; and Bukar-Sadong *l > /r/, while shared with many (but not all) Bidayŭh languages, is not shared by Rejang.

[4] The following is the statement that Lynch was addressing: 'The Bukar-Sadong version of PMP *a Raising is not found in other Bidayŭh dialects, in contrast to *l > /r/ which is fairly widespread. It follows that *l > /r/ must have preceded *a Raising in Bukar-Sadong; and therefore no version of *a Raising can possibly be assigned to any subgroup containing Proto Rejang and Proto Bukar-Sadong as members. Our most interesting comparison, therefore, must be due to borrowing (language contact) or chance (phonetic drift). But the likelihood of chance must be considered extremely low given the unusual nature of the conditioning (*a underwent raising 'except before velars') in exactly these two languages. Therefore, I shall argue for borrowing as the more likely explanation.' (McGinn 2003:49)

If the above comment is relevant to the concerns of this paper, then what is needed is evidence that *l > r in Proto Bidayŭh is indeed a weak innovation. Accordingly, we shall investigate the second conjunct in the following statement to determine if perhaps it might be too strong.

(6) The characteristic features of the Land Dayak languages which Hudson mentions include distinctive numerals for 'eight', 'nine' and 'ten' *and /r/ as the reflex of PAN *l* (Kroeger 1998:150–151, citing Hudson 1978; emphasis ours)

According to Paul Kroeger (1998:139), 'The Land Dayak languages do not appear to be closely related to any other language in Sarawak, but they do form a linguistic subgroup with the many Land Dayak languages spoken across the border in West Kalimantan (Indonesian Borneo)'. This statement provides the context for Hudson's assertion, quoted above, claiming that PMP *l > /r/ is an important diagnostic feature of Land Dayak (= Bidayŭh) languages. Examination of the available evidence within this language group suggests, however, that the pattern of regular *l > /r/ is not as widespread throughout the family as was initially thought; therefore, any attempt to construct a protolanguage for the Bidayŭh group must do so on the basis of something other than regular *l > *r. This claim will be substantiated below.

## 2 The status of PMP *l > /r/ in Bidayŭh languages

In fact, the reflexes of PMP *l are problematic among Bidayŭh languages. Five sources of evidence support this claim.

First, the data displayed in Ray (1913) indicates that whereas most of the thirteen Land Dayak languages surveyed consistently show expected PMP *l > r, two of the languages—Grogo and Milikin—regularly retain PMP *l as /l/.

Second, in the Biatah-Bidayŭh language spoken in Sarawak, whereas most dialects regularly reflect PMP *l as /r/, the Mbaan dialect regularly retains PMP *l as /l/. (Kroeger Ms, 1994:22)

Third, Adelaar (2007), citing unpublished field notes, suggests that in another Bidayŭh language of West Kalimantan (Sungkung), /l/ appears to reflect intermediate *r which itself reflects the merger of PMP *R, *r and *l, implying that PMP *l > *r was ancient and /l/ a recent innovation. If this model is proven by future research to account for the organic /l/s in other Bidayŭh languages, it would strengthen the case for *l > /r/ as a diagnostic for Bidayŭh languages, and weaken if not break the case for a Rejang connection. On the other hand, logically and phonetically it is just as easy to assume that *R, *r and *l merged as intermediate *l in some language, then split into /l/ and /r/ in daugher languages, or changed unconditionally to /r/. Clearly, what are needed are detailed historical phonologies for individual dialects, where the results can be tested against the strict demands of the regularity hypothesis (see McGinn 1997:91–92 for discussion). In the meantime, Lynch's comments about Oceanic, cited above, remain as a valid cautionary note.

Fourth, among the six Bukar-Sadong dialects surveyed by this writer during field work in 2000 and 2001 and partially displayed in McGinn (2003), the reflexes of PMP *l seem to vary unpredictably between /l/ and /r/ for five of the dialects; and for the sixth dialect (Mujat) there is a three-way alternation between /l/, alveolar /r/ and uvular /ɣ/. In fact, out of a total of thirty-one potential reflexes of PMP *l, Proto Bukar-Sadong shows fifteen cases of *r (see McGinn (2003) examples 38, 39, 46, 103, 107, 108, 136, 158, 163, 168, 186, 187, 200, 242) and fifteen of *l (see the same source for examples 12, 19, 27, 50, 66, 93, 101, 105, 109, 135, 149, 153, 189, 233, 243). Whereas it is likely that some (perhaps all) putative B-S /l/ < PMP *l are Malay borrowings, this remains to be demonstrated in a definitive historical phonology of the Bukar-Sadong language group. For starters, a candidate for borrowing includes one member or the other of the doublet /jʌrʌn/ 'road' and /jalʌn/ 'walk' in the Tapŭ, Ranchan, Bedŭp and Mawang dialects, which show contrastive /r/ and /l/ corresponding to PMP *l; another may be the PBS outcome *m-ilih 'choose' corresponding to PMP *piliq 'choose'. PBS *milih contains suspicious /l/ and suspicious -/h/ (expected -/ʔ/), exactly like Malay *pilih* 'choose'. Other possibly problematic examples include PMP *bales = PBS *malɔs 'reply'; PMP *gatel = PBS *gatəl 'scratch', PMP *palaqepaq = PBS *kilapa 'palm frond', and PMP *balaŋa = PBS *b[a,i]laŋaʔ 'clay pot'.



Fifth, Bukar-Sadong dialects overwhelmingly agree with respect to /r/ and /l/ as apparent reflexes of PMP *l. Therefore, even if all the unexpected /l/s are the result of massive borrowing, the borrowing would have occurred very early, in Proto Bukar-Sadong; otherwise the dialect uniformity is unexplained.

These comparative problems justify the decision to re-weigh the subgrouping value of *l > /r/ in relation to the Bidayŭh languages in general, and Bukar-Sadong in particular. We now have grounds to set the problem to one side; it is an anomaly to be investigated, not yet a counterexample; it is too weak to bear any weight for subgrouping purposes. By implication, much more importance can and should be given to the set of comparisons (3a,b,c), especially (3a). Just how much weight to assign to (3a) will be taken up in the next section.

## 3 Back to the basic comparison

As mentioned above, the most important change for subgrouping purposes is (3a) because the conditioning factor ('except before velars') is unusual and phonetically unmotivated. (See next section for arguments.) If so, then the discovery that both the Rejang dialect group in Sumatra and the Bukar-Sadong dialect group in Sarawak, show unmistakable traces of a similar change in their phonological histories, constitutes *prima facie* evidence for a shared innovation.

However, change (3a) is directly relevant only for Rejang, because it resulted in partial merger of PMP *a and *e as *ə (schwa). By contrast, the Bukar-Sadong counterpart did not result in merger. Therefore, to be consistent with hypothesis (1), the Bukar-Sadong facts require an additional (and perfectly natural) assumption to be added to the phonological history of Rejang, namely, that (3a) occurred in two steps as shown in (7ii) and (7iv) below.

| | | | | |
|---|---|---|---|---|
| (7) | i. | (3b) | *-'VCa# > 'VCə# | Unstressed *a > /ə/ in word-final position. |
| | ii. | (3a-1) | *-'VCaC# > 'VCʌC[-velar]# | Unstressed *-aC > *-ʌC except before velars. |
| | iii. | (3c) | Stress Shift (Language split) | Vowels in final syllables became stressed |
| | iv. | (3a-2) | *-VC'ʌC# > VC'əC[-velar]# | Stressed *ʌC > -əC in Rejang (partial merger) |

Thus, only the first step (7ii) is claimed to be a shared innovation, whereas the partial merger (7iv) occurred in pre-Rejang after language split.

### 3.1 Summary of PMP last-syllable **a* Raising in pre-Rejang

The following formula represents five early changes in the historical phonology of pre-Rejang.

(8) i.

| PMP | | pre-Rejang | (Unstressed *a > *ə except before velars.) |
|---|---|---|---|
| *a | > | *ə / ˈVC__(C[-velar])# | |

ii.

| | PMP | | pre-Rejang | PMP | Kebanagung | Gloss |
|---|---|---|---|---|---|---|
| a. | *a | > | *ə / ˈVC __ # | *duha | dui | 'two' |
| b. | *a | > | *ʌ / ˈVC __ C[-velar] # | *taŋan | *taŋʌn | 'hand' |
| c. | *-aw | > | *əw | *danaw | danɔa | 'lake' |
| d. | *-ay | > | *əy | *matay | matɔe | 'die' |
| e. | *ʌ | > | *ə | *taŋan | taŋən | 'hand' |

### 3.2 PMP last-syllable **a* Raising in pre-Bukar-Sadong

The set of pre-Rejang changes shown by the formula in (8i) *almost* works for reconstructed pre-Bukar-Sadong as well—only the partial merger of *ʌ and *ə is missing (8ii,e). Consider the following set of changes in Bukar-Sadong, illustrated by the Tibakang dialect.

(9)

| | PMP | | pre-Bukar-Sadong | PMP | Tibakang | Gloss |
|---|---|---|---|---|---|---|
| a. | *a | > | *ə / ˈVC __ # | *duha | duəh | 'two' |
| b. | *a | > | *ʌ / ˈVC __ C[-velar] # | *taŋan | tʌŋʌn | 'hand' |
| c. | *-aw | > | *əw . . . > u | *danaw | danu | 'lake' |
| | *-ay | > | *əy . . . > i | *punay | puni | 'dove' |

To help explain all of these changes, we assume that pre-Bukar-Sadong (like pre-Rejang) had a Malay-type stress system: i.e. the accent fell on the ultimate when the penult was schwa; otherwise on the penult. Another assumption is that all contemporary Bukar-Sadong dialects have ultimate stress, again like Rejang.

#### 3.2.1 Neutralisation of PMP word-final *a in open final syllables

Both languages show evidence of early neutralisation of PMP *-a in open final syllables.

(10)

| PMP | Pre-Rejang | Pre-Bukar-Sadong | Tibakang | Gloss |
|---|---|---|---|---|
| *duha | *ˈduə | *ˈduə | duˈəh | 'two' |
| *mata | *ˈmatə | *ˈmatə | baˈtəh | 'eye' |
| *naŋa | *ˈnaŋə | *ˈnaŋə | naˈŋəh | 'fork of river' |
| *limə | *ˈlimə | *ˈlimə | liˈməh | 'five' |
| *nia | *ˈniʔə | *ˈniʔə | niˈʔəh | 'he/she' |

#### 3.2.2 Neutralisation of PMP word-final *-a in diphthongs

Both languages show evidence that *a raised to *ə in PMP *aw and *ay.



(11)

| PMP | Pre-Bukar-Sadong | Proto Bukar-Sadong and Tibakang | Proto Rejang and Rawas | Gloss |
|---|---|---|---|---|
| *danaw | *danəw | danu | daniw | 'lake' |
| *punay | *punəy | puni | punuy | 'dove' |
| *qatey | *atəy | ati | atuy | 'liver' |
| *hapuy | *apuy | apuy | apuy | 'fire' |
| *kahiw | *kaiw | kayu | kiiw | 'wood' |

#### 3.2.3 Raising of PMP *a in closed final syllables 'except before velars'

Data like that shown below is what first drew my attention to the comparison of Rejang and Bukar-Sadong. The unusual conditioning of PMP *a *except before velars* was first reported for the Mĕntu-Tapŭ dialect of Bukar-Sadong by Christopher Court (1967), and for the Musi dialect of Rejang by Robert Blust (1984).

(12)

| PMP | Rejang (Rawas) | Bukar-Sadong (Tibakang) | Gloss | McGinn (2003) (Appendix) |
|---|---|---|---|---|
| *bulan | buˈlən | buˈrʌtn | 'moon' | 38 |
| *quzan | uˈjən | uˈjʌtn | 'rain' | 214 |
| *tawad | taˈwəa | tawˈʌr | 'haggle' | 184 |
| *anak | aˈnak | aˈnak | 'child' | 3 |
| *bataŋ | baˈtaŋ | baˈtakŋ | 'trunk (of tree)' | 15 |
| *hasaq | aˈsah | ŋ-aˈsaʔ | 'sharpen' | 6 |

In the same Appendix, see also items: 13, 34, 44, 46, 93, 112, 146, 147, 165, 173, 182, 186, 203, 204, 217, 232 242.

This comparison is the strongest evidence of a greater-than-chance subgrouping relationship between Rejang and Bukar-Sadong. Section 4 provides the justification for this claim.

Finally, there are drift-theoretical comparisons between Rejang and Bukar-Sadong dialects which were listed in McGinn (2003) and repeated in Sets I and II of (13) below. David Zorc (2006:509) reviewed this evidence and judged it as 'indeed plausible' with respect to an out-of-Borneo migration theory for Rejang. Much more specifically, the following comparisons are also consistent with hypothesis (1) of this paper.

(13)

| Rejang and Bukar-Sadong | Widespread in Borneo | Shared by Malay |
|---|---|---|
| **Set I** | | |
| *Ca- > *Cə- in trisyllables | YES | YES |
| *qa- > Ø in trisyllables | YES | NO |
| *z > *j (except Rejang d- in 'road' and 'needle') | YES | YES |
| *-eq > -aC; elsewhere *-eC > *-əC | YES | POSSIBLY |
| *-q > *-ʔ | YES | NO |
| *-mb-, -nd- > -$m^b$-, -$n^d$-, etc. ('barred nasals')[5] | YES | NO |
| *-m, *-n > $^p$m, $^t$n, etc. (pre-stopped nasals)[6] | YES | NO |

[5] The distinction between Plain and Barred (post-stopped) nasals precludes the necessity of recognizing phonemic nasalised vowels. See Court (1967) and Scott (1964) for discussion in relation to Bukar-Sadong; see Coady and McGinn (1983) for analysis of a similar issue in Rejang.

| | | |
|---|---|---|
| **Set II = (7)** | | |
| Stress on final syllable | YES | NO |
| *-a > *-ə | PROBLEMATIC[9] | PROBLEMATIC[7] |
| *-aC > *-ʌC except before velars | NO | NO |
| **Set III (morphology)** | | |
| Loss of suffixes in a language with word-final stress | YES | NO |
| Retention of PMP completive infix *-in- reanalysed as passive morpheme | YES | NO |
| Loss of *p- and *b- in transitive active verbs, e.g. *piliq > *m-ilih* 'choose'; *pinzem > *m-inʒəm*; *bili > *m-irih* 'buy' | YES | NO |

In the aggregate, the evidence presented in McGinn (2000, 2003) and this paper offers compelling reasons to believe that the Rejangs originated in Borneo (rather than, say, Taiwan, the Philippines, Sulawesi, or the Malay peninsula). Furthermore, the evidence is consistent with the much stronger claim represented by (1), namely, that Rejang belongs in a lower-order subgroup with Bukar-Sadong Bidayüh.[8]

## 4 Arguments against a drift-theoretical explanation of the basic comparison

The crucial sound change (7ii) can and must be reconstructed independently for pre-Proto Rejang and pre-Proto Bukar-Sadong, and for no other languages (including no other Bidayüh languages), as far as is known at present. Therefore, the research question concerns how this comparison should be explained. The only possibilities are chance (phonetic drift) and shared innovation—implying direct inheritance from a common ancestor language, as illustrated in (1) above. (Borrowing has been excluded owing to the paucity of lexical evidence showing intimate contact between the two languages.) This section presents arguments against the drift-theoretical explanation.

The basic claim is that the conditioning of change (7ii), namely that the change occurred except before velars, is phonetically unexpected and therefore unlikely to have occurred in both Rejang and Bukar-Sadong as the result of mere chance. The crucial issue

[6] For discussion of Rejang's pre-stopped nasals, see Coady and McGinn (1983:442) and Voorhoeve (1955).

[7] Tadmor (2003) argues that *-a > *-ə* spread by borrowing from Sanskrit via Javanese during the Majapahit period (1293–1520), and subsequently affected scores of Malay dialects and numerous other western Austronesian languages, regardless of whether the affected vowel was stressed or unstressed, under the political sway of the Majapahit empire. Our claim is that Rejang and Bukar-Sadong underwent a similar change indpendently, and much earlier, which affected only unstressed vowels.

[8] A note on reflexes of PMP *j is in order owing to astute comments by a reviewer. Rejang dialects reflect PMP *j as -/g/- between vowels and -t/-k word-finally, with -/k/ being the most frequent; however, -/t/ < *-j is the outcome for the dialect judged by McGinn (2005) to be the most conservative (Rawas). The solution to this problem adopted by McGinn (2003, 2005) was to retain *-j at the level of Proto Rejang, as a direct retention from PMP. The reviewer noted that whereas most Bidayüh languages reflect PMP *j as -/d/- and -/d/, this is not the case for Lara', Bekati' and possibly Lundu, which reflect *j as -/g/- between vowels. It is apparent to me that the dialect/language splits in the Rejang and Bidayüh groups point to a shared retention. PMP *j was simply retained as *j at the highest level (Proto Sadong-Rejang). After pre-Rejang split from Proto Sadong-Rejang, the pre-Bidayüh languages developed independent reflexes for PMP *j as noted by the reviewer, and much the same thing happened in the Rejang group. This assumption accounts for the comparisons and preserves the hypothesis.



is whether the comparison (7ii) is *rare enough* to indicate a subgrouping hypothesis, given that the supporting evidence consists of common sound changes. Theoretically, the subgrouping value of (7ii) falls towards zero to the extent that its basis is phonetic and natural, but by the same token, if the basis is phonological and arbitrary, then its subgrouping value rises accordingly.[9] So why did final velar consonants block raising/neutralisation of PMP *a in pre-Bukar-Sadong and pre-Rejang?

Throughout, we shall assume the generalised version of PMP *a Raising shown in (3) above and analyzed further in (7) above, which involves three shared innovations in Rejang and Bukar-Sadong. Crucially, all of the raised reflexes of PMP *a that underwent Raising were unstressed; and the Raising of *-aC to *-ʌC occurred in all environments except before velars. Our proposal is that Stress Shift occurred after *a Raising had begun to spread, but before the spreading process was complete. In other words, Stress Shift (7iii) interrupted the spread of *a Raising (7i) and (7ii).

Recall that *a Raising only affected unstressed vowels—a phonetically well-motivated assumption. The primary change was probably (7i) affecting unstressed word-final position *-a before word boundary. Next, *a Raising spread to include word-final *-aC except when the final -C was a velar (*-q, *-k, *-ŋ). Left unchecked, the spreading should have generalised totally; so why, indeed, did the spreading stop? We doubt very much that it had anything to do with phonetic naturalness. The relevant question is: why did velar consonants check the spread of PMP *a neutralisation? Is it because velars offer more resistance to airflow from the lungs than, say, labials and alveolars? Does *a neutralisation require more air than the anticipation of a velar can provide? Such a line of questioning seems unlikely to lead to a satisfactory explanation.

A more likely explanation is that the spread of *a Raising was blocked by a competing sound change, namely, (7iii)—Stress Shift. This rule altered the stress pattern from trochaic to iambic, and in the process, would have affected negatively *any* rule in the process of spreading among unstressed vowels. This introduces the element of arbitrariness which is so important in a subgrouping argument. The outcome was 'unnatural' in the sense of Blevins (2004), since the synchronic rule shows neutralisation of a stressed vowel; however, the individual (and sometimes competing) sound changes which produced the outcome were all perfectly natural.

## 5 Phonetic and phonological effects in sound change

The explanation just offered has in part a phonetic basis and in part a phonological one. Phonetically, it is necessary to assume that *a neutralisation rules, such as (7i) and (7ii), affected only unstressed vowels. What cannot be motivated phonetically is the actual form of (7ii), namely, the fact that velar consonants blocked the spread of the change, whereas labials, dentals, alveolars, and even semivowels and zero, did not. See (11)–(12) above.

The phonological part of the argument benefits from the assumption that the changes raising PMP *a affected a phonological system, and that the system was disrupted by a competing prosodic change, Stress Shift. This is the assumption behind (3) and (7) above. If changes (7i-iv) were systemically connected, then it is convenient to assume that change (7i)—raising of *-a in open final syllables—was the primary change. After all, this change was clearly *phonologically* motivated; it completed the distribution of PMP *ə (schwa),

[9] This is essentially the form of the argument put forward by Blust (2006 and earlier work) in support of a subgroup he called Proto North Sarawak. See §7 and fn 11.

which did not occur word-finally in PMP. Second, a classic structuralist assumption holds that sound changes are regular because they tend to generalise (or spread) within a phonological system, allophone by allophone (Bloomfield 1929). The model allows for the situation that any generalizing sound change may compete with other sound changes, producing unexpected effects and even sometimes 'crazy rules' in contemporary languages (Bach and Harms 1972; Blevins 2004). Such rules are not caused by any lack of regularity of sound change, but by the effects of competing sound changes. As expressed by Blevins (2004:44–45), 'Changes which occur in the course of evolution are random ... and (do) not necessarily result in a more symmetrical, more stable, or generally improved phonological system.'

## 6 After-effects of rule (7ii) in Bukar-Sadong and Rejang

The Stress Shift change (7iii) had important drift-theoretical effects in the two languages. Most importantly for this paper, after the protolanguage split into pre-Rejang and pre-Bukar-Sadong, the output of (7ii), namely *-ʌC from PMP *-aC, developed differently. At one and the same time, however, both languages developed seven-vowel systems,[10] and vowel harmony rules which appear to have operated regressively at first, but evolved into synchronic phonological rules operating progressively (and somewhat unnaturally in the sense of Blevins (2004)).

### 6.1 Bukar-Sadong: a new vowel phoneme /ʌ/

In Bukar-Sadong, rule (7ii) added a new allophone [ʌ] which subsequently evolved into a new phoneme /ʌ/ (contemporary orthographic ĕ). *Ex hypothesi* the new phoneme /ʌ/ began as an allophone in word-final (unstressed) position before Stress Shift, and later, after becoming a stressed vowel, gained phonemic status owing to the effects of a vowel harmony rule. In particular, after Stress Shift had converted allophonic [ʌ] into a stressed vowel, it served as trigger for a harmony rule which targeted the destressed reflexes of PMP *a, e.g. *zalan > *jarʌn > /jʌrʌn/ [jʌrʌtn] 'road' (all dialects). A full analysis of this harmony rule remains for future research.

### 6.2 Rejang: merger of *[ʌ] with *ə

By contrast, in Rejang the outcome of (7ii), namely *-ʌC, merged with the reflex of PMP *-eC after the break-up of the protolanguage, becoming Rejang -əC in all dialects. This outcome converged with the outcome of rule (7i), which also partially merged PMP *-a and *-e as *-ə before splitting into Proto Rejang *-əy, *-i, and *-o (McGinn 1997, 2005). These changes yielded two further, and closely-related, effects: (a) schwa came to bear a heavy functional load in the inherited four-vowel system, and (b) the lexicon became governed by height harmony based on the feature [±low] (McGinn 1999:226), as follows. Firstly, all words containing the neutral vowel (schwa) became harmonised by default, since schwa was harmonic with every vowel. Secondly, words lacking a schwa underwent eight harmonic changes, e.g. *manuk > *monok chicken'; *lanit > *läɲät 'sky';

[10] Court (1967) ) and Topping (1990) ascribe seven-vowel systems to a number of Bukar-Sadong dialects, including the two non-peripheral (central) vowels we have transcribed as /ʌ/ and /ə/ (traditional orthographic ĕ and ŭ respectively). However, Topping uses the symbol ə to represent ĕ (my /ʌ/) and the symbol + to represent ŭ (my /ə/).



*sapu > *supu 'broom'; *tali > *tili 'rope' (McGinn 1997, 2005; cf. Blust 1984). In the process, Rejang added two new vowels to the phonemic inventory: mid-back /o/ and low front /ä/, owing to the effects of vowel harmony. (A third new vowel, mid front /e/, was added via borrowing from unknown sources (McGinn 2005), resulting in a seven-vowel system for Proto Rejang, and attested in contemporary Rawas.

An interesting twist is that Rejang's harmony rules applied more or less simultaneously with Stress Shift, affecting the newly stressed final vowels and de-stressed penult vowels. But phonologically, the pattern evolved into a set of inviolable 'crazy' rules. In contemporary Rejang, as first noted by Blust (1984), penultimate mid-vowels /e/ and /o/ always co-occur with like vowels in ultimate syllables. According to McGinn (1997, 2005), Rejang's synchronic mid-vowel harmony rule, which applies progressively, evolved from a historical rule that applied regressively. The synchronic rule is 'unnatural' in the sense of Blevins (2004), because unstressed vowels trigger harmony in stressed vowels; but as expected, the historical explanation consists entirely of natural changes.

## 7 Conclusion

This paper has attempted to explain an unusual comparison by hypothesizing that it was a shared innovation between the Rejang dialect group of Sumatra, Indonesia, and the Bukar-Sadong dialect group of Sarawak, Malaysian Borneo. The comparison involves a change neutralizing PMP *a in word-final syllables 'except before velars'. In our current state of knowledge, only these two languages show evidence of this comparison, which we have attributed to a common ancestor consisting of just these two languages, named Proto Sadong-Rejang, which was a daughter of Proto Bidayŭh-Rejang. (See (1).)

The principal arguments of this paper are of two types—both concerned with the problem of weighing evidence in comparative linguistics. The first argument concluded that one piece of evidence, namely *l > r in Bidayŭh languages but not Rejang, is virtually weightless on phonetic grounds, i.e. because it is far too 'natural', unpredictable, and ubiquitous to provide useful subgrouping information; hence this evidence has been ignored. The second argument was just the opposite, contending that another piece of evidence should be weighted heavily, namely, PMP *a Raising (neutralisation) in word-final syllables except before velars. What is odd about this change is the conditioning, which (we contend) cannot be explained on phonetic grounds. Moreover, *a Raising occurred in a 'real' phonological system being buffeted by a pair of competing sound changes: the spread of *a Raising among unstressed vowels, and Stress Shift, which caused unstressed vowels to become stressed (and vice versa). The competition from Stress Shift blocking the spread of *a Raising resulted in the odd conditioning ('except before velars') of *a Raising, traces of which are very much in evidence in contemporary Rejang and Bukar-Sadong dialects, and in no other languages, as far as is known. (See (13).)

Two precedents in the literature lend some theoretical support for the form of our argument. First, Adelaar's (1992) reconstruction of Proto Malayic demonstrates that a valid subgrouping hypothesis may be supported solely on the basis of common sound changes if there is a sufficient variety of them, in effect interpreting a rich enough *array* of changes as typologically unusual and therefore significant for subgrouping purposes (Adelaar lists eleven such changes as the basis of reconstructing Proto Malayic. In fact, as shown in (13), Proto Rejang and Proto Bukar-Sadong share an impressively rich array of common changes. At the other extreme, Blust (2006 and earlier work) presents a subgrouping hypothesis for Proto North Sarawak based almost exclusively on evidence of

languages exhibiting reflexes and drift-theoretical effects attributed to a series of rare voiced aspirates.[11]

Our approach shares features with each of these two precedents. Like the evidence for Proto North Sarawak, the evidence for Proto Sadong-Rejang is weighted heavily in favor of a single odd comparison. And like the evidence for Proto Malayic, Proto Sadong-Rejang is also supported by an interesting array of sound changes that are not at all unusual, especially in Borneo. Hypothesis (1) proposes an explanation.

## Appendix: The reconstructed phonemes of Proto Rejang and Proto Bukar-Sadong

Proto Bukar-Sadong phonemes are based on the data presented in Appendix B of this paper. Proto Rejang inventories are from McGinn (2005).

**Table 1:** Proto Rejang phonemes

| PR Consonants (23) | | | | | | PR Vowels (7) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Stops and Affricates | *p | *t | *c | *k | *ʔ | High | *i | | *u |
| | *b | *d | *z | *g | *j [gʸ] | | | | |
| Fricative | | *s | | | | Mid | *e | *ə | *o |
| Plain Nasals | *m | *n | *ñ | *ŋ | | | | | |
| 'Barred' Nasals | *m̄ | *n̄ | *ñ̄ | *ŋ̄ | | Low | ä | | *a |
| Liquids | *l | | *r | | | | | | |
| Semivowels | *w | | *y | | | Diphthongs (2) | *iw | *uy | |

PR *ʔ was glottal stop; PR *r was presumably a velar or uvular liquid (reflected as *h* or *ʔ* or zero in contemporary dialects); PR *ä was low, front and unrounded (reflected as /ä/ in Rawas); and the series /*m̄, *n̄, *ñ̄, *ŋ̄/ represents the 'barred nasals' (Coady and McGinn 1983). They are regular reflexes of PMP consonant sequences *-mb-, *-nd-, *-nz- and *-ŋg-, respectively.

See McGinn (2005) for extensive discussion of the evidence for Proto Rejang based on data from five contemporary dialects.

[11] According to Blust (2000:285), 'The 15– 20 languages of northern Sarawak form a linguistic subgroup ... defined primarily by a single sound change that left typologically unusual traces in the phonology of its members, including a set of true phonemic voiced aspirates (not murmured stops) *bh, dh, gh* in Bario Kelabit, corresponding to implosive stops in Bintulu and various Lowland Kenyah dialects, and a synchronic alternation of *b* and *s* in Kiput, reflecting *bh.' As noted by Kroeger (SMJ 1998:145), '(Blust) argues that even though no other significant phonological changes have been found, the Vowel Deletion rule is so well-attested and so unlikely to have spread by borrowing that it must be regarded as outweighing all other kinds of evidence, e.g. lexical isoglosses (Blust 1974a:220).'



**Table 2:** Proto Bukar-Sadong phonemes

| PBS Consonants (23) | | | | | | PBS Vowels (7) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Stops and Affricates | *p | *t | *c | *k | *ʔ | High | *i | | *u |
| | *b | *d | *j | *g | | | | *ə | |
| Fricative | | *s | | | *h | Mid | *e | | *o |
| Plain Nasals | *m | *n | *ñ | *ŋ | | | | *ʌ | |
| 'Barred' Nasals | *m̄ | *n̄ | *ñ̄ | *ŋ̄ | | Low | | | *a |
| Liquids | (*l) | *r | | | | | | | |
| Semivowels | *w | | *y | | | Diphthongs (3) | *ʌy | *w | *uy |

Symbols have the usual phonetic values except the 'barred' nasals (for which see Scott 1964 and Court 1967a, b and 1970).

## References


Adelaar, K. Alexander. 1992. *Proto Malayic: a reconstruction of its phonology and part of its morphology and lexicon*. Canberra: Pacific Linguistics.

—— 2007. Review of John Lynch, ed. *Issues in Austronesian historical phonology*. Canberra: Pacific Linguistics. *Bijdragen tot de Taal-, Land- en Volkenkunde* 163/1:139–146.

Asmah Haji Omar. 1983. *The Malay peoples of Malaysia and their languages*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

—— 1992. An overview of linguistic research on Sarawak. In Peter W. Martin, ed. *Shifting patterns of language use in Borneo*. BRC Proceedings Series vol. 3. Williamsburg, VA: The Borneo Research Council.

Bach, Emmon and Robert Harms. 1972. How do languages get crazy rules? In R. Stockwell and R. Macaulay, eds *Linguistic change and generative theory*, 1–21. Bloomington: Indiana University Press.

Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1929. A note on sound change. *Language* 4:99–100.

Blust, Robert A. 1984. On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140.4:422–450.

—— 2000. Low-vowel fronting in Northern Sarawak. *Oceanic Linguistics* 39.2:285–319.

—— 2006. The origin of the Kelabit voiced aspirates: a historical hypothesis revisited. *Oceanic Linguistics* 45.2:311–338.

—— (MS, n.d.) *Austronesian comparative dictionary*. Electronic manuscript.

Coady, James and Richard McGinn. 1983. On the so-called implosive nasals of Rejang. In Raine Carle et. al., eds *GAVA'. Studies in Austronesian languages and cultures dedicated to Hans Kähler*. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg Band 17; 437–449. Berlin: Dietrich Reimer Verlag.

Court, Christopher. 1967a. Some areal features of Mĕntu Land Dayak. *Oceanic Linguistics* 6:46–50.

1967b. A distinctive feature analysis of the phonemes of Měntu Land Dayak. *Phonetica* 17:202–207.

1970. Nasal harmony and some Indonesian sound laws. In *Pacific Linguistic Studies in honour of Arthur Capell*, 203–217. Canberra: Pacific Linguistics.

Dahl, Otto Christian, 1951, *Malgache et Maanyan. Une comparaaison linguistique.* Olso: Egede Instituttet.

Hudson, A.B. 1970. A note on Selako: Malayic Dayak and Land Dayak languages in Western Borneo. *Sarawak Museum Journal* XVIII (36–37):301–318.

—— 1978. Linguistic relations among Bornean peoples with special reference to Sarawak: an interim report. In *Sarawak: linguistics and development problems*. Williamsburg, VA: Studies in Third World Societies No.3, 1–45.

Kroeger, Paul R. 1994. MS. The dialects of Biatah. Borneo Research Council, 3rd Biennial Meeting, 10–14 July, 1994.

—— 1998. Language classification in Sarawak: a status report. *The Sarawak Museum Journal* 74:137–173.

McGinn, Richard. 1997, Some irregular reflexes of PMP vowels in Rejang. *Diachronica* XIV.1:67–106.

—— 2000. Where did the Rejangs come from? In Marlys Macken, ed. *Proceedings of the Tenth Annual Conference of the Southeast Asia Linguistics Society*, 247–262. Tempe: University of Arizona.

—— 2003. Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang. In John Lynch, ed. *Issues in Austronesian historical phonology*, 37–64. Canberra: Pacific Linguistics.

—— 2005. What the Rawas dialect reveals about the linguistic history of Rejang. *Oceanic Linguistics* 44.1:12–64.

Ray, Sidney H. 1913. The languages of Borneo. *The Sarawak Museum Journal* 1(4):1–196.

Scott, N.C., 1964, Nasal consonants in Land Dayak (Bukar-Sadong). In David Abercrombie et. al., eds *In honour of Daniel Jones*. London: Longmans, 432–436.

Tadmor, Uri. 2003. Final /a/ mutation: a borrowed feature in Western Austronesia. In John Lynch, ed. *Issues in Austronesian historical phonology*, 15–36. Canberra: Pacific Linguistics.

Topping, Donald M. 1990. A dialect survey of the Land Dayaks of Sarawak. In James T. Collins, ed. *Language and oral traditions in Borneo*, 247–274. Borneo Research Council Proceedings Series vol. 2. Williamsburg, VA: The Borneo Research Council.

Voorhoeve, P. 1955. *Critical survey of studies on the languages of Sumatra*. Martinus Nijhof: 's Gravenhage.

Zorc, David. 2006. Review of John Lynch, ed. *Issues in Austronesian historical phonology*. Canberra: Pacific Linguistics. In *Oceanic Linguistics* 45.2:505–516.

# 25 *The position of Makuva among the Austronesian languages of East Timor and Southwest Maluku*

AONE VAN ENGELENHOVEN

*For Bob Blust. May this be a small shoot that can be grafted somewhere on the Austronesian Tree that you have been pruning and cultivating so well.*

## 1 Introduction[1]

Makuva is spoken in the easternmost sub-district of Tutuala in the Republic of East Timor. Speakers are concentrated in the villages of Loiquero and Porlamano, which together make up the administrative centre of Mehara municipality. Makuva is known in the literature under several names: Loikera, Lóvaia or Lóvaia Epulu and Maku'a.[2] In this paper we will refer to this language as Makuva, a term which was introduced by Hull and Branco (2002/03).

The main language in Tutuala district is a form of Fataluku, which is non-Austronesian. A third language in this region, Rusenu, used to be spoken directly on and around Ilikerekere. It was generally considered to be extinct, but two remaining elderly semi-speakers were found in January 2007. Research is being conducted to determine whether this isolect is a language of its own or a dialect of either Fataluku or Makuva. Recent literature generally acknowledges Hajek's (1995) claim that Makuva is Austronesian,[3] although there is a

1 The research underlying the present paper was done within the framework of the Fataluku Language Project (2005–08) funded by the Netherlands Organisation for Scientific Research and initially with a pilot grant from the Hans Rausing Endangered Languages Programme (2003–04). I would like to thank Juliette Huber, Justino Valentim Cailoru and the editors of this volume for their input and help.

2 See Engelenhoven and Valentim (2006).

3 For a discussion on the older perception of Makuva being non-Austronesian, refer to Engelenhoven and Valentim (2006).

Alexander Adelaar and Andrew Pawley, eds
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 425–442.
Pacific Linguistics, 2009.

# 26 *Words of Eastern Polynesia: is there lexical evidence for the origin of the East Polynesians?*

PAUL GERAGHTY

## 1 Eastern Polynesian[1]

The existence of an Eastern Polynesian subgroup comprising all Polynesian languages from Hawai'i in the north to Rapanui (Easter Island) in the east and Aotearoa (New Zealand) in the south, and excluding all Polynesian languages from Pukapuka westward, has long been accepted. Hale (1846:117) was probably the first to moot such a subgroup, noting a number of features exclusive to the Eastern Polynesian languages that he studied, including the desiderative and reciprocal forms of the verb, the passive voice, and the plural of the possessive and demonstrative pronouns.[2]

Elbert (1953:154) and Haudricourt (1964:389) noted possible phonological innovations of Eastern Polynesian, then Pawley (1966:59–61) used innovations in grammatical morphemes (tense and other verbal markers, demonstratives, interrogatives etc) to lay the foundation for a more rigorous definition of the subgroup, while Green (1966:12–15) added lexicostatistical evidence. More recently, Marck (2000:131–132) has summarised the more compelling morphological innovations. Although Eastern Polynesian is a well-defined and generally accepted subgroup, there is a problem in reconstructing the lexicon, since one of its two first-order subgoups consists of only one language, Rapanui, spoken in a relatively impoverished natural environment and for which limited data is available. For

[1] It gives me great pleasure to dedicate this paper to my longtime friend and colleague Robert Blust. Since we first met at the University of Hawai'i in the 1970s, I have found him an unfailing source of support, unstintingly generous with information and advice, and a model of dedication. Many thanks, Bob, and may you long continue to flourish and reconstruct! Many thanks also to Andrew Pawley, who suggested the topic, and gave much help and advice during the writing of this paper; and to Pila Wilson and Erik Pearthree for useful discussions.

[2] Hale (1846:118, 175) may also be said to have anticipated the modern subgrouping of Polynesian languages into Tongic and Nuclear Polynesian, in that he noted that Tongan (he was unfamiliar with Niuean) 'differs strikingly, in several points, from the others, especially in the article, the pronouns, and the passive voice of the verb.'

Alexander Adelaar and Andrew Pawley, eds
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 445–460.
Pacific Linguistics, 2009.

want of a Rapanui reflex many items can only be reconstructed to Proto Central Eastern Polynesian (PCEP), the immediate ancestor of the other first-order group.

It has long been argued—or assumed—that East Polynesia was settled from Samoa (Hale 1846:119–125, 148), or somewhere in the region of Samoa and Tonga (Kirch 2000:231, 245). Apart from geographical proximity, a reason frequently cited has been that Savai'i, the largest island of Samoa, is believed to be the source of the place-name Hawaiki, commonly referred to in oral traditions as the Eastern Polynesian homeland.

An unexpected challenge to this assumption was made by Wilson (1985), who pointed out a number of shared innovations in the pronouns of East Polynesian and those of the North Central Outliers (Nukuria, Takuu, Nukumanu and Luangiua—situated north of Papua New Guinea and the Solomon Islands), and to a lesser extent Kapingamarangi and Nukuoro, in contrast to the pronominal system of Samoan, which shares no innovations exclusively with that of Eastern Polynesian.[3] Wilson suggested (1985:122–123) as a possible explanation that Eastern Polynesia was settled directly from the North Central Outliers, despite the distance of some 4000 miles, pointing out that in historic times it has been the inhabitants of small islands of scant resources, the Carolineans, the Tuamotuans, and the Tongans, who have been the most prone to sailing long distances.

Some fifteen years later, Marck (2000:xix) claimed to have found that 'Ellicean [North Central] outliers shared sporadic sound changes with Eastern Polynesian and Samoan that other Polynesian languages did not share …, a stunning bit of support for Wilson's (1985) suggestion of "Ellicean", composed of those same languages, on the basis of the pronoun prehistory.' However, the three sporadic sound changes referred to are, in my view, less than stunning, and in any case point to three distinct affiliations for Eastern Polynesian, only one of which matches that proposed by Wilson. In the first case, *fu?aŋa 'whetstone' > *fo?aŋa, the change is shared with Kap, MFa, Nkr, and Sam.[4] In the second, *kiu 'plover, wading bird' > *kiwi, the change is shared with Oja, Nkr, Sik and Tak—that is, the North Central outliers, as proposed by Wilson. In the third, *mafo 'healed' > *mafu, the change is shared with Ren, Tik, WUv, Nkr, Sam and Tok. In particular, the inclusion of Samoan in two of these three proposed innovations is at odds with Wilson's proposal, which specifically excludes Samoan.

[3] I would query just one of Wilson's proposed shared innovations: 6. replacement of **ki-* initiated pronouns, e.g. **ki-taaua* 'we' inclusive dual, with forms in which **ki-* has been deleted, e.g. *taaua*. My objection is that both types of independent pronoun form can be reconstructed for PPn, given that unprefixed forms like *taaua* are also found in both Western and Eastern Fijian, e.g. Yasawa (WF) *tatou* 'first person inclusive paucal/plural' (Triffitt 2000:320), Dogotuki (EF) *mutou* 'second person paucal' (Geraghty 1983:199).

[4] Abbreviations and default sources: All Polynesian languages Biggs and Clark (n.d.) unless otherwise specified. All Fijian, PCP (Proto Central Pacific) and PEO (Proto Eastern Oceanic) from my own notes; EF unspecified communalect of Eastern Fijian, EFu East Futuna (Moyse-Faurie 1993), EP Eastern Polynesian, EUv East Uvea (Rensch 1984), Kap Kapingamarangi, Mae, Mao New Zealand Maori, MFa Mele-Fila (Clark 1998), Mqa Marquesan, Mia Mangaia, Mki Manahiki, Mva Mangareva, Niu Niue (Sperlich 1997), Nkr Nukuria, Nuk Nukuoro (Carroll and Soulik 1973), Oja Ontong Java (Luangiua), PCE Proto Central Eastern Polynesian, Pen Penrhyn, Pohnpei (Rehg and Sohl 1979), PEP Proto Eastern Polynesian, PNP Proto Nuclear Polynesian, PPn Proto Polynesian, PTa Proto Tahitic, Puk Pukapuka, Rar Rarotongan (Buse and Taringa 1995), Ren Rennell (Elbert 1975), Rot Rotuman (Inia et al. 1998), Rpn Rapanui (Salles and Pizarro 2001), Sam Samoan (Milner 1966), Sik Sikaiana, Tah Tahitian (Lemaître 1986), Tak Takū, Tik Tikopia (Firth 1985), Tok Tokelau, Ton Tongan (Churchward 1959), Tua Tuamotu (Stimson 1964), Tuv Tuvalu (Besnier 1981), WF unspecified communalect of Western Fijian, WFu West Futuna, WUv West Uvea (Hollyman 1987).



Most recently, Wilson (2008) has pointed to some lexical innovations that indicate a period of shared development between Eastern Polynesian and the Northern Outliers, including the following: PPn *kawiki 'ghost crab' > *kawiti (shared by EP with Kap, Tak, Sik); *watuke 'pencil urchin' > *fatuke (Tak); *fasi 'split' > *wasi (Tak, Oja, Sik); *ta?e 'faeces' > *tūta?e (Kap, Nuk, Tak); *fanaŋa 'story' > *wanaŋa (Kap); *?ulupoko 'skull' > 'head' (Nuk); *la?e 'forehead' > 'headland' (Oja).

## 2 Words and things

The purpose of this paper is to apply the *Wőrter und Sachen* (henceforth 'words and things') method of historical reconstruction to Proto Eastern Polynesian, to determine the geographical origin of the ancestral Eastern Polynesians, thus reinforcing or challenging Wilson's hypothesis of a Northern Outlier source. I will focus on plant-names, looking for evidence as to whether the ancestral Eastern Polynesians lived on a high volcanic island, such as Samoa, in which case they may have tended to retain names of high island plants, or a low coral island or atoll, such as Pukapuka, Tokelau, Tuvalu, or any of the northern outliers, in which case they may have become unfamiliar with high island plants, and have had to reinvent names for them on arrival in the high islands of Eastern Polynesia.

The 'words and things' technique of historical reconstruction is based on the assumption that if a word is reconstructable for a protolanguage, then the speakers of that language must have been familiar with its referent.

In what was probably the first systematic application of this technique in the Pacific, Pawley and Green (1971) studied relevant vocabulary reconstructable to Proto Polynesian and concluded that the speakers of Proto Polynesian 'occupied or lived near an environment where, for example, mountains, cliffs, rivers, lakes, landslides and, probably, volcanic rock were found. That is, the community lived on or near a high island or large land mass, rather than a remote atoll.' They added: 'the presence in PPN of many terms for plants characteristic of the Indo-Pacific tropical zone indicates that the location lay within this zone' (Pawley and Green 1971:17). More specifically, 'the PPN speech community were fishermen-horticulturalists, familiar with a typical tropical Indo-Pacific high island environment and also with certain objects found natively only on certain islands of this category, including the *balolo* worm, the pearl oyster, such land animals as snakes, pestiferous mosquitoes, bats, owls, rails, pigeons, parrots. [It is] highly unlikely that the homeland lay anywhere in East Polynesia, or in marginal regions of West Polynesia' (Pawley and Green 1971:23).

In the same paper, Pawley and Green proposed a number of postulates, i.e., ways of drawing further historical inferences from the linguistic data. They suggested, for instance, that the corollory of the first tenet of 'words and things' is that if a protolanguage had no name for a thing, then it was probably absent from their environment. They cited (following Biggs) the example of 'seal', a concept for which there is no Proto Polynesian reconstruction, although some individual Polynesian languages do have words for it. This suggests that seals were not present in the Polynesian homeland. Again following Biggs, they noted that reflexes of PPn *namu 'mosquito' in some Eastern Polynesian languages do not mean 'mosquito' (e.g. 'biting midge' in Maori and Marquesan), and inferred that mosquitoes did not exist in these places at the time of settlement (Pawley and Green 1971:19).

More recently, there have been some striking successes of the 'words and things' technique in and around Polynesia. One concerns the word for 'megapode', a flightless bird

which buries its eggs in the sand to hatch, hence also known as the 'incubator bird'. Clark (1982:126) noted that the name for the megapode in Tonga, *malau*, is related to the names for similar birds in Vanuatu, Solomon Islands, and New Guinea. Clark argued that since we must reconstruct *malau as the word for 'megapode' in Proto Oceanic, and there must have been an unbroken transmission of this word from Proto Oceanic to Proto Polynesian for it to appear in Tongan, then it must also have been part of the lexicon of Proto Central Pacific, which is believed to have been spoken in Fiji. So 'words and things' requires that during the Lapita period, when Proto Central Pacific was spoken, megapodes must have been present in Fiji, even though there are currently none, nor have there been in recorded history. Soon after Clark's observation, the archaeologist Simon Best unearthed the remains of at least two different species of megapode in Fiji, both of which became extinct soon after initial human settlement (Clunie 1984:140)—a dramatic vindication of 'words and things'.

The 'words and things' technique has, of course, limitations. While the logic of 'if they have a word for it they must know it, so it must be there' is unassailable, there is always the possibility that the reconstruction is in some way flawed, and it is possible to reconstruct apparently ancient words that are not ancient at all. Geraghty (2004a:65–66) notes that a word for 'motor-car' can be reconstructed for Proto Micronesian (it is actually a loan from Japanese), and similarly *sāmala 'hammer' can be reconstructed for Proto Polynesian—both cases are relatively recent loanwords that have been 'etymologically borrowed' (Geraghty 2004a:77–78) among related languages. Similarly, with regard to the parent language of the family to which Pacific languages belong, Proto Austronesian, Mahdi (1994) has shown that while words for 'iron', 'gold', 'silver' and some other metals and useful plants can be reconstructed, it is highly unlikely that the speakers of Proto Austronesian had any knowledge of them—they were all introduced well after the break-up of Proto Austronesian. Another potential source of confusion is that related words can undergo parallel semantic extension: Crowley (1994:87) has pointed out that *tusi 'book' can be reconstructed for Proto Polynesian (it originally meant 'mark or adorn with colour'), to which can be added *faʔo 'nail' (originally a wooden peg used for fastening). It is important, then, that this method be applied with caution.

Bearing in mind these provisos, we now turn to plant-names, and other words relevant to the environment that can be reconstructed for Proto Eastern Polynesian. Not all are of interest. For example, we can reconstruct PEP *futu 'k large coastal tree, *Barringtonia asiatica*', but this is not particularly useful, since it does not distinguish between volcanic islands and atolls as the homeland of the Eastern Polynesians, this particular tree being found in profusion in both ecosystems. What do concern us are words in the PEP lexicon that refer to high island entities and are *not* continuations of PNP vocabulary, which would suggest that the referents had not been part of the environment of pre-PEP speakers, since they lived in an atoll environment. Conversely, if we find that names of plants that are confined to high islands and absent from atolls continue from PPn and PNP into PEP, then that would suggest that the speakers of PEP came from a high island environment.

## 3 Regular changes and tendencies

Curiously for such a well-defined subgroup, Eastern Polynesian shows no regular phonological innovations, other than the purely phonetic, and in any case debatable, *l > *r (Marck 2000:23–25). Elbert (1953:154) pointed out the tendency for EP languages, and also some outliers, to merge *f and *s as *h, but this strictly speaking does not constitute a unique shared innovation of PEP; and Haudricourt's (1964:389) observation that *faf- became *vah- holds only for Proto Central Eastern Polynesian, in other words, is not valid for Rapanui.[5]



The plant-name data presented below do, however, when combined with data from other domains, suggest another phonological tendency of Proto Eastern Polynesian and its close relatives and daughter languages: for pretonic high vowels to become a low (or mid) vowel. I do not intend to explore this in detail in this paper, but the following are suggestive.[6]

*ŋiŋie 'k. coastal shrub, *Pemphis acidula*' > Puk, Tok, PEP *ŋaŋie
*(hs)ulufe 'k. fern, *Dicranopteris linearis*' > PTa *aruhe
*muti(a,e) 'grass' > PTa *matie (Pollex *mutie; *mutia is indicated by Sam, Tok *mutia* and the external evidence of EF *vūtia, mūtia* 'sea-grass')
*tufuŋa 'expert, priest' > PTa *tahuŋa, Haw *kahuna*,[7] Tua *tohūŋa* 'priest', Mao *tohuŋa*

Another possible tendency is for a final mid back vowel to be lowered:

*hoŋohoŋo 'k. nettle, *Laportea interrupta*' > PCE *oŋaoŋa (Mqa *okaoka, onaona*, Tua *oŋaoŋa*, Mao *oŋaoŋa*). (Pollex PPn *hoŋahoŋa 'nettle or other stinging plant' is incorrect, since EF *soŋa* 'sago palm' is probably not cognate, whereas Rot *usoŋo* '*Laportea interrupta*' is)
*(ka) ŋaoso 'k. shrub, *Caesalpinia*' > Mqa *kaeha, keoha* (but Rpn *ŋaoho*)

Data presented below also bear witness to a tendency for PPn words (usually, or maybe exclusively, nouns) to acquire in PEP, and its close relatives and daughter languages, a prefix consisting of a stop and a vowel, *kō- being the most common.[8] In the following comparisons the first reconstructions are all PPn.

*fai 'k large tree, leguminid' > PCE *kofai 'pod-bearing plant, *Sesbania*' (Tah *ʔofai*, Haw *ʔohai*, Tua *kohai*)[9]
*fatu 'stone' > PCE *pōfatu, *kōfatu (Mia, Maò)
*felo '*Ficus tinctoria*, k. banyan with yellow-red berries' > Haw *ʔōhelo* 'k. shrub, *Vaccinium* spp., bears yellow or red berries' (cf. PPn *felo 'yellowish, reddish')
*(fua)fua 'young mullet' > Haw *ʔōhua* 'young of certain fish'
*fue 'k shore creeping vine' > PCE *pōfue (also Puk)
*kili 'saw, file' > PEP *kōkili 'triggerfish' (also Puk)
*kisi 'k. grass, *Oxalis*' > Rar *kōkiʔi*, Mao *kōkihi* '*Tetragonia*'
*polo '*Solanum nigrum*' > PCE *kōporo
*reŋa 'turmeric' > Haw *ʻōlena* '*Curcuma*'
*taʔe 'faeces' > PEP *tūtaʔe (also Kap, Nuk, Tak)
*ʔura 'crayfish' > PEP *kōʔura

5 Nor is it valid for an apparently older stratum of Mangareva vocabulary (Fischer 2001).
6 Three of these were noted as sporadic sound changes of Proto Tahitic by Marck (2000:134).
7 Marck (2000:134) claims that this is a loan from Tahitian, but offers no evidence, so appears to be begging the question.
8 Biggs (1994:22) notes that New Zealand's first settlers made extensive use of prefixation of *poo-* and *koo-*, which seem to have had the meaning 'pseudo-' or 'like'.
9 Ton, EUv *ʔōhai* '*Delonix*' are presumably nineteenth century loans from Hawai'ian.

Finally, a small number of forms suggest a tendency for reduplication in PPn to become deleted or reduced in PEP and/or PCE:

*kakamika '*Sigesbeckia, Ageratum*' > *kamika '*Sigesbeckia*'
*kisikisi '*Oxalis*' > *kisi
*palapala 'k. tree-fern, *Cyathea*' > *pala 'k. fern, *Marattia*'
*talatala?amoa '*Caesalpinia*' > *tātara?amoa

while two show the reverse:

*kaso 'reed, *Miscanthus* sp' > *kākaso
*kawa '*Pittosporum*' (WF *kawa*) > *ka(wa)kawa (Haw *'a?awa*, Rar *kavakava*)

## 4 Results of survey

I collected and compared plant-names for most Polynesian languages, using the standard dictionaries in most cases, and compared them to plant-names of Rotuma and Fiji and some other Oceanic languages. In addition I referred to Pollex (Biggs and Clark n.d.), Göthesson (1997), and Rensch (2005), all very useful sources of information on Polynesian plants and their names. The only major language I did not study in detail was New Zealand Maori, which I judged to be less useful because of its non-tropical location, and in any case is well covered in Pollex and works such as Biggs (1991, 1994). Doubtless a detailed study of New Zealand Maori plant names, and older dictionaries and word-lists of Polynesian languages, would yield more results and cognate sets, but probably not affect the major conclusions of this study.

A large number of the plants of Western Polynesia are simply not found in Eastern Polynesia, so are irrelevant to our discussion. Nevertheless, I list below those that have a PPn reconstruction. For those reconstructions which are not found in Pollex, or differ in some way from the Pollex entry, I add some supporting data.

**Table 1:** Coastal plants not found in Eastern Polynesia

| | | |
|---|---|---|
| *lekileki | *'Xylocarpus granatum'* | |
| *saŋale | *'Lumnitzera littorea'* | (EF *saŋale*; Ton *haŋale* 'k. tree which like the mangrove grows in the sea', Tuv *saŋale* 'k. tree', Ren *saŋaqe*) |
| *sinu | *'Excoecaria agallocha'* | (EF *sinu*; MFa *sinu*, WUv *sinu*) |
| *tātāŋia | *'Acacia simplicifolia'* | (EF *tātāŋia*; Sam *tātāŋia*) |
| *toŋo | 'mangrove, *Bruguiera* and *Rhizophora*' | |

**Table 2:** Non-coastal plants not found in Eastern Polynesia

| | | |
|---|---|---|
| *aka | *'Pueraria lobata'* | |
| *alu/walu | *'Epipremnum pinnatum'* | (EF *yalu*, *walu*; Ton *alu*, Tik *valu*; Rot *rauvaru* loaned from Pn *rau-walu) |
| *asi | '*Syzigium* sp., not cultivated, not fragrant or edible' | (WF *yasi* 'S curvistylum', EF *yasi*; EFu *asi* '*S clusiifolium*', Sam *asi* '*S inophillodes*') |



| | | |
|---|---|---|
| *ate | *'Wedelia'* | (also *kofekofe) |
| *filimoto | *'Flacourtia rukam'* | |
| *fiso | *'Saccharum edulis'* | |
| *fukafuka | *'Kleinhovia hospita'* | (Ton *fukofuka*, Sam *fu?afu?a*, Tik *fukafuka*) |
| *ŋī | *'Imperata'* | (EF *ŋī*; WUv *ŋī*, WFu *ŋi* 'k. tall thin-bladed grass or sedge', Mae *ŋī* 'k. plant') |
| *kala?apusi | *'Acalypha grandis'* [10] | |
| *kanume | *'Diospyros elleptica, D ferrea'* | |
| *kofekofe | *'Wedelia'* | (EF *kovekove*; Tik *kofekofe*; also *ate, possible convergent development from PCP *kove 'bamboo') |
| *laŋakali | *'Aglaia saltatorum'* | |
| *loŋoloŋo | *'Cycas rumphii'* | |
| *manaui | *'Garuga* sp.' [11] | |
| *mapa | *'Diospyros* sp.' | |
| *moli | *'Citrus* spp.' | |
| *nukanuka | *'Decaspermum vitiense'* | |
| *pau | 'k. large hardwood tree, *Palaquium, Planchonella'* | |
| *pele | *'Abelmoschus manihot'* (probably a recent introduction from Fiji to Polynesia, see Geraghty 2004a:85) | |
| *poumuli | *'Flueggea flexuosa'* | (EF *baumuri*; Ton, Sam *poumuli*, Tik *poumuri*) |
| *salato | *'Laportea/Dendrocnide harveyi'* | |
| *sea | *'Parinarium insularum'* | |
| *talafalu | *'Micromelum'* | (WF *tavolali*; Sam *talafalu*; Rot *tarfai* Pn loan) |
| *tamanu | *'Calophyllum vitiense, C samoense* and other inland species' | |
| *tanetane | *'Polyscias multijuga'* | |
| *taputoki | *'Alectryon grandifolius'* | |
| *tawa | *'Pometia pinnata'* | |
| *tawahi | *'Rhus taitensis'* | |
| *usi | *'Evodia* sp.' | |

[10] Marck (2000:64) cites Ton *kalakala?apusi* as an example of Pn *s 'which has yet to change into Tongan *h*'; I believe a more likely explanation is that a former *kalakala?apuhi has been quite recently changed to *kalakala?apusi* by analogy with Ton *pusi* 'cat', the tail of which the flower of this plant resembles.

[11] Note that this appears to be a direct reflex of PCP *?ui (<PEO *?uRi *Spondias dulcis*), with the mala-/mana- prefix meaning 'like, false', though Tongan should be *mana?ui. The widespread Polynesian *wī *Spondias dulcis* is clearly a loan from Fijian (Geraghty 2004a:87).

A similar list compiled by Biggs (1994:23) includes a number of taxa that are not listed here. In some cases it is because they are names that were replaced in Eastern Polynesia (e.g. *aŋo '*Curcuma*' replaced by *reŋa); in others, I believe the plants and/or their names to be relatively recent introductions, so not reconstructible to PPn (*fesi '*Intsia bijuga*', *mosokoi '*Cananga odorata*', *tono '*Centella asiatica*', *wālai 'a liana', see Geraghty 2004a); and some (e.g. *makari 'k. tree') are not sufficiently defined to determine whether or not they are also found in Eastern Polynesia.

Below are two lists of names of plants that are found in both Western and Eastern Polynesia, and have relatively secure reconstructions at both levels (or at least to Proto Central Eastern Polynesian). Note that neither list is claimed to be complete. The first is of coastal plants, i.e. those that can be found on atolls; the second is non-coastal plants, i.e. those that are not found on atolls.

**Table 3:** Atoll plants found in both West and East Polynesia

*aloalo/walowalo '*Premna taitensis*' > *warowaro (Tah, Mva)
*fano '*Guettardia speciosa*' > *fano (cf. PPn *puapua)
*fao '*Ochrosia*' > *fao (doubtful, only reflex being Haw *hao Rauvolfia*)
*fara '*Pandanus*' > *fara
*fatai '*Cassytha filiformis*' > *tainoka
*fau '*Hibiscus tiliaceus*' > *fau
*felo '*Ficus tinctoria*, k. banyan with yellow-red berries' > Haw *'ōhelo* 'k. shrub, *Vaccinium* spp., bears yellow or red berries' (cf. PPn *mati '*Ficus tinctoria*', *felo 'yellowish, reddish')
*feta?u '*Calophyllum inophyllum*' > *tamanu[12]
*fiso?a '*Colubrina asiatica*' > *tutu (cf. PPn *tutu)
*fue '*Convolvulus*, Ipomoea pes-caprae' > *pōfue (cf. PEP *fue '*Lagenaria vulgaris*')
*futu '*Barringtonia asiatica*' > futu
*ŋasu '*Scaevola*' > *ŋasu (Pen, Rar), *naupata (Tah, Haw)
*ŋiŋie '*Pemphis*' > *ŋaŋie (also Tok, Puk; but Tua *ŋieŋie*)
*(ka)ŋa?oso '*Caesalpinia*' (Sam *?anaoso*, *?āŋoso*) > *ŋa?oso (Mqa *keoha*, Rpn *ŋa?oho*) (cf. *talatala?amoa)
*kanawa, *fakanawa, *tou '*Cordia subcordata*' > *tou (probably from PPn *tou 'tapa paste', made from *Cordia* fruit)
*katafa '*Asplenium nidus*' > *katafa
*katuli 'coastal herb *Portulaca, Boerhavia*' > *katuri (Haw, Tah, Pen, Mki)
*kaute '*Hibiscus rosa-sinensis*' > *kaute
*kie 'k. *Pandanus* used for fine mats' > *kie 'sail'
*kofe 'bamboo' > *kofe
*kulu '*Artocarpus*' > *kuru (cf. *mei)
*lala '*Vitex*' > *rara (Rar)

[12] The replacement of PPn *feta?u '*Calophyllum inophyllum*' by PCE *tamanu (from PPn *tamanu 'inland sp of *Calophyllum*') is, to say the least, unexpected. By a strict 'words and things' interpretation, this single item suggests that Eastern Polynesia was first colonised by inland dwellers—perhaps from Samoa—who had lost knowledge of *feta?u, so called it by the name of the inland species they were familiar with. No other evidence I have come across points to this conclusion.



*mahuku 'grass' > *mauku
*maile '*Alyxia, Polypodium*' > *maire
*mati '*Ficus tinctoria*' > *mati (cf. *felo)
*ma?utofu '*Urena, Sida*' > *kulima *Sida* (Haw *?ilima*, Tua *kurima*)
*mei 'breadfruit' > *mei (Mqa, Mva) (cf. *kulu; probably introduced from Micronesia, see Geraghty 2004a:87–88)
*milo '*Thespesia populnea*' > *miro
*muti(a,e) 'grass' > *mutie
*niu '*Cocos nucifera*' > *niu
*nonu '*Morinda citrifolia*' > *nonu
*pia '*Tacca*' > *pia
*pipi '*Hernandia*' > *puka
*pua '*Fagraea berteriana*' > *pua (Tah, Rar)
*puapua '*Guettardia speciosa*' > *fano (cf. PPn *fano)
*puka '*Pisonia*' > *puka, *pukatea
*puka '*Hernandia*' > *puka
*rewa '*Cerbera*' > *rewa (*Pollex* *lewa in error, there being no Tongic reflex)
*tafifi '*Ipomoea littoralis*' (EF *sovivi*; Niu *tefifi* '*Ipomoea* sp.', Sam *lautafifi*) > *tafifi 'k. creeper' (Tua *tāhihi* Alyxia, Tah *tāfifi* Alyxia, Mqa *tāfifi*)
*talatala?amoa '*Caesalpinia*' > *tātara?amoa (cf. *(ka)ŋa?oso)
*tamole '*Portulaca*' > *katuri (cf. PCE *tamole '*Polygonum*' (Tah, Haw))
*tausun(i,u) '*Tournefortia argentea*' > *tausinu
*tiale '*Gardenia taitensis*' > *tiare
*toa '*Casuarina equisetifolia*' > *toa
*tutu '*Colubrina asiatica*' > *tutu (cf. *fiso?a)

**Table 4:** High island plants found in both West and East Polynesia

*aloŋā '*Pipturus aregenteus*' > *oroŋā
*aŋo '*Curcuma*' > *reŋa (from PPn *reŋa 'tumeric')
*asi 'sandalwood' > *asi
*falasola '*Pandanus* var' > *farasora (dubious, only reflex Mqa *fa?aho?a* 'pineapple')
*fai 'k. large tree, *Leguminid*' > PCE *kofai 'pod-bearing plant, *Sesbania*' (Tah *?ofai*, Haw *?ohai*, Tua *kohai*)
*fau?ui '*Grewia crenata*' > *faupā (Tah, Tua)
*fenua '*Macaranga harveyana*' (EF *venua*) > *fenua (Tah *fenia* '*Omalanthus nutans*' (same family, Euphorbiaceae), Rar *?enua*. Note also Niu *lēhau* '*Macaranga*'; and Mqa *ferua*, Haw *lehua*, both '*Metrosideros collina*'.
*fanuamamala '*Omalanthus*' > *fenua (Tah *fenia*)
*futi '*Musa*' > *maika
*ŋase '*Geniostoma*' (EUv) > *saŋe (Rar)
*ŋase 'k. fern' (EF *qase* '*Dendrobium*') > *ŋahe (Tua *ŋahe* 'k. giant fern')
*ŋatae '*Erythrina indica*' > *ŋatae
*hiti '*Grewia*' (EF *siti*; EFu *iti*) > *faupā (Tah, Tua)

*hoŋohoŋo 'k. nettle, *Laportea interrupta*' > *oŋaoŋa (Mqa *okaoka, onaona*, Tua *oŋaoŋa*, Mao *oŋaoŋa*)
*(hs)ulufe 'k. fern, *Dicranopteris linearis*' > PEP *urufe (Haw), PTa *aruhe
*ifi '*Inocarpus edulis*' > *ifi
*kafika '*Syzygium malaccense*' > *kafika
*kakamika '*Sigesbeckia, Ageratum*' > *kamika *Sigesbeckia*
*kalaka '*Planchonella*' > *karaka
*kape '*Alocasia*' > *kape
*kaso 'reed, *Miscanthus* sp.' > *kākaso
*kawa '*Pittosporum*' (WF *kawa*) > *ka(wa)kawa (Haw *ʔaʔawa*, Rar *kavakava*)
*kawa '*Piper methysticum*' > *kawa
*kawakawaʔatua '*Piper latifolium*' > *kawakawaʔatua
*kawasusu '*Tephrosia*' > *kaususu, *sora
*kiekie '*Freycinetia*' > *kiekie
*kisikisi '*Oxalis*' > *kisi
*koka '*Bischofia javanica*' > *koka (Rar)
*kuta 'k sedge, *Eleocharis*' > *kuta 'k reed' (Mia, Mao)
*laupata '*Macaranga*' > *naupata '*Scaevola*' (problematic on both phonological and semantic grounds)
*maŋele '*Trema*' > *maŋele 'k. tree' (Mqa '*Alphitonia*', Haw 'k. tree')
*mako '*Trichospermum, Melochia*' > *mako '*Melochia*' (Tah)
*manon(o,u) '*Tarenna sambucina*' > *manono
*masame '*Glochidion ramiflorum*' > *masame (Tah, Rar)
*maʔota '*Dysoxylum*' > *maʔota (Mao)
*palapala 'k. tree-fern, *Cyathea*' > *mamaku, *pala 'k. fern, *Marattia*', cf. *poŋa
*palai '*Dioscorea nummularia*' > *parai (Tah, Rar)
*pilita '*Dioscorea pentaphylla*' > *pirita (Tua, Rar)
*poŋa 'k tree-fern, *Cyathea*' > *poŋa (Rar, Mao), cf *palapala (perhaps from PPn *poŋa 'hole, orifice', with reference to hollow trunk)
*polo '*Solanum nigrum*' > *poro (Rar), *pōporo (Haw, Rpn), *poroporo (Haw, Tua), *kōporo (Tah, Mqa, Rap)
*siapo '*Broussonetia*' > *kaute (Tah), cf. *kaute '*Hibiscus rosa-sinensis*'
*soaka '*Musa fehi*' > *fekī, fuatū
*soi '*Dioscorea bulbifera*' > *soi
*tewe '*Amorphophallus campanulatus*' > *tewe
*tī '*Cordyline fruticosa*' > *tī
*toi '*Alphitonia*' > *toi (Tah, Rar)
*toto '*Euphorbia*' > *(ka)toto (Haw, Tah)
*tuitui '*Aleurites moluccana*' > *tuitui, *tutui (Haw, Tah)
*wawae '*Gossypium barbadense*' > *wawai
*ʔafatea '*Nauclea*' > *ʔafatea (Tah)
*ʔufi '*Dioscorea alata*' > *ʔufi



The above tables leave little doubt that many names of high island plants persisted from Proto Polynesian and Proto Nuclear Polynesian into Proto Eastern Polynesian and Proto Central Eastern Polynesian. Particularly striking are *kalaka '*Planchonella*', *kawa '*Piper methysticum*', and *koka '*Bischofia javanicus*', all relatively widespread in Eastern Polynesia, and none of which is found on any of the northern outliers that Wilson (1985) has proposed as the source for the earliest settlers of Eastern Polynesia.

A quick glance at other semantic fields suggests that results would be similar. For instance, the names of freshwater fish *hinaŋa 'whitebait' and *tuna 'freshwater eel' continue into Proto Eastern Polynesian, as do high island topographical features such as *kalā 'hard, black volcanic stone', *mato 'precipice, steep place, cliff', and *solo 'landslide'.

## 15 Discussion

Even though there are many etyma that indicate that PEP speakers had knowledge of exclusively high island plants and other features of the environment, this does not entirely rule out the possibility that EP was settled initially from an atoll or atolls. There are at least two scenarios, not necessarily mutually exclusive, which would allow the speakers of PEP to originate from an atoll environment, yet for words for high island phenomena to be reconstructible for PEP:

1. The atoll dwellers had knowledge of high island environments; or
2. An initial colonisation by atoll dwellers was followed by a colonisation from a high island or high islands, at such an early date as to be perceived as simultaneous linguistically, that is, before any sound change[13] or further significant population movement.

Regarding option one, that atoll dwellers, in this case in the northern outliers, at the time of the first settlement of East Polynesia, had detailed knowledge of high island environments, there is no way that we can be sure whether or not this was the case. Extrapolating from relatively recent times, it is true that Tuvaluans and Tokelauans apparently had no knowledge of Samoa around 1840 (Hale 1846:153, 5, 65), but they did know of at least some high island produce. As Hale (1846:166) noted in Vaitupu, Tuvalu: 'Yams and bananas they knew by name, but had none.' Tuamotuans in the late 18th and early 19th centuries often paid extended visits to Tahiti (Haddon and Hornell 1975:79)—though there is no indication of how well they knew the Tahitian environment.

Postulate four of Pawley and Green (1971:17) stated that 'the presence in a proto-language of a term denoting a category of objects is taken as indicating that the referents were familiar to the speakers of the language, either as part of their own immediate environment or as part of a nearby environment.' In Postulate five, Pawley and Green then defined 'nearby', taking their cue from Sharp's (1963) theory of accidental voyaging: 'deliberate two-way voyaging over distances exceeding two or three hundred miles across open sea, and using indigenous craft, is... unlikely to have occurred. We thus postulate a radius of three hundred miles around any point as the upper range of 'nearby environments'. Even by this very parsimonious estimate, inhabitants of the North Central Outliers would have had in their 'nearby environment' the high islands of the Solomons,

[13] Even after a sound change, words borrowed 'etymologically' would be linguistically invisible (Geraghty 2004a:77–78).

likewise Nukuoro, which is less than three hundred miles from Pohnpei, and Kapingamarangi, which is approximately three hundred miles from New Ireland.

Since those isolationist days ushered in by Sharp and others, the pendulum has swung the other way (Finney 1994:255–259) and there are few now who would deny that prehistoric Polynesians were adventurous long-distance voyagers, and that the 300 mile limit is an underestimate. 'They are a race of navigators, and often undertake long voyages in vessels in which our own sailors would hesitate to cross a harbour [..] not only is a constant communication kept up among the different islands of each group of Polynesia, but perilous voyages of many days between different groups are frequent' (Hale 1846:14). We know from the Ra'iatean Tupa'ia and other navigators encountered by Spanish and British explorers in the eighteenth century that Tahitians knew at least the names and approximate locations of all the islands of triangle Polynesia (except the extremities of Hawai'i , Mangareva, Rapanui and New Zealand) as well as Fiji and Rotuma (Hale 1846:124, Dening 1962:103, 135). Rotuma is about 4000 km (2400 miles) distant from Tahiti. Moreover Tupa'ia indicated that his father had even greater knowledge of the islands of the Pacific (Beaglehole 1968:157, Dening 1962:105), and we know from other sources that Polynesian navigation had been in decline since the 'little ice age' that began around 1350 AD (Nunn 2008). In Western Polynesia, Tongans told Cook of islands they knew as far as Kiribati and probably also the Solomons (Geraghty 2004b), and we can infer from linguistic and other evidence that Tongans, or other western Polynesians, travelled to and from places as far away as Vanuatu, Pohnpei (Geraghty 1994), and the Carolines. The Marquesans also have legends of voyages to and from Rarotonga to procure red feathers (Langridge and Terrell 1988:11–31). In sum, I would venture to suggest that when Polynesian voyaging was at its peak, the 'nearby environment' with which Polynesians were familiar could well have stretched to a thousand miles or even more.

The acceptance of such voyaging capabilities explains some apparent 'words and things' anomalies. For example, the Proto Central Pacific and Proto Polynesian reconstruction *lulu 'owl', has come to refer to a sea bird, usually the booby (*Sula* sp.), in Eastern Polynesian languages spoken where there are no owls (that is, all except Hawai'ian and New Zealand Maori). However, in New Zealand Maori, the referent is again the owl. While it is possible that the name for booby was transferred back to the owl, and even remotely possible that owls once existed in Central Eastern Polynesia, the most likely explanation is simply that the Eastern Polynesians who colonised New Zealand were familiar with owls, and their name, from voyaging to Western Polynesia.

Similarly, the Hawai'ian word for the tree *Myoporum*, *naio*, corresponds exactly to *ŋaio*, its name in the Austral Islands, Cook Islands, and New Zealand (it is only found in Eastern Polynesia). However the genus is absent not only in the Marquesas, whose languages subgroup with Hawai'ian within Eastern Polynesian, but also in the Society Islands, where part of the Hawai'ian lexicon is believed to have originated (Whistler 1995:51). The mystery of this 'words and things' conundrum dissipates when we acknowledge that the prehistoric Polynesians' world was far from confined to their own island group. It is hardly surprising that such well-travelled people should be familiar with useful plants—*Myoporum* was used as sandalwood and in house construction in Hawai'i (Degener 1973:267–268), while in Rarotonga the flowers are used to scent coconut oil—in neighbouring island groups.

The second scenario of Eastern Polynesian colonisation (not mutually exclusive with the first) that is consistent with Wilson's thesis is that an initial colonisation by atoll dwellers was followed by colonisation from a high island or high islands, at such an early



date as to be perceived of as simultaneous linguistically, that is, before any sound change or further significant population movement. In other words, that Eastern Polynesia was not colonised only once, but twice or a number of times, from different Western Polynesian sources, initially from atolls, but subsequently from high islands, and that Proto Eastern Polynesian could have been lexically enriched by later colonists from high islands of Western Polynesia.

As with the notion of limited voyaging ability, the 1960s notion that each Polynesian island or group was colonised only once has succumbed over the years to the weight of evidence from many fields (Finney 1994:263–270). As noted by Kirch (2000:244–245), recent work on language relationships, voyaging, and long-distance interaction spheres (to which could be added oral traditions) all suggest that 'rather than a single population movement into one island or archipelago of central Eastern Polynesia, which then served as a primary dispersal center … the process of expansion out of the Ancestral Polynesian homeland was more complex, involving at least three separate movements, each resulting in interaction spheres and dialect chains that persisted over significant time periods.' While I do not agree with Kirch's specific proposals, I believe that the idea of initial colonisation from northern outliers followed by a number of intrusions from elsewhere is substantially correct.[14]

## References

Beaglehole, J.C., ed. 1968. *The journals of Captain James Cook on his voyages of discovery.* vol.1. *The voyage of the Endeavour 1768–1771.* Published for the Hakluyt Society. Cambridge: Cambridge University Press.

Besnier, Niko. 1981. *Tuvaluan lexicon.* Funafuti: United States Peace Corps.

Biggs, Bruce. 1991. A linguist revisits the New Zealand bush. In Andrew Pawley, ed. *Man and a half: essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer*, 67–72. Auckland: The Polynesian Society.

—— 1994. New words for a new world. In Pawley and Ross, eds 1994:21–29.

Biggs, Bruce and Ross Clark. n.d. Pollex.

Buse, Jasper and Raututi Taringa. 1995. *Cook Islands Maori dictionary.* Cook Islands Ministry of Education et al.

Carroll, Vern and Tobias Soulik. 1973. *Nukuoro lexicon.* PALI Language Texts: Polynesia. Honolulu: University Press of Hawaii.

Churchward, C. Maxwell. 1959. Tongan dictionary. Oxford: Oxford University Press.

[14] The name of the island Tahiti suggests that it was settled from Samoa, which lies almost due west, being composed of *ta 'the' + *fiti (> *hiti*) 'east'. Pollex does not currently list *ta as an alternative form of the definite article *te, but the evidence for it in at least Nuclear Polynesian is compelling: WFu *ta*, Haw *ka*, Mqa *ta* (Hale 1846:134–135), and loans such as Rot *tarau* 'hundred' (PPn *rau) and Pohnpei *sakau* 'kava' (PPn *kawa).

A possible explanation for the shared innovations pointed out by Wilson (1985), which I will not explore in this paper, is that Eastern Polynesia was colonised from Samoa at a time when the language spoken in Samoa subgrouped with those now spoken in the Equatorial Outliers, and that Samoan has changed dramatically since then due to both internal change and innovations spreading from other languages in the Western Polynesian interaction zone (Pawley 1996:401–403).

Clark, Ross. 1982. Proto Polynesian birds. In J. Siikala, ed. *Oceanic Studies: essays in honour of Aarne A. Koskinen*, 121–43. Helsinki: The Finnish Anthropological Society.

—— 1998. *A dictionary of the Mele language (Atara Imere), Vanuatu*. Canberra: Pacific Linguistics.

Clunie, Fergus. 1984. *Birds of the Fiji bush*. Suva: Fiji Museum.

Crowley, Terry. 1994. Proto-who drank kava? In Pawley and Ross, eds 1994:87–100.

Degener, Otto. 1973. *Plants of Hawaii National Parks illustrative of plants and customs of the South Seas*. Ann Arbor: Braun-Brumfield.

Dening, G.M. 1962. The geographical knowledge of the Polynesians and the nature of inter-island contact. In Jack Golson, ed. *Polynesian navigation*, 102–153. Wellington: A.H. and A.W. Reed, for the Polynesian Society.

Elbert, Samuel H. 1953. Internal relationships of the Polynesian languages and dialects. *Southwestern Journal of Anthropology* 9:147–173.

—— 1975. *Dictionary of the language of Rennell and Bellona*. Copenhagen: National Museum of Denmark.

Finney, Ben. 1994. *Voyage of rediscovery: a cultural odyssey through Polynesia*. Berkeley: University of California Press.

Firth, Raymond. 1985. *Tikopia–English dictionary*. Auckland and Oxford: Auckland University Press and Oxford University Press.

Fischer, Steven Roger. 2001. Mangarevan doublets : preliminary evidence for Proto-Southeast Polynesian. *Oceanic Linguistics* 40,1:112–24.

Geraghty, Paul A. 1983. *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawaii Press.

—— 1994. Linguistic evidence for the Tongan Empire. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world*. Trends in Linguistics Studies and Monographs 77, 223–349. Berlin: Mouton de Gruyter.

—— 2004a. Borrowed plants in Fiji and Polynesia: some linguistic evidence. In Jan Tent and Paul Geraghty, eds *Borrowing: a Pacific perspective*, 65–98. Canberra: Pacific Linguistics.

—— 2004b. Polynesian loans in the Solomon Islands. *Rongorongo Studies* 14,2:43–68

Göthesson, Lars-Åke. 1997. *Plants of the Pitcairn Islands including local names and uses*. Sydney: Centre for South Pacific Studies, University of New South Wales.

Green, Roger. 1966. Linguistic subgrouping within Polynesia: the implications for prehistoric settlement. *Journal of the Polynesian Society* 75,1:6–38.

Haddon, A.C. and James Hornell. 1975. *Canoes of Oceania*. Special Publications 27, 28, and 29. Honolulu: Bernice P Bishop Museum.

Hale, Horatio. 1846. *United States exploring expedition during the years 1838, 1839, 1840, 1841, 1842, under the command of Charles Wilkes U.S.N: ethnography and philology*. Philadelphia: Lea and Blanchard.

Haudricourt, A.G. 1964. Comment on G.W. Grace, 'Movement of the Malayo-Polynesians: 1500 B.C. to A.D. 500. The linguistic evidence.' *Current Anthropology* 5:389–390.



Hollyman, K.J. 1987. *De Muna Fagauvea — I: Dictionnaire fagauvea–français*. Te Reo monographs. Auckland: Linguistic Society of New Zealand.

Inia, Elizabeth K., Sofie Arntsen, Hans Schmidt, Jan Rensel and Alan Howard. 1998. *A new Rotuman dictionary*. Suva: Institute of Pacific Studies, University of the South Pacific.

Kirch, Patrick Vinton. 2000. *On the road of the winds: an archaeological history of thePacific Islands before European contact*. Berkeley: University of California Press.

Langridge, Marta and Jennifer Terrell. 1988. *Von den Steinen's Marquesan myths*. Canberra: Target Oceania and The Journal of Pacific History.

Lemaître, Yves. 1973. *Lexique du tahitien contemporain*. Paris: Editions de l'ORSTOM.

Mahdi, Waruno. 1994. Some Austronesian maverick protoforms with culture–historical implications — I. *Oceanic Linguistics* 33,1:167–229.

Marck, Jeff. 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.

Milner, G.B. 1966. *Samoan dictionary*. Oxford: Oxford University Press.

Moyse-Faurie, Claire. 1993. *Dictionnaire futunien–français*. Paris: Peeters.

Nunn, Patrick D. 2008. A shock to the system: climatic disruption to Pacific Island societies around AD 1300. Paper presented at the Pacific History Association Conference, University of the South Pacific, Suva.

Pawley, Andrew. 1966. Polynesian languages: a subgrouping based on shared innovations in morphology. *Journal of the Polynesian Society* 75,1:39–64

—— 1996. On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In J.M. Davidson et al., eds *Oceanic culture history: essays in honour of Roger Green*, 387–410. New Zealand Journal of Archaeology Special Publication.

Pawley, Andrew and Kaye Green. 1971. Lexical evidence for the Proto-Polynesian homeland. *Te Reo* 14:1–35.

Pawley, Andrew and Malcolm Ross, eds. 1994. *Austronesian terminologies: continuity and change*. Canberra: Pacific Linguistics.

Rehg, Kenneth L. and Damian G. Sohl. 1979. *Ponapean–English dictionary*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

Rensch, Karl H. 1984. *Tikisionalio Fakauvea–Fakafalani/Dictionnaire Wallisien–Français*. Canberra: Pacific Linguistics.

—— 2005. *Plant names of Eastern Polynesia*. Canberra: Archipelago Press.

Sallés, Arturo Hernández, Nelly Ramos Pizarro, et al. 2001. *Diccionario Illustrado Rapa Nui, Español, Ingles, Francés*. Santiago: Pehuén.

Sharp, Andrew. 1963. *Ancient voyagers in Polynesia*. Wellington: Angus and Robertson.

Sperlich, Wolfgang B. 1997. *Tohi vagahau Niue: Niue language dictionary, Niuean–English with English–Niuean finderlist*. Government of Niue.

Stimson, J. Frank with the collaboration of Donald Stanley Marshall. 1964. *A dictionary of some Tuamotuan dialects of the Polynesian language*. The Hague: Martinus Nijhoff.

Triffitt, Geraldine. 2000. The dialects of the Yasawa Islands of Fiji. In Bill Palmer and Paul Geraghty, eds *SICOL: Proceedings of the Second International Conference on Oceanic Linguistics:* vol. 2, *Historical and descriptive studies*, 315–327. Canberra: Pacific Linguistics.

Whistler, W. Arthur. 1995. Folk plant nomenclature in Polynesia. *Pacific Studies* 18,4:39–59

Wilson, William H. 1985. Evidence for an outlier source for the Proto Eastern Polynesian pronominal system. *Oceanic Linguistics* 24:85–133.

—— 2008. Source of Tongic-like developments in PPN *r and *h in Eastern Polynesia. Paper presented at the Pacific Roots Symposium, University of the South Pacific, Suva, Fiji.

# 27 *Some clan names of the Chuukic-speaking peoples of Micronesia*

JEFF MARCK

## 1 Introduction

This report considers the antiquity of the clan names of the Chuukic-speaking peoples. The Chuukic-speaking islands (Figure 1) constitute the largest region of cognate matrilineal or patrilineal clan names in Oceania. We are presently confronted with a diverse Chuukic clan situation. Clan numbers are small on the atolls, usually less than ten. New clans abound in Chuuk Lagoon where there are now more than 80 clans. Saipan and other Mariana Islands have more than others due to immigration in the historic period. Twenty-seven clan names were found that occur in two or more of the Chuukic-speaking islands (counting Chuuk Lagoon as a single island) (Table 1). As few as six show evidence that allow attribution to Proto Chuukic, the language spoken around the Chuuk Lagoon[1] ca. AD 500–1000 before spreading on to the atolls of what is now Yap State, Federated States of Micronesian, and the atolls of the Republic of Palau.

The 27 clan names reconstructed with what may have been their Proto Chuukic[2] sounds are listed in Table 1 along with a guess at what may have been their meanings in English. I do not believe all 27 are as old as Proto Chuukic but it is convenient to begin by indexing each with a Proto Chuukic spelling. Table 2 then gives these same names alphabetically with their distributions according to islands for which I found regularly or irregularly agreeing forms. Table 3 then gives these same names grouped according to pattern of island distributions.

1 And perhaps the Mortlocks, Chuuk State's 'Western Islands' and Chuuk State's northern atolls.

2 The language spoken around Chuuk Lagoon some 1000 years ago or thereabouts at a time when those peoples were on the threshold of establishing permanent settlements on the atolls between Chuuk and Yap.

Alexander Adelaar and Andrew Pawley, eds
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 461–487.
Pacific Linguistics, 2009.

rituals were disappearing or were already lost when ethnographers first became interested in their collection early last century. Any re-emergence of the activity as in Nauru will include new patterns and associations with non-traditional objects. So although it seems entirely possible that string games and frequently used moves were known to Proto Oceanic speakers, it is unlikely that comparative linguistics can ever offer lexical proof beyond the somewhat tentative reconstruction of POc **paRi* as the generic term for string games and the activity of playing them.

## References


Davidson, D.S. 1941. Aboriginal Australian string figures. *Proceedings of the American Philosophical Society* 84(6):763–901.

de Coppet, Christa. 1978. Preface to Honor Maude, *Solomon Islands string figures*. xvii-xix.

Emory, K.P. and Honor Maude. 1979. *String figures of the Tuamotus*. Canberra: The Home Press.

Firth, Raymond and Honor Maude. 1970. *Tikopia string figures*. London: Royal Anthropological Institute of Great Britain and Ireland.

Haddon, K. 1930. *Artists in string*. London: Methuen and Co. Ltd.

Handy, Willowdean Chatterton. 1925. String figures from the Marquesas and Society Islands. *Bernice P. Bishop Museum Bulletin* 18:1–92.

Hornell, James. 1927. String figures from Fiji and Western Polynesia. *Bernice P. Bishop Museum Bulletin* 39:1–88.

Jenness, D. 1920. Papuan cat's cradle. *Journal of the Royal Anthropological Institute* 50:299–326.

Kirch, Patrick and Roger C. Green. 2001. *Hawaiki, Ancestral Polynesia*. Cambridge: Cambridge University Press.

Maude, Honor. 1978. *Solomon Islands string figures*. Canberra: The Home Press.

—— 1984. *String figures from New Caledonia and the Loyalty Islands*. Canberra: The Home Press.

—— 2001 (revised first edition 1971 from fieldnotes made 1937–38). *The string figures of Nauru Island*. The University of the South Pacific Centre in Nauru and Institute of Pacific Studies.

Maude, H.C. and H.E. Maude. 1958. *String figures from the Gilbert Islands*. Wellington: The Polynesian Society.

Maude, Honor and Camilla Wedgwood. 1967. String figures from Northern New Guinea. *Oceania* 37:202–229.

Noble, P.D. 1979. *String figures of Papua New Guinea*. Boroko: Institute of Papua New Guinea Studies.

Rivers, W.H.R. and A.C. Haddon. 1902. A method of recording string figures and tricks. In *Man* 1–2:146–153.

Rosser, W.E. and J. Hornell. 1932. String figures from British New Guinea. *Journal of the Royal Anthropological Institute* 62:39–50.


# 30 *The role of the Solomon Islands in the first settlement of Remote Oceania: bringing linguistic evidence to an archaeological debate*

ANDREW PAWLEY

## 1 Introduction

This paper looks at some problematic aspects of the history of human settlement of the Solomon Islands over the last three millennia.[1] The initial spread of Oceanic languages into Remote Oceania[2] can be strongly associated with the movement into the Reefs/Santa Cruz group and Vanuatu, at about 3200–3100 BP, of bearers of the archaeological culture known as Lapita. Lapita is first attested in the Bismarck archipelago and on geographic grounds one would expect the islands in the main Solomons group (extending from Bougainville to Makira) to have been stepping stones for the Lapita expansion eastwards into Remote Oceania. Thus, archaeologists have been puzzled as to why no early Lapita archaeological sites been found in the main Solomons group, and why almost no pottery-bearing sites of any kind have been found in the southeastern part of the group. Does this mean that the main Solomons group was bypassed in the initial Lapita colonisation of Remote Oceania, as was suggested by Sheppard and Walter (2006), or is the archaeological record too fragmentary to allow any firm conclusions to be drawn?

[1] I am delighted to contribute to a volume honouring Bob Blust's distinguished and diverse contributions to Austronesian historical linguistics and culture history. An earlier version of this paper was presented at the 7th International Conference on Oceanic Linguistics, Noumea, July 2007. The paper has benefited from discussions with Roger Green, Stuart Bedford, Bethwyn Evans, Frank Lichtenberk, Malcolm Ross, Matthew Spriggs and Darrell Tryon.

[2] Whereas 'Near Oceania' consists of New Guinea, the Bismarck Archipelago and the main Solomons Archipelago (ending at Makira), which form a chain of largely intervisible islands, 'Remote Oceania' consists of the remaining, much more widely dispersed islands and island groups of the SW and Central Pacific (chiefly those of Vanuatu, New Caledonia, Fiji, Micronesia and Polynesia).

Alexander Adelaar and Andrew Pawley, eds
*Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 515–540.
Pacific Linguistics, 2009.

I will address three questions concerning the history of the Oceanic languages of the Solomons that have a bearing on this issue:

(1) Given that there is no major geographic barrier that would account for an early and sharp separation of these subgroups, what circumstances created the major subgroup boundary that runs through the centre of the Solomons archipelago, separating Northwest Solomonic from Southeast Solomonic?

(2) How long have the Northwest and Southeast Solomonic groups been in their present locations?

(3) Why have the Northwest Solomonic languages replaced a much higher percentage of Proto Oceanic core basic vocabulary items than Southeast Solomonic languages?

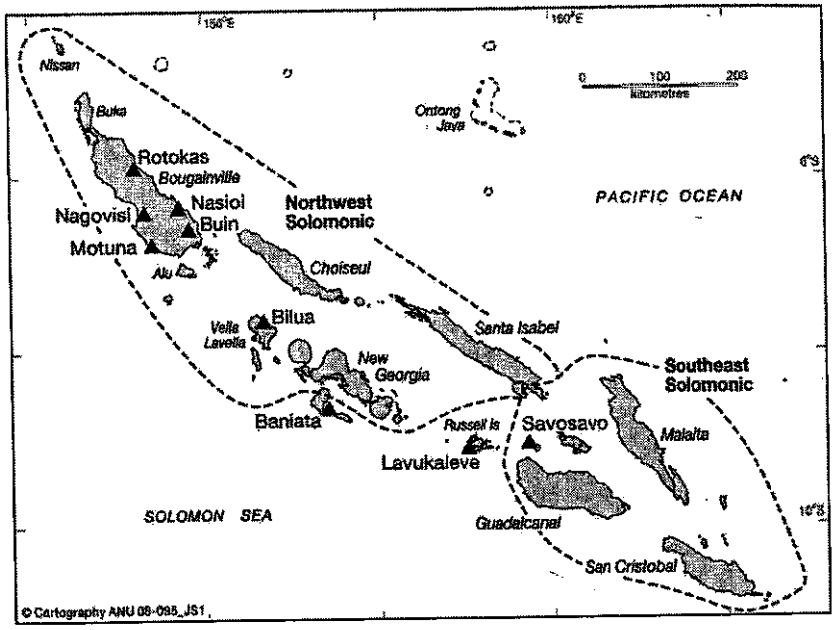



**Figure 1:** Boundaries of Northwest Solomonic and Southeast Solomonic and locations of non-Austronesian languages

## 2 The Solomons archipelago

Because the geographic span of the main group of Solomon Islands differs markedly from that of the nation called 'the Solomon Islands' I will refer to the former as 'the Solomons archipelago' or 'the main Solomons group'. The archipelago consists of a chain of closely spaced large islands that extends for about 1000 km from northwest to southeast (see Figure 1). The main islands are quite large: Bougainville is about 10,000 square



kilometres, Guadalcanal 6500, Makira 4600, Malaita 3900, Choiseul 3000, and New Georgia 2100. All the large islands are mountainous and heavily forested. Typically there is a narrow coastal strip of strand forest of sandy soil with light forest of salt-resistant trees and patches of mangrove and sago swamp. Man-made grasslands occur in some areas, most extensively in the plains of northern Guadalcanal. In several regions there are extensive fringing coral reefs and lagoons carrying a rich biota.

It will be convenient to distinguish between a Northwest Solomons region, including Buka, Bougainville, Choiseul, the New Georgia group and Santa Isabel, and a Southeast Solomons region, including Guadalcanal, Florida, Malaita and Makira. Buka and Bougainville are separated from New Ireland by an ocean gap of 180 km, with only the small island group of Nissan (aka Nehan or the Green Is.) in between. Some 400 km of open sea separate Makira from the small Santa Cruz-Reef Is. group to the east.

Humans reached New Guinea, New Britain and New Ireland by 40,000 years ago and by about 30,000 years ago had settled Greater Bougainville in the NW Solomons (Kirch 2000; Spriggs 1997; Specht 2005) at a time of lower sea levels, when the island of Bougainville extended from what is now Buka almost to Guadalcanal. However, the Solomons archipelago remained the limit of human expansion into the Southwest Pacific until just over 3000 years ago. Until then it appears that people lacked sailing craft capable of making the long crossings to islands further east, against the prevailing trade winds and currents.

## 3 The spread of Lapita as a marker of the dispersal of Oceanic languages

In the second half of the 2nd millennium BC people bearing a new language and technology entered Northwest Melanesia. These were fishermen–farmers from Southeast Asia, who by 3400–3300 BP had settled in various parts of the Bismarck Archipelago, chiefly on small islands, where they established the first nucleated villages known in Melanesia (Green 2003; Kirch 2000; Specht 2005; Spriggs 1997; Summerhayes 2000, 2001). The most visible archaeological marker of this Neolithic culture is its highly distinctive decorated pottery, with elaborated motifs impressed by dentate-stamping. In sites representing permanent habitations the decorated pottery is part of a cluster of distinctive elements: settlement patterns, rectangular houses raised in stilts, an array of ceramic vessel forms, mainly undecorated, fishing gear, adze/axe kit, shell ornaments and evidence of long distance exchange of obsidian. The pottery tradition is known as Lapita, after which the cultural complex as a whole is named. Changes in the styles and proportions of decorated pots lend themselves to the construction of a fine-grained seriation chronology which can supplement $C^{14}$ dating of Lapita assemblages. Many elements of the Lapita complex have close parallels in Neolithic cultures that appear in Taiwan, the Philippines and the Marianas and parts of Indonesia in the early to mid 2nd millennium BC (Bellwood 1997; Bellwood and Dixon 2005; Green 2003; Kirch 1997, 2000).

The sudden appearance of this distinctive cultural complex in the Bismarck Archipelago can be strongly associated with the arrival there of Austronesian languages, and specifically with the separation of the large Oceanic branch from its nearest relatives, spoken in the Cenderawasih Bay area at the western end of New Guinea, and in South Halmahera (Blust 1978a). Oceanic is a well-defined subgroup which contains all the Austronesian languages of Melanesia except the western end of New Guinea, plus those of Polynesia and (with two exceptions) Micronesia. The lexicon of Proto Oceanic has been

reconstructed in considerable detail (Ross et al. 1998–2008, in prep.) and, when compared with the lexicon reconstructed for Proto Malayo-Polynesian (Blust 1995) shows a fairly high degree of continuity in terminologies for various domains of material culture and social organization (Green 2003; Pawley 2007).

The earliest attested phase of Lapita in the Bismarcks is known as Early Western Lapita, which appears between 3400 and 3300 BP. Around 3200 BP or soon after bearers of the Early Western Lapita culture moved east of the Bismarck Archipelago into Remote Oceania. The Reefs/Santa Cruz group, some 400 km east of Makira, contains one of the earliest and most extensively excavated Lapita sites in Remote Oceania. Site SZ–8R, with initial occupation dated to between 3200–3100 BP (Green 1991, 2003, pers. comm.) is among 19 Lapita sites in Reefs/Santa Cruz. For some time the Lapita occupants of this group kept importing considerable quantities of obsidian from Talasea in New Britain, an indication that initially they maintained trade links with the homeland. Some Talasea obsidian appears in early Northern Vanuatu Lapita sites, a strong indication that this region was settled at about the same time as Reefs/Santa Cruz (Bedford 2003, Bedford et al. 2006). By 3050 BP, Lapita people had occupied New Caledonia (Sand 2001) and Fiji (Nunn et al. 2004). By 2950 BP they were in Tonga (Burley et al. 2007) and by 2800–2700 BP they were in Samoa and some of the other islands in the Tonga–Samoa voyaging corridor (Kirch 1997; Green 2003). In each of these island groups in Remote Oceania the distinctive Lapita decorated ware disappeared within a few centuries of first settlement but in most regions some other features of the Lapita cultural complex including, as a rule, the plain ware ceramic vessel forms, continued for much longer.

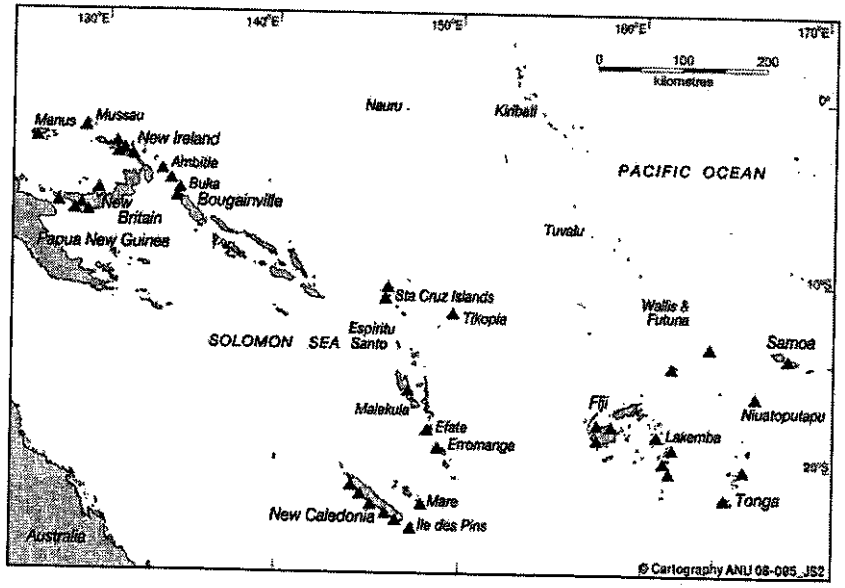



**Figure 2:** The distribution of important Lapita sites (after Spriggs 1995:113)



It appears that there was a pause of about 200–300 years in the Bismarcks before bearers of the Lapita culture moved eastwards into Remote Oceania. The final stages in the development of Proto Oceanic (POc) can be associated with this pause (Blust 1998; Pawley 2003a, 2008). The initial eastward migrations of Lapita people mark the spread of Oceanic languages into Remote Oceania. All but two of the 180–190 indigenous languages spoken in Remote Oceania at time of first European contact belong to the Oceanic subgroup. The two exceptions are two languages on the western margin of Micronesia, Chamorro and Palauan; both are Austronesian languages that probably stem from movements out of the Philippines or Indonesia before 3000 BP

## 4 Archaeological debates over Lapita settlement of the Solomons archipelago

Given the position and size of the main Solomons group one would expect colonies to have been established there during the first Lapita movements eastward from the Bismarcks. However, although Early Western Lapita sites have been found immediately to the west of the Solomons, on Nissan (3200 BP, Summerhayes 2000, 2001), and slightly later sites on Buka (3000 BP, Wickler 2001), no Lapita sites associated with the initial Lapita expansion of 3200–3000 BP have so far been identified in the main Solomons group east of Buka. The nearest approximations are various sites in the New Georgia group, chiefly in the Roviana Lagoon, which contain the remnants of stilt-house settlements built over the intertidal zone. These are evidenced by residues of potsherds and some stone tools in shallow water, one to two metres below the surface (Felgate 2001, 2003, 2007). The Roviana Lagoon sites are dated by seriation chronologies of ceramic styles as being late Lapita, around 2700–2400 BP.

The absence of Early Western Lapita pottery from the NW Solomons, and the almost complete absence of any pottery finds in the SE Solomons, has led to a lively debate among archaeologists about the role of the Solomons archipelago in the early Lapita settlement of Remote Oceania. Two competing sets of proposals have emerged, which I will refer to as the 'early settlement' and 'late settlement' hypotheses.

In a recent review of Solomons archaeology Sheppard and Walter (2006) put forward the following proposals:

(i) The early Lapita colonists leapfrogged the main Solomons group, moving directly to the Reefs/Santa Cruz Is. about 3200–3100 BP. (A similar proposal had been adumbrated by Roe 1993.) For a time the Reefs/Santa Cruz settlers maintained long distance obsidian trade connections with the Bismarck archipelago, as well as obtaining chert from Malaita or Ulawa and basalt for adzes from southeast Guadalcanal.

(ii) Several centuries later, ca 2700 BP, the NW Solomons were settled by Austronesian-speaking, farming, pottery-making populations who moved from the west (the Bismarcks) and whose languages in time became dominant over the non-Austronesian autochthonous languages.

(iii) More tentatively, they propose that Austronesian speakers did not settle the southeastern islands in the main Solomons chain (Guadalcanal, Malaita and Makira) until some 800–1000 years after the initial Lapita dispersal into Remote Oceania. Around 2300–2200 BP, these islands were settled by an a-ceramic, farming population coming from the Reefs/Santa Cruz group and/or Utupua and Vanikoro, where manufacture of pottery ceased about 2100 BP.

This scenario would of course explain the sharp linguistic boundary between the NW and SE Solomonic groups.

Felgate (2001, 2003, 2007) takes a more cautious view regarding the absence of early Lapita sites in the NW Solomons. He suggests that early Lapita occupation of the NW Solomons is likely to have been low density, because of the presence there of established non-Austronesian populations and perhaps because of malaria. He points out that archaeological surveys there have been mainly terrestrial, whereas Lapita settlements are likely to have consisted of stilt houses built over the edge of the lagoon, a pattern attested for late Lapita sites in the New Georgia region, as it is for a number of regions further west (Kirch 2000; Spriggs 1997). Felgate (2001:57) favours the view that:

> a pattern of intertidal settlement [in the Lapita period] has created the dual conditions of low site preservation/visibility and unexpected site location. Implicit in this proposition is a suggestion that early Lapita may have been continuously distributed across the Near Oceanic Solomon Islands in the past, as a shifting network of interacting settlements, located exclusively over the tidal zone, of which we are likely to find only rare traces in settings favourable to their preservation.

Felgate's critics feel that he overstates the domination of intertidal sites in the Lapita settlement of the New Georgia group. Sheppard has recently reanalysed the geomorphic context of inter-tidal sites there and concludes that it is unlikely that an Early Lapita record has been obliterated by submersion (Sheppard pers. comm.). Insofar as there is a consensus on this matter, it is that the earliest material in the Roviana Lagoon dates to around 2700 BP and represents the late end of dentate-stamped pottery, after which decorations on pots were made using a different technique.

Archaeological surveys of the SE Solomons from Guadalcanal to Makira have so far found almost no ceramics. This stands in sharp contrast with the NW Solomons, where pot sherds are highly visible on all the main islands, and it is clear that pots continued to be made long after the Lapita period. The pollen record for Guadacanal gives evidence of intensive slash and burn horticulture there beginning around 2300–2200 BP (Haberle 1996; Roe 1993) and the faunal record also points to increased predation and extinction of larger species about that time (Spriggs 1997). Comparing these indicators of the first appearance of large scale shifting argriculture in Guadacanal with earlier dates for similar signs in Aneityum and New Caledonia, Spriggs (1997:149) comments '[t]he nearly 800 year time lag on Guadalcanal and the lack of pottery in any of the sites so far investigated suggests that Austronesian settlement here was delayed until pottery was no longer in use in the region'.

However, there is reason to think this suggestion is premature. The best surveyed of the main islands in the SE Solomons is Guadalcanal but even there the archaeological record is poor. Malaita remains virtually an archaeological blank. A few small excavations have been carried out on Makira, Uki and and Ulawa, yielding no pottery or early dates. The solitary exception is a rock shelter on Santa Ana which contained plain (undecorated) ware ceramics of late Lapita type, dating to about 2900 BP (Green pers. comm.).

While it seems clear that the inhabitants of the SE Solomons have not made pottery during the past 2000 years, the scarcity of Lapita pottery in a region with a poor archaeological record should not necessarily be taken to indicate that the rest of the Lapita cultural complex was also absent. While pottery is an invaluable aid in finding sites and in dating assemblages, it was just one component in a rich Lapita cultural tradition. Phases 2 and 3 of Vatulama Posovi, a cave site in the Poha Valley, near Honiara on Guadalcanal,



have yielded an assemblage of artefacts dated to around 3250–2900 BP and 2750–2550 BP which has been described as 'Lapita without pots' (Roe 1993). Around 3000 BP the Lapita settlers of Reefs/Santa Cruz were importing basalt for adzes from Marau Sound on SE Guadalcanal, chert for blades from Ulawa and/or Malaita and temper for pots from part of the Florida group, off N. Guadalcanal, and it would be strange if they did not establish settlements or interact with sister Lapita colonists in these places. The Santa Ana rock shelter site is presumably the byproduct of one such settlement.

In the sections that follow I will discuss some linguistic evidence that bears on these archaeological issues.

## 5 The language groups of the Solomons Archipelago

### 5.1 Overview

In many cases dialect chaining makes it hard to draw language boundaries without some degree of arbitrariness, but on a conservative estimate there are 60 or so mutually-unintelligible languages spoken in the Solomons archipelago. Some 50 of these languages belong to the large Oceanic branch of Austronesian. Another 12 or so are non-Austronesian ('Papuan') and fall into at least four different families that cannot, on present evidence, be convincingly shown to share a common origin (Ross 2001; Dunn et al. 2002, 2005).

Except on Bougainville, where they occur in coastal pockets, the Oceanic languages in the Solomons have a continuous distribution over all the habitable parts of the larger islands. Two major subgroups of Oceanic are represented there: Northwest Solomonic and Southeast Solomonic. The boundary between them runs roughly north–south between Santa Isabel in the west, and Guadalcanal and Malaita in the east. SE Solomonic languages are spoken on Guadalcanal and the Florida group, Makira, and Malaita. A single SE Solomonic language, Bugotu, is spoken on the south-eastern tip of Santa Isabel, where it is clearly represents an intrusive settlement from the Florida group or Guadalcanal within the last 1000 years. NW Solomonic comprises the Oceanic languages of Santa Isabel (other than Bugotu), the New Georgia group, Choiseul, Bougainville, Buka and the small Nissan island group which lies between New Ireland and Buka.

The few surviving non-Austronesian languages in the Solomons Archipelago are plainly the residue of a larger number that were present in this region when speakers of Oceanic Austronesian arrived. The surviving languages are genetically very diverse (Ross 2001; Dunn et al. 2002). According to Ross (2001), Bougainville contains two families of non-Austronesian languages with four members each. There are two non-Austronesian languages in the New Georgia group and two occupying the small islands of Russell and Savo to the northwest of Guadalcanal.[3]

Presumably, non-Austronesian languages were once spoken on all the main Solomon islands at least as far east as Guadalcanal, and possibly on Malaita and Makira as well. The pre-Austronesian populations were probably mobile foragers and this mode of life, in combination with the rugged and densely forested nature of the islands, and the lack of large terrestrial animals to hunt and, in some islands, the scarcity of fringing reefs, would have severely limited their numbers and distribution.

[3] The non-Polynesian languages of Santa Cruz, and Äiwoo of the Reefs, have sometimes been classified as non-Austronesian but recent work has strengthened the case made in Lincoln (1978) that they are Oceanic languages that have undergone an unusual amount of phonological and morphological change. It is likely that they fall together in a single first-order subgroup of Nuclear Oceanic (Ross and Næss 2007).

## 5.2 Southeast Solomonic and its subgroups

### 5.2.1 Southeast Solomonic

The existence of a SE Solomonic (SES) subgroup is uncontroversial. Milke (1958) and Grace (1959) observed that this group is defined by the merger of POc *l and *R, an unusual merger in the Austronesian family. A larger body of morphological innovations defining SE Solomonic was set forth in Pawley (1972), e.g. development of a special suffix marking inanimate 3rd person plural pronouns: Proto SES *-ki (direct object), *-ni possessor; replacement of POc preverbal subject markers *ku '1SG', *ko '2SG', *na '3SG' by Proto SES *u, *o and *e; replacement of the POc possessive pronoun *-da '1inc.pl.' by the independent form *-kita, used as a possessive.

However, the quantity of shared innovations defining SES is quite small. This indicates that the period of unified development of SES after it diverged from other Oceanic languages was no more than a few centuries, after which its two primary subgroups, Makira-Malaitan and Guadalcanal-Gelic, began to diverge.

### 5.2.2 Makira-Malaitan

Makira-Malaitan (MkMl) consists of some 13 languages. Seven are spoken on Malaita and its satellites (including Ulawa and Ugi, lying between Malaita and Makira), four on Makira, and two at the eastern end of Guadalcanal (the latter are both clearly intruders from Malaita or Makira). This subgroup is marked by a number of changes to the Proto SES sound system (Lichtenberk 1988, 1994; Pawley 1972; Tryon and Hackman 1983): *t was lost in Proto MkMl, *s > *t except before high vowels, *k > glottal stop in most cases and there was accretion of a prothetic consonant *y- before initial *a. There are also a few irregular changes in particular grammatical forms. POc *-kita '1inc.pl.' in Proto MkMl reduced to *-ka (presumably via *kia, after regular loss of *t). The Pre MkMl 1st inclusive trial form *kita-tolu reduced to *kaolu, and the 1st exclusive trial form reduced from *kami-tolu to Proto MkMl *ʔamelu.

From the pattern of overlapping isoglosses it is pretty clear that Proto MkMl persisted for many centuries as a chain of dialects extending over both Malaita and Makira (Lichtenberk 1988, 1994). While the geographic extremes in this chain began to diverge very early they remained connected by intermediate dialects. (See §8 for further discussion.)

### 5.2.3 Guadalalcanal-Gelic

Guadalalcanal-Gelic (GG) contains about seven languages. On Guadalcanal (where dialect chaining complicates the count) there are perhaps five languages. There is one (Gela) in the Florida group and another, Bugotu, is spoken at the eastern end of Santa Isabel.

Two phonological innovations mark GG: POc *w is lost in word initial position; *m and *mw merge as *m*. There are a few morphophonemic or irregular phonological changes, e.g. when certain disyllabic roots are reduplicated the second consonant drops out in the first root, e.g. Gela *taitahi* 'salt' instead of **tahitahi*. Proto SES *no- 'marker of general possessive relation' irregularly became Proto GG *ni-. It is clear that Proto GG was spoken on Guadalcanal and probably also on Florida.



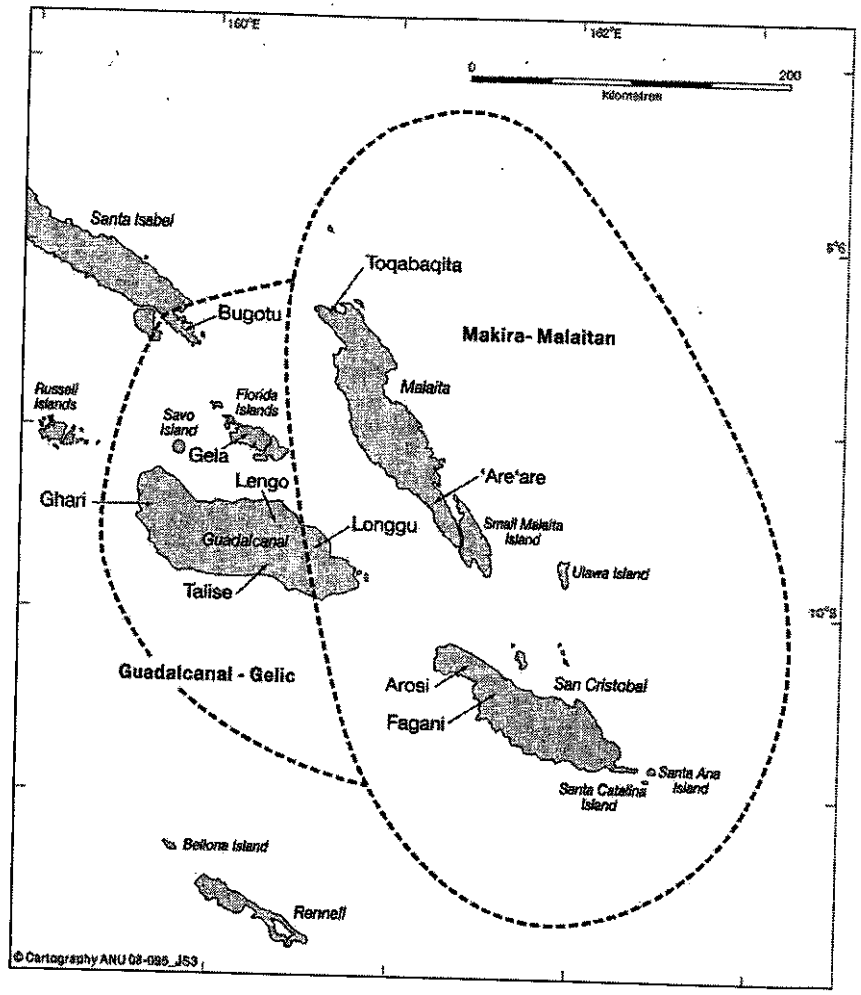



**Figure 3:** The primary subgroups of SE Solomonic, with languages mentioned in text

## 5.3 Northwest Solomonic and its subgroups

### 5.3.1 Northwest Solomonic

The Northwest Solomonic group was not recognised until the early 1980s. Tryon and Hackmann (1983) showed that all the languages from the Shortland Islands to Santa Isabel share a few innovations defining them as a single, though very heterogenous subgroup which they called 'Western Solomons'. Ross (1986, 1988) showed that this group also includes the languages of Bougainville, Buka and Nissan.

Three regular sound changes are attributed to Proto NWS: (i) POc *w is lost in all positions, (ii) an 'echo' vowel is added after word-final consonants, e.g. *onom 'six' > PNWS *onomo, (iii) POc word-final *q becomes PNWS *k, whereas initial and medial *q was either lost or merged with *γ. The POc 1st person singular independent pronoun *au was replaced in PNWS by *(a)rau. The relatively small number of innovations defining NWS indicates that the period of unified development was quite short.

### 5.3.2 Subgroups of Northwest Solomonic

Ross distinguished five primary branches of NWS: (1) Nissan-Buka-North Bougainville (10 languages), (2) Piva-Banoni (W. Bougainville) (two languages), (3) S. Bougainville-Shortlands (three languages), (5) Choiseul (four languages), and, more tentatively, (5) New Georgia-Santa Isabel (16 languages). Although New Georgia and Santa Isabel are each well-defined groups the evidence for uniting them is slender and any period of common development must have been very brief. For our purposes it is more useful to treat New Georgia (nine languages) and Santa Isabel (seven) as separate primary groups.

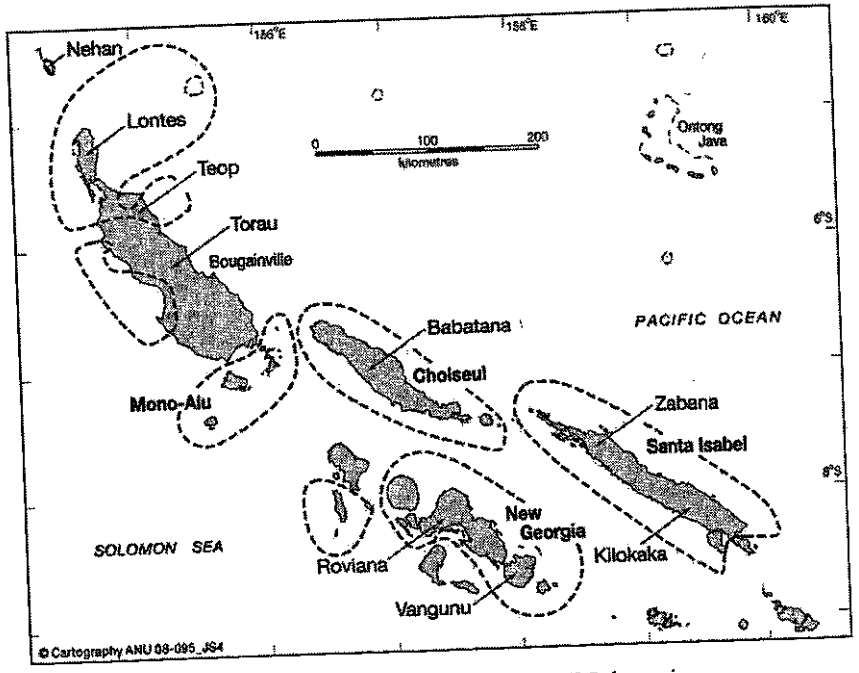



**Figure 4:** The primary subgroups of NW Solomonic, with languages mentioned in text.



## 6 Why is there a deep boundary between NW Solomonic and SE Solomonic?

Let us return now to the question of why there is a major subgroup boundary between Northwest Solomonic and Southeast Solomonic. There is no major geographic barrier that would account for this boundary. Ocean gaps between Santa Isabel and Malaita, and between Santa Isabel and the Florida group are on the order of 50 km—i.e. no greater than some of the distances separating islands within the NW Solomonic or the SE Solomonic regions. Oceanic speakers who settled the Solomons certainly had the sailing capacity to maintain regular communication across such ocean gaps. Nor are there other obvious environmental factors, such as periods of explosive vulcanism or the absence of key natural resources, which might account for the boundary

It seems, then, that we must look for an explanation of this boundary in terms of historical and social factors. An obvious question is: do NWS and SES belong to different branches of Oceanic, each with members elsewhere?

Our understanding of the high-order subgrouping of the Oceanic languages of western Melanesia rests largely on two important studies. Blust (1978b) showed that the 20 or so languages of the Admiralty and Western Is. form a closed subgroup. He also pointed to a single phonological change undergone by all other Oceanic languages except the Admiralities, namely the merger of Proto Austronesian *j and *s, and on this basis assigns all non-Admiralties languages to a single subgroup of Oceanic (Blust 1978b, 1998), which I will refer to here as 'Nuclear Oceanic'. Ross's (1988) monumental study encompassed all the Oceanic languages in 'western Melanesia' (defined as extending as far east as the boundary between NW Solomonic and SE Solomonic). He found evidence indicating that, within the Bismarck archipelago, there was an early two-way split between two primary branches of Oceanic: (i) an Admiralties subgroup, well defined by shared innovations, and (ii) a Western Oceanic (WOc) 'linkage', which includes all or almost other Oceanic languages of the Bismarcks and those of Papua New Guinea. A linkage is an imperfect subgroup, defined by innovations that link different sections of the chain, rather than by innovations shared by all members. A linkage derives from a well-differentiated dialect chain rather than a relatively homogeneous ancestor. Ross (1988) also noted the possibility that there was a third primary branch of Oceanic in western Melanesia, consisting of the small Mussau subgroup. He said little about Oceanic languages of the SE Solomons and Remote Oceania. However, he inclined to the view that these languages separated very early from Oceanic languages spoken in the Bismarcks, as the result of a single eastward movement from the Bismarcks through the Solomons and beyond into Remote Oceania.

Ross concluded that the Western Oceanic languages remained confined to the Bismarcks for some time, initially as a complex of dialects represented in parts of coastal north New Britain east of the Willaumez Peninsula, in Bali-Vitu (the French Is.) off the coast of New Britain, and in New Ireland and its offshore islands. At some point Western Oceanic dialects spread beyond this region in two directions: to the New Guinea mainland and to the NW Solomons. He found that the NW Solomonic languages share some innovations with Western Oceanic languages found in the Bismarcks that are not present in the Oceanic languages of the New Guinea mainland. These innovations define an imperfect subgroup that he called the Meso-Melanesian (MM) linkage. The diagnostic innovations are (i) merger of POc *r and *R as *r, (ii) merger of *d and dr as *d, (iv) merger of POc *c and *s as *s, (iv) the split of *k into *k and *γ, and (vi) the split of *p into *p and *v.

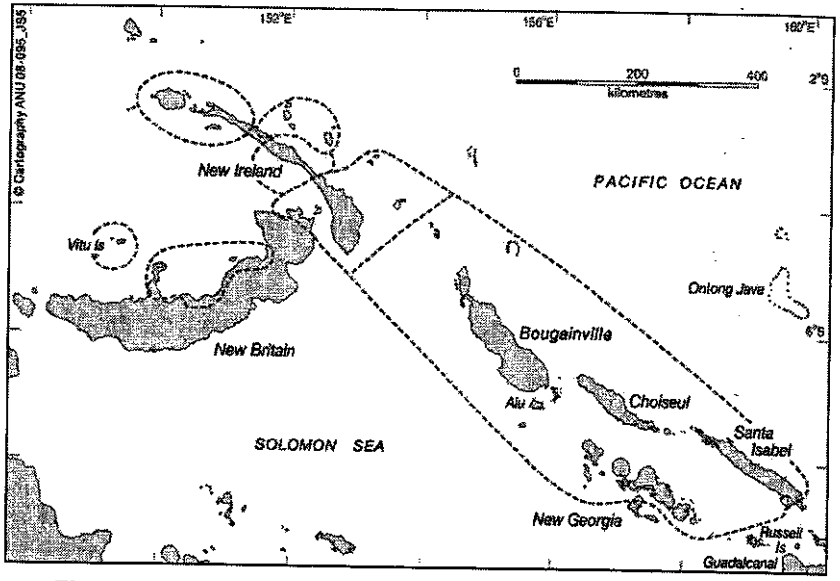



**Figure 5:** The Meso-Melanesian linkage and its subgroups (after Ross 1988)

There are fragments of evidence indicating that NW Solomonic stemmed from a particular area in the Meso-Melanesian linkage, namely a dialect network centred in southern New Ireland and perhaps extending to the nearby Tanga and Feni groups and to Nissan (Ross 1982, 1988). The evidence consists of a few innovations common to languages of that region and to the North Bougainville members of NW Solomonic. NW Solomonic then developed separately from the S. New Ireland/Tanga/Feni languages. The likely dispersal centre of NW Solomonic is the area consisting of Buka, N. Bougainville and Nissan.

Although Ross' work indicates that the ancestral NW Solomonic language arrived in the western Solomons some centuries after the breakup of POc, it does not explain why the expansion of NW Solomonic stopped at New Georgia and Santa Isabel. As part of the groundwork for tackling this question, I turn now to another vexing question: How intensive were interactions between incoming speakers of Oceanic languages and autochthonous speakers of non-Austronesian languages in different parts of the Solomons archipelago? Some evidence bearing in this question can be found in patterns of lexical replacements.

## 7 Evidence that NW Solomonic languages have replaced basic lexicon faster than SE Solomonic languages

### 7.1 Identifying the most stable 60 POc words

It has long been the impression of Oceanicists that SE Solomonic languages are among the most conservative members of the Oceanic group in respect of lexicon and that their sister languages in the NW Solomons have been more innovative. The usual explanation



for this difference is that the NW Solomonic languages have been strongly influenced by contact with non-Austronesian languages whereas SE Solomonic languages have not.[4] However, as far as I know no one has tried to measure rates of lexical replacement in the languages in question, or to pinpoint the periods when particular lexical changes took place. In order to achieve these two objectives, the rates of replacement in 60 highly stable words were investigated for a sample of SES, NWS and other Oceanic languages.

The following procedure was used to identify the 60 most stable POc words, i.e. the words with the highest retention rates in the daughter languages. (i) A first approximation was made by examining a table in Dyen et al. (1967) that ranks *word meanings* (not forms) on the Swadesh list of 200 basic lexical concepts according to how often pairs of languages had cognate forms for these meanings, using a sample of some 200 Austronesian languages. (ii) The 65 meanings yielding the highest percentages of cognate pairs were then extracted and the POc lexical form(s) reconstructable for each of these meanings were listed. (In five cases it was necessary to reconstruct pairs of synonymous forms and to count a retention of either etymon as a plus). (iii) A few problematic meanings were eliminated from the list, reducing it to 60. (iv) Retentions and losses for these etyma were recorded in 40 contemporary Oceanic languages drawn from various major subgroups.[5] (v) From these comparisons an average retention rate for each POc etymon was computed.

This procedure proved to have some flaws. It turned out that at least two of the lexical items that are among the most 20 stable items in Oceanic languages were missing from the variant of the Swadesh 200 word list used by Dyen et al. (1967), namely '(woman's) breast' and 'excrement'. In addition, several other etyma that are among the 60 most highly stable items in our Oceanic comparisons have meanings that do not appear in the top-ranked 65 items in Dyen et al.'s list. These included 'cry', 'night', 'tail', 'moon', 'star'

---

4 I have found no works specifically addressing the differences between NW Solomonic and SE Solomonic but there is a large literature on the effects of contact between Austronesian and Papuan languages in various parts of Melanesia. See Dutton and Tryon (1994), Pawley (2006), Blust (2005, 2008), Donohue and Denham (2008) for recent discussions.

5 The 40 languages in the sample used to calculate retention rates were:

| | | |
|---|---|---|
| SE Solomonic: | Guadalcanal-Gelic: | Bugotu, Gela, Talise |
| | Makira-Malaitan: | Arosi, Toqabaqita (To'aba'ita) |
| NW Solomonic: | Bougainville-Mono: | Mono, Teop, Torau, Lontes (Halia) |
| | Nehan: | Nehan |
| | Choiseul: | Babatana |
| | New Georgia: | Roviana, Vangunu |
| | Santa Isabel: | Kilokaka |
| Polynesian: | Samoan, Niuean | |
| Fijian: | Bauan (E. Fijian) | |
| Micronesian: | Marshallese, Woleai | |
| S. Vanuatu: | Erromangan (= Sye) | |
| N. Vanuatu: | Mota, Nguna, Raga | |
| Eastern Outer Is.: | Malo, Vano, Asumboa | |
| N. New Ireland: | Lihir, Tigak | |
| S. New Ireland: | Sursurunga, Kuanua | |
| New Ireland islands: | Anir | |
| W. New Britain: | Bali, Nakanai | |
| Manus: | Kele, Titan | |
| N. New Guinea: | Manam, Takia, Lote (= Pomio) | |
| Papuan Tip: | Motu, Galea | |

and 'where?'[6] No doubt these discrepancies arise in part from the different language samples used in the two studies but they are likely to be due mainly to the fact that Dyen et al. dealt with cognate percentages for meanings whereas my study deals with the retention rate of individual word forms. The discrepancies were not noticed until the analysis was well advanced and time constraints have prevented me from redoing the calculations. However, the fact that a few highly stable words were omitted from the list of 60 used in this study does not matter—given a list of highly stable items the important thing is how different languages behave with regard to these.

Average retention rates for the 60 items in a sample of 40 Oceanic languages are shown in Table 1.

**Table 1:** Retention rates for POc reconstructions for 60 highly stable items on the basic vocabulary list, based on 40 languages

| | | POc | % retained | | | POc | % retained |
|---|---|---|---|---|---|---|---|
| 1 | eye | mata | 97 | 31 | fruit | puaq | 60 |
| 2 | we excl | kami | 97 | 32 | new | paqoRu | 57 |
| 3 | we incl. | kita | 95 | 33 | dig | kali-, keli- | 56 |
| 4 | two | rua | 92 | 34 | bird | manuk | 56 |
| 5 | father | tama- | 90 | 35 | inside | lalom | 56 |
| 6 | you pl. | kam(i)u | 90 | 36 | path | jalan | 53 |
| 7 | they | ira | 90 | 37 | name | [ŋ,q]ajan | 52 |
| 8 | mother | tina- | 82 | 38 | head | qulu, pʷatu | 50 |
| 9 | louse | kutu | 82 | 39 | tooth | nipon | 50 |
| 10 | die | mate | 82 | 40 | woman | papine | 50 |
| 11 | five | lima | 82 | 41 | to fear | matakut | 50 |
| 12 | thou | iko, koe | 80 | 42 | root | wakaR, lamut | 50 |
| 13 | three | tolu | 80 | 43 | one | tasa, sakai | 50 |
| 14 | hear | roŋoR | 77 | 44 | liver | qate | 50 |
| 15 | four | pati | 75 | 45 | blood | draRaq | 47 |
| 16 | tongue | maya | 73 | 46 | water | waiR | 46 |
| 17 | I | [i]au | 73 | 47 | far | sauq | 46 |
| 18 | come | (lako) mai | 73 | 48 | skin | kulit | 43 |
| 19 | ear | taliŋa | 72 | 49 | feather | pulu | 43 |
| 20 | nose | ŋicu | 70 | 50 | rain | qusan | 42 |
| 21 | eat | kani | 70 | 51 | fire | api | 42 |
| 22 | drink | inum | 70 | 52 | leaf | draun | 40 |
| 23 | vomit | luaq, mumutaq | 70 | 53 | sky | laŋit | 40 |
| 24 | tree | kayu | 70 | 54 | thin | manipis | 40 |
| 25 | he/she | ia | 67 | 55 | ashes | drapu(R) | 37 |
| 26 | stone | patu | 67 | 56 | egg | katoluR | 36 |
| 27 | hand | lima | 66 | 57 | day | qaco, raqani | 36 |
| 28 | fish | ikan | 66 | 58 | right(hd) | mataqu | 36 |
| 29 | what | sapa | 66 | 59 | bone | suRi | 23 |
| 30 | who | sai | 60 | 60 | heavy | (ma)mapat | 23 |

[6] Retention rates (in percentages) for some additional stable POc etyma in the 40 language sample: *susu 'breast' 85, *taŋis 'cry' 80, *taqe 'excrement' 75, *pai, *pea 'where?' 57, *pituqun 'star' 55, *boŋi 'night '52'. Percentages for 'breast' and 'excrement', are based on samples of 34 and 22 languages, respectively, as some wordlists do not include these items.



### 7.2 Results

Retention rates for the 60 POc etyma were then calculated for each of the 40 languages in the sample plus a further dozen or so languages.[7] Table 2 shows retention rates for the NW and SE Solomonic languages in the sample.

**Table 2:** Retention rates for 60 highly stable words in some SE Solomonic and NW Solomonic languages

| | | Items retained | Percentage retained |
|---|---|---|---|
| **SE Solomonic** | **Guadalcanal-Gelic** | | |
| | Gela | 52 | 86 |
| | Lengo | 48 | 80 |
| | Ghari | 47 | 78 |
| | Talise | 45 | 75 |
| | Bugotu | 41 | 68 |
| | **Makira-Malaitan** | | |
| | Fagani | 48 | 80 |
| | Longgu | 46 | 77 |
| | Arosi | 43 | 71 |
| | 'Are'are | 41 | 68 |
| | Toqabaqita | 39 | 65 |
| **NW Solomonic** | **Nehan-Buka-N. Bougainville** | | |
| | Nehan | 31 | 52 |
| | Teop | 24 | 40 |
| | Lontes | 32 | 53 |
| | **S. Bougainville** | | |
| | Mono-Alu | 27 | 45 |
| | Torau | 34 | 57 |
| | **Choiseul** | | |
| | Babatana | 25 | 42 |
| | **New Georgia** | | |
| | Roviana | 35 | 59 |
| | Vangunu | 33 | 55 |
| | **Santa Isabel** | | |
| | Kilokaka | 27 | 45 |
| | Zabana (Kia) | 26 | 44 |

[7] It is noteworthy that Proto Central Pacific (PCP) retained all 60 of the POc items we are concerned with here. Put another way, the forms for meanings 1-60 reconstructed by comparing just Fijian, Rotuman and Polynesian are the same as those reconstructed by comparing the full range of Oceanic groups. For two POc etyma, *katoluR 'egg' and *sauq 'far', there are reflexes only in Rotuman, not in Polynesian or Fijian. I do not suggest that such a high level of retention would hold for the PCP lexicon as a whole, but this is evidence that the early Oceanic language(s) that reached the Central Pacific region had changed rather little from POc itself. It indicates that the interval between the breakup of POc and the breakup of PCP was at most a few centuries. In PPn six items have been replaced: *draRaq 'blood > *toto, *qajan 'name' > *hiŋoa, *maya 'tongue' > *qalelo, *katoluR 'egg' > *fua, *lalom 'inside' > *loto.

Among SES Solomonic languages the average percentage of retentions is 73, the highest being 86 (Gela) and the lowest 65 (Toqabaqita). Bugotu scores much lower (68) than other GG languages. This is mainly because it has borrowed some basic lexical items from Santa Isabel neighbours which show high replacement rates. It is also noteworthy that the languages of Makira (represented here by Arosi and Fagani) are in general somewhat more conservative than the Malaitan languages (represented by 'Are'are and Toqabaqita).

Among NW Solomonic languages the average percentage of retentions is 49, the highest being 59 (Roviana) and the lowest 40 (Teop).

It can be seen that all the NW Solomonic languages have replaced more of the POc basic lexicon than any of the SE Solomonic languages. However, there is considerable variation within each group and the most conservative NW Solomonic language, Roviana (59 per cent) scores only a few per cent less than the most innovative SE Solomonic language, Toqabaqita (65).

### 7.3 Determining when lexical replacements occurred in SE Solomonic and NW Solomonic

It is clear that NW Solomonic languages have replaced much more basic vocabulary than SE Solomonic. But can we determine when the changes occurred? To answer this question it is necessary to reconstruct particular interstages (intermediate protolanguages) in order to see which items were replaced between earlier and later stages. This has been done for some interstages.

#### 7.3.1 Lexical changes in Proto SE Solomonic and Proto NW Solomonic

The proto-languages of the SE Solomonic and NW Solomonic groups were both lexically quite conservative. Proto SE Solomonic (PSES) replaced just three of POc items 1-60: *draRaq 'blood' > *kabu; *matakut 'be afraid' > *matolo; *laŋit 'sky' was replaced, probably by *masawa(ŋ).[8] (In POc the primary sense of *masawa(ŋ) was apparently 'the open sea, far from land', with a secondary sense 'vast open space(s)'.) Proto NW Solomonic (PNWS) replaced just four of items 1–60: *draun 'leaf', *api 'fire', *papine 'woman' (retained only in the sense of 'man's sister') and *waiR 'water'.[9]

It is noteworthy that no replacements of POc reconstructions for items 1–60 are shared by PSES and PNWS. This is strong evidence that the two protolanguages had independent histories after they diverged at the level of Proto Nuclear Oceanic. However, in later times some borrowing occurred between certain neighbouring languages across the NWS/SES boundary, and this occurred even in a few items of basic vocabulary.[10]

[8] The same replacement, *laŋit > masawa(ŋ), is found also in some Vanuatu languages.

[9] For each of the four items replaced it is hard to reconstruct a Proto NW Solomonic etymon because the replacements differ across subgroups.

[10] Evidence for borrowing between Guadalcan-Gelic languages and the nearer NW Solomonic languages is suggested by the following comparisons, among others. In most S. Isabel/Guadalcanal-Gelic languages, POc *talinŋa 'ear' is replaced by the type of Gela *kuli* and *pisiko 'flesh' by the type of Gela *vinasi*, Poc *tam$^{w}$ata and *tau 'person' are replaced by the type of Gela *tinoni* in most New Georgia/SE Solomonic languages. See Blust (2007:411–412) for a fuller list.



#### 7.3.2 Lexical change in subgroups of SE Solomonic

The proto-languages of the major subgroups of SE Solomonic remained lexically conservative. In addition to the three replaced in Proto SE Solomonic, Proto Makira-Malaitan replaced four to five items: *draun 'leaf' > *ʔa[f,p]a, ʔapa [Toqabaqita has *rau* 'leaf, leaflet']; *api 'fire' > *kiu or *[d,t]una; *mataqu 'right hand' > *matolo or *katolo; *pulu 'feather' > *(wara)ifu. POc *mapat 'heavy is lost but a reflex of POc *(b,p)ita 'heavy' is retained in a few languages.

Besides the three items replaced in Proto SE Solomonic, Proto Guadalcanal-Gelic replaced four items: *talinŋa 'ear' > *kuli; *maya 'tongue' > *lapi, *api 'fire' > *lake, *waiR 'water' > *kolo. In addition, Proto GG lost *wakaR, the most general term for 'root' but retained POc *lamut 'root', a term that probably referred specifically to fibrous roots and root hairs.

The contemporary languages in both SE Solomonic groups, as we saw in §7.2, also remain lexically more conservative than NW Solomonic languages. This relatively small number of lexical replacements strongly suggests that neither Proto GG nor Proto MkMl nor their descendants were much influenced by contact with non-Austronesian languages. Evidently at the time Oceanic speakers arrived in the Southeast Solomons non-Austronesian speaking populations in this region were small and were easily absorbed or displaced.

#### 7.3.3 Lexical change in subgroups of NW Solomonic

Once speakers of early varieties of NW Solomonic dispersed across the NWS region each local variety underwent rapid lexical change. Thus, of the 56 POc items retained by Proto NW Solomonic in the 60 item list, Proto Choiseul, as we reconstruct it, retains only 30. That is to say, in the period between Proto NW Solomonic and Proto Choiseul almost half of the highly stable lexicon was replaced, Proto S. Isabel retains 36/56, having replaced more than a third. Proto New Georgia retains 47/56 but this is still a loss of almost 20 percent. I have not calculated percentages for the other NWS subgroups.

This very high rate of replacement in the most stable part of the lexicon indicates extensive borrowing from non-Austronesian sources. A reasonable inference is that in each of these regions the speakers of incoming NW Solomonic languages encountered substantial populations of non-Austronesian languages and that sustained bilingualism, especially in Choiseul and Santa Isabel but also in the New Georgia group, led to many non-Austronesian loanwords entering the basic vocabulary of the NW Solomonic languages. It remains to be seen to what extent putative borrowings from non-Austronesian sources can be associated with particular surviving non-Austronesian languages of the Solomons group.

### 7.4 Comparison with other Oceanic languages

Comparison of replacement rates in Oceanic languages spoken outside of the Solomons Archipelago reveals a pattern consistent with the hypothesis that higher rates correlate with more intensive contact between Oceanic and non-Oceanic languages. Table 3 gives retention rates for the 60 most stable items in a sample of languages from Polynesia, Fiji, Micronesia and Vanuatu. All are spoken on islands in Remote Oceania and probably had no direct contact with non-Austronesian languages after these islands were settled.

**Table 3:** Retention rates for the 60 most persistent words items in some languages of Remote Oceania

| | | Items retained | Percentage retained |
|---|---|---|---|
| **Polynesian** | | | |
| | Tikopia | 53 | 88 |
| | Tongan | 51 | 85 |
| | Samoan | 50 | 83 |
| | Niuean | 50 | 83 |
| | Maori | 48 | 80 |
| **Fijian** | | | |
| | Bauan (E Fijian) | 49 | 81 |
| | Wayan (W Fijian) | 49 | 81 |
| **Central and Northern Vanuatu** | | | |
| | Raga (Pentecost Is.) | 47 | 78 |
| | Nguna (Efate) | 45 | 75 |
| **Southern Vanuatu** | | | |
| | Erromangan | 37 | 61 |
| **Nuclear Micronesian** | | | |
| | Woleai | 43 | 71 |
| | Marshallese | 39 | 65 |

The range of retention rates in these particular languages is similar to that found in SE Solomonic. All have retained more of the POc basic lexicon than any of the NW Solomonic languages.

Next is a set of languages also spoken in Near Oceania which, at certain periods in their history, are likely to have had sustained contact with non-Austronesian languages. It can be seen that scores for these languages fall within the range of the NW Solomonic languages.

**Table 4:** Retention rates for some languages of Near Oceania likely to have had fairly high contact with non-Austronesian languages

| | Items retained | Percentage retained |
|---|---|---|
| **North New Guinea subgroup** | | |
| Takia | 29 | 48 |
| Sengseng | 19 | 31 |
| **Southern New Ireland subgroup** | | |
| Kuanua | 30 | 50 |

## 8 How long have SE Solomonic languages been in the Solomons archipelago?

Let us now return to the hypothesis (Sheppard and Walter 2006, Spriggs 1997) that when the Oceanic-speaking Lapita people first colonized Remote Oceania just over three millennia ago they bypassed the Solomons Archipelago, and that it was another 800 years or so before speakers of Oceanic languages established permanent settlements in the Guadalcana-Malita-Makira region. The archaeological evidence bearing on this proposal is equivocal, as was noted in §2.



Historical linguistics could throw light on this matter if a way could be found of dating the nodes on the SE Solomonic branch of the Oceanic family tree. The chief absolute dating method developed in linguistics is the much-maligned '(lexicostatistical) glottochronology', which uses cognate percentages in basic lexicon to date the length of time since particular related languages diverged.[11] In the foundation research on glottochronology the mean replacement rate for items in the 200 list was initially calculated to be about 19.5% per millennium. Rounding this to 20% yields the following predictions for a single language: 80% of the original 200 items will be retained after 1000 years, 64% after 2000, 51% after 3000, 41% after 4000. When estimating separation dates from cognate percentages between contemporary languages, the equations based on 20% replacement per millennium are: 64% cognates = 1000 years separation, 41% = 2000, 28% = 3000, 17% = 4000.

In the case of Austronesian languages, these estimates can be tested against an independent chronology that can be established for particular intermediate proto-languages (the ancestors of particular subgroups) by correlating linguistic and archaeological events. Austronesianists have a valuable external point of reference when estimating the dates at which particular subgroups broke up, namely, several cases where archaeological dates for the settlement of a particular can, with high confidence, be matched with the arrival of a particular language in that region, a language ancestral to a large subgroup. Thus, one can date the breakup of Proto Malayo-Polynesian to about 4000 BP, because the emergence of the Malayo-Polynesian branch of Austronesian can be connected with the movement of people from Taiwan across the Bashi Channel into the Batanes Is. and Luzon at about that time (Bellwood pers. comm.; Bellwood and Dizon 2005; Ross 2005). The breakup of POc can be placed at between 3400 and 3100 BP (see §3). We can be confident that the Central Pacific languages (Fijian, Rotuman and Polynesian) diverged from both the NW Solomonic and SE Solomonic groups no later than about 3000 years ago. This is because the foundation settlement of Fiji and Tonga is rather securely dated to about 3050–2950 years ago. An earliest possible date for the split is that assigned to the breakup of POc itself.

Although it has been shown that Malayo-Polynesian languages vary greatly in their retention rates (Blust 1981, 1999), there is reason to think that the standard glottochronological estimates are about right for lexically conservative Oceanic languages. Assuming that Proto Malayo-Polynesian broke up about 4000 BP, we get results close to the mark for the most conservative Oceanic languages, such as Gela, Samoan and Fijian. Each is known to retain about 40% of the reconstructed Proto Malayo-Polynesian items for 200 item basic lexicon. And although the calculations have not been done for the full range of languages, we can be reasonably sure that quite similar results will be obtained for almost all the SE Solomonic languages, all the Fijian languages and many of the Polynesian languages.

Given this method, it is possible to assign approximate dates to the breakup of Proto SE Solomonic and its daughter subgroups, Guadalcanal–Gelic and Makira-Malaitan. The following account of lexical diversity exhibited by languages in the SE Solomons and neighbouring areas draws on the percentages given in Tryon and Hackman (1983) for the Swadesh 200 item basic lexicon.

[11] Russell Gray and his associates have in recent years been developing an alternative dating method (Gray 2005; Gray and Atkinson 2003; Greenhill and Gray this volume).

Let us first consider how SE Solomonic languages score with other Oceanic languages that are known to be fairly conservative and compare these agreements with those between Guadalcanal-Gelic and Makira-Malaitan languages.[12] Recall that the split between SE Solomonic and Polynesian is dated to no later than 3000 years ago. Cognate percentages between SE Solomonic languages and five Polynesian Outlier languages in the Solomons region (Rennellese, Tikopia, Sikaiana, Luangiua and Pileni) fall between 25 and 36, with a median of 29.[13] Percentages between Guadalcanal-Gelic and Makira-Malaitan languages fall between 28 and 43, with a median of 36.

The differences between the SE Solomonic-Polynesian agreements and the agreement between Guadalcanal-Gelic and Makira-Malaitan are thus on the order of 7 percent. This is consistent with about 500 years elapsing between the SE Solomonic-Polynesian split, and the breakup of SE Solomonic into incipient Guadalcanal-Gelic and Makira-Malaitan branches.

Next let us consider agreements within the Makira-Malaitan group. The Makira–Malaitan languages are clearly descended from a dialect chain that extended over most of the Makira-Malaitan region. Today the lexical diversity of languages from opposite ends of this region is almost as great as the divergence between Makira-Malaitan and Guadalcanal-Gelic. The most differentiated Makira-Malaitan languages show percentages in the 34–40% range, e.g. Toqabaqita of N. Malaita has the following percentages with Makira languages: 34 with Santa Ana, 35 with Kahua and Bauro, 40 with Arosi. These are about the same as Toqabaqita shares with Guadalcanal-Gelic (32–36%). All this suggests that the opposite ends of the Proto Makira–Malaitan region began to diverge into dialects soon after Makira–Malaitan split off from Guadalcanal-Gelic but that the divergence proceeded gradually because the central dialects of Makira-Malaitan remained in close contact with the extremes.

Guadalcanal-Gelic is more homogeneous than Makira-Malaitan. Excluding Bugotu, the most differentiated GG languages show cognate percentages in the range 50–55% and some pairs of languages score 60–70%. This strongly suggests that the ancestral GG dialect chain remained fairly cohesive for much longer than Makira-Malaitan, with most dialects remaining mutually intelligible until about 1000 years ago. Table 5 gives approximate divergence dates for pairs of groups based on the median percentage, using the standard glottochronological equations.

[12] Excluded from the intra-SE Solomonic comparisons are Marau and Longgu, two MkMl languages spoken on Guadalcanal, whose percentages are inflated by loans from GG neighbours. Also excluded is one GG language, Bugotu, whose percentages with MkMl and with other GG languages are much lower owing to sustained contact with Santa Isabel languages. Bugotu's agreements with MkMl are in the range 26–32%, i.e. almost 10% lower than other GG languages.

[13] For example, the lexically most conservative GG language, Gela, scores 31–36% with Polynesian Outliers. It scores just a bit higher, 34–43%, with MkMl languages. Its sister language Ghari scores 28–32% with Polynesian Outliers, compared to 33–40% with MkMl languages. The most conservative MkMl language, Fagani (of Makira), scores 28–33% with Polynesian Outliers compared to 36–43% with GG. The least conservative MkMl language, Toqabaqita (of Malaita), scores 25–27% with Polynesian Outliers, compared with 32–36% with GG. The most conservative MkMl language, Fagani (of Makira), scores 28–33% with Polynesian Outliers compared to 36–43% with GG.



**Table 5:** Cognate percentages for inter-group comparisons with approx. divergence dates

| | percentages | median | approx. divergence date for median |
|---|---|---|---|
| SES-Polynesian | 25–36 | 29 | 2900 BP |
| MkMl-GG | 28–43 | 36 | 2400 BP |
| extremes of MkMl | 34–40 | 37 | 2300 BP |

These figures do not, of course, tell us how long the ancestral SE Solomonic language was in the SE Solomons before it diverged into GG and MkMl. However, it is reasonable to assume that the innovations defining SES were accumulated over a few centuries when pre-SES was a single language—no doubt with dialect variants—spoken in a string of mainly coastal and small island settlements in parts of Makira, Malaita and Guadalcanal.[14]

But where was pre SE Solomonic spoken before it was carried to the SE Solomons? Does this group fall into a subgroup with any other branch of Nuclear Oceanic?

From time to time it has been argued that SES falls into an Eastern Oceanic group together with most or all of the Oceanic languages of Remote Oceania, especially those of Vanuatu, New Caledonia and the Loyalties, Fiji, Polynesia and possibly Micronesia. There are a few scraps of evidence supporting such a group but the hypothesis remains highly problematic and this is not the place to review the evidence.[15]

## 9 Conclusions

We are led to the following conclusions.

1. The sharp boundary between NW and SE Solomonic is not the product of *in situ* divergence. The NW and SE Solomons regions were settled independently by two different populations of Oceanic speakers.

2. The position of the NW Solomonic languages on the Oceanic family tree is consistent with Sheppard and Walter's proposal that that the NW Solomons was bypassed in the initial movement of Lapita people into Remote Oceania. NW Solomonic is a division of the Meso-Melanesian branch of Oceanic. The centre of diversity within Meso-Melanesian, and its original site is clearly in the New Britain-New Ireland area. At some point speakers of a Meso-Melanesian language moved to the Nissan-Buka-N. Bougainville region. There the language developed the few innovations that define the NW Solomonic subgroup. After a short period of unified development Proto NW Solomonic spread to the Shortlands, Choiseul, New Georgia and Santa Isabel. Linguistic methods do not allow us to date precisely the spread of NW Solomonic. However, it is clear, from the archaeological

[14] Recently Lynch et al. (2002:110ff.) have suggested that Proto SE Solomonic was confined to the Bugotu-Gela-North Guadalcanal region and that its descendants later moved from Guadalcanal into Makira and Malaita. However, this scenario rests on a very flimsy argument.

[15] Re the Eastern Oceanic hypothesis, see Grace (1976), Pawley (1972), Lynch and Tryon (1985). The Oceanic languages of the Eastern Outer Islands region are not known to share any innovations with Southeast Solomonic. Although their histories are still poorly understood it seems likely that the better known languages of Utupua and Vanikoro form a first-order subgroup of Nuclear Oceanic, to which Aiwoo of the Reef Is. may also belong (Ross and Næss 2007). In that case, they are likely to be relics of the first Lapita movement into Remote Oceania. All this does not rule out Greater Reefs/Santa Cruz as a source for pre-SE Solomonic. It simply implies that if it was, pre-SE Solomonic speakers left Reefs/Santa Cruz quite soon after Oceanic speakers first arrived there.

evidence, that the breakup of Proto Oceanic must have occurred between about 3350 BP, by which time Lapita settlements had been established in various parts of the Bismarck Archipelago, and 3100 BP, by which time Lapita settlements had been established in Remote Oceania. The innovations marking off Meso-Melanesian from the rest of Oceanic, and those marking off NW Solomonic from the rest of Meso-Meso-Melanesian are relatively few, and in all, probably took no more than three or four centuries to accumulate. This estimate would place the breakup of NW Solomonic as occurring between about 3000 and 2700 BP.

3. Subsequently, in the course of dispersing across the NW Solomons, the ancestral NW Solomonic language developed regional variants that underwent very rapid lexical change. Many words not known to have Austronesian antecedents entered their core lexicons. A reasonable explanation is that in each locality small populations of immigrant Oceanic speakers came into contact with established populations of non-Austronesian speakers, leading to extensive intermarriage, bilingualism and lexical borrowing from non-Austronesian languages.

4. Over the next couple of millennia Austronesian languages replaced non-Austronesian languages over most of the NW Solomons. An exception is Bougainville, where non-Austronesian languages remain dominant over most of the island.

5. The scenario sketched in 2–4 above does not preclude the possibility that speakers of NW Solomonic were not the first speakers of an Oceanic language to settle in the NW Solomons. However, if there were earlier Oceanic-speaking colonists, they left no surviving daughter languages in the region. This fact suggests that, at best, any earlier Oceanic-speaking populations must have been small.

6. The SE Solomonic languages show few signs of influence from non-Austronesian languages, an indication that the pre-Austronesian populations were sparse in the SE Solomons. However, non-Austronesian languages survive on two small islands near Guadalacanal: Savosavo and Russell.

7. The linguistic evidence weighs strongly against Sheppard and Walter's suggestion that the islands from Guadalcanal to Makira were not settled until around 2300–2200 BP, around the time when the making of ceramics had ceased in the Reefs/Santa Cruz area. Southeast Solomonic is a fairly well defined subgroup of Oceanic, without obvious close relatives elsewhere and it must have separated from the language ancestral to the Fijian and Polynesian groups no later than 3000 BP. The set of phonological, morphological and lexical innovations that define Southeast Solomonic indicate several centuries of unified development in the Southeast Solomons region. The internal diversity of Southeast Solomonic is also considerable. In comparisons of a 200 item basic lexicon the two primary subgroups of SE Solomonic (Guadalcanal-Gelic and Makira-Malaitan) diverge from each other almost as sharply as they diverge from Fijian and Polynesian. This degree of difference points to the two subgroups as having followed separate paths since about the middle of the first millennium BC. Furthermore, the languages at opposite ends of the Makira-Malaitan subgroup differ from each other, lexically, almost as sharply as they do from Guadalcanal-Gelic languages, indicating that internal differentiation within Makira-Malaitan began around the same time (although the rate was slowed by the persistence of a dialect chain). I conclude that the SE Solomonic languages have been present in Makira, Malaita and Guadalcanal for well over 2500 years and probably for around 3000 years.



8. It is uncertain where the immediate ancestor of SE Solomonic came from. There is no decisive evidence to subgroup SE Solomonic with any other branch of Nuclear Oceanic. On archaeological grounds an immediate origin from the east, from the Eastern Outer Islands of the Solomons, or from Vanuatu, is perhaps more likely than direct settlement directly from the Bismarcks. Over the years a number of linguists have pointed to scraps of evidence suggesting a brief shared history with certain other languages of Remote Oceania, especially those of Vanuatu, Fiji, Polynesia and Micronesia but the evidence is far from decisive.

9. If SE Solomonic speakers dispersed over the coasts and offshore islands of Makira, Malaita and Guadalcanal in the first half of the 1st millennium BC one may ask why did they not also settle the nearest parts of the Western and Central Solomons, such as Santa Isabel and New Georgia. I think a good part of the answer is that at that time the latter islands were populated exclusively, or almost exclusively by non-Austronesian speakers and that they remained largely non-Austronesian speaking for many centuries after that. In Santa Isabel and New Georgia, as well as on the small islands of Russell and Savo, non-Austronesian speaking areas for a time formed a buffer between NW Solomonic and SES Solomonic languages. However, once the two subgroups came into contact there was a good deal of borrowing between the languages closest to each other.

## References


Bedford, Stuart. 2003. The timing and nature of Lapita colonization in Vanuatu: the haze begins to clear. In Sand, ed. 147–158.

Bedford, Stuart, Matthew Spriggs and Ralph Reggenvanu. 2006. The Teouma site and the early human settlement of the Pacific Islands. *Antiquity* 80:812–828.

Bedford, Stuart, Shaun Connaughton and G. Clark, eds. 2007. Oceanic explorations: Lapita and western Pacific settlement. Terra Australis 26. Canberra: ANU E–Press.

Bellwood, Peter. 1997. *Prehistory of the Indo–Malaysian Archipelago.* (Revised edition.) Honolulu: University of Hawaii Press.

Bellwood, Peter and Eusebio Dixon. 2005. The Batanes archaeological project and the 'out of Taiwan' hypothesis for Austronesian dispersal. *J. of Austronesian Studies* 1(1):1–32.

Blust, Robert A. 1978a. Eastern Malayo-Polynesian: a subgrouping argument. In S.A. Wurm and L. Carrington, eds *Second international conference on Austronesian linguistics: Proceedings*, 181–234. Canberra: Pacific Linguistics.

—— 1978b. *The Proto-Oceanic palatals.* Polynesian Society monograph No.43. Auckland: Polynesian Society.

—— 1981. Variation in retention rate among Austronesian languages. Paper presented at 3rd International Conference on Austronesian Linguistics, Bali, January 1981.

—— 1995. Austronesian comparative dictionary. Computer file. Department of Linguistics, University of Hawaii.

—— 1998. A note on higher-order subgroups in Oceanic. *Oceanic Linguistics* 37:182–188.

—— 1999. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. In Colin Renfrew, A. McMahon and L. Trask, eds *Time depth in historical linguistics*, 311–332. Cambridge: MacDonald Institute for Archaeological Research.

—— 2005. Review of Lynch, Ross and Crowley 2002. *Oceanic Linguistics* 34(1):191–202.

—— 2007. Proto-Oceanic *mana revisited. *Oceanic Linguistics* 46:404–423.

—— 2008. Remote Oceanic: one history to two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47:433–444.

Burley, David and S.P. Connaughton. 2007. First Lapita settlement and its chronology in Vava'u, Kingdom of Tonga: New data on old problems in the kingdom of Tonga. *Radiocarbon* 49(1):131–137.

Donohue, Mark and tim Denham. 2008. The language of Lapita: Vanuatu and an early Papuan presence in the Pacific. *Oceanic Linguistics* 47:433–444.

Dunn, Michael, Ger Reesink and Angela Terrill. 2002. The East Papuan languages: a preliminary typological appraisal. *Oceanic Linguistics* 41:28–62.

Dunn, Michael, A. Terrill, G. Reesink, R. Foley, S. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 22 September 2005, vol. 309:2072–2075.

Dutton, T.E. and D.T. Tryon, eds. 1994. *Language change and contact in the Austronesian world.* Berlin: Mouton de Gruyter.

Dyen I., A.T. James and J.W.L. Cole. 1967. Language divergence and estimated retention rate. *Language* 43(1):150–171.

Felgate, Mathew. 2001. A Roviana ceramic sequence and the prehistory of Near Oceania: work in progress. In G.R. Clark, A.J. Anderson and T. Vunidoilo, eds *The archaeology of Lapita dispersal in Oceania. Papers from the 4th Lapita conference, June 2000, Canberra, Australia*, 39–60. Terra Australis 17. Canberra: Pandanus Books.

—— 2003. Reading Lapita in Near–Oceania: intertidal and shallow–water pottery scatters, Roviana Lagoon, New Georgia, Solomon Islands. PhD thesis, University of Auckland.

—— 2007. Leap–frogging or limping? Recent evidence from the Lapita littoral fringe, New Georgia, Solomon Islands. In S. Bedford, S. Connaughton and G. Clark, eds, 121–140.

Grace, George W. 1959. *The position of the Polynesian languages within the Austronesian (Malayo–Polynesian) language family.* Bloomington: International Journal of American Linguistics, Memoir 16.

—— 1966. Austronesian lexicostatistical classification: a review article. *Oceanic Linguistics*, 5(1):13–31.

—— 1976. Review of R.C. Green and Marion Kelly, eds, *Studies in Oceanic culture history*, vol. 3. *J. Polynesian Society* 85(1):103–112.

Green, Roger C. 1991. A reappraisal of the dating of some Lapita sites in the Reef/Santa Cruz Group of the Southeast Solomon Islands. *J. Polynesian Society* 100:197–207.



—— 2003. The Lapita horizon and traditions — signature for one set of Oceanic migrations. In Sand, ed. 95–120.

Haberle, Simon. 1996. Explanations for palaeoecological change on the northern plains of Guadalcanal, Solomon Islands: the last 3200 years. *The Holocene* 6:333–338.

Kirch, Patrick V. 2000. *On the road of the winds: an archaeological history of the Pacific Islands before European contact.* Berkeley: University of California Press.

Lichtenberk, Frantisek. 1988. The Cristobal–Malaitan subgroup of Southeast Solomonic. *Oceanic Linguistics* 27:24–62.

—— 1994. Reconstructing heterogeneity. *Oceanic Linguistics* 33:1–36.

Lincoln, Peter. 1978. Reef-Santa Cruz as Austronesian. In Stephen Wurm and Lois Carrington, eds *Second international conference on Austronesian linguistics: Proceedings*, 929–967. Canberra: Pacific Linguistics.

Lynch, John, Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. Richmond, Surrey: Curzon.

Milke, Wilhelm. 1958. Zur inneren Gliederung und geschichtlichen Stellung der ozeanisch-austronesischen Sprachen. *Zeitschrift für Ethnologie* 83:58–62.

Nunn, Patrick D., Roselyn Kumar, Sepeti Matararaba and 10 others. 2004. Early Lapita settlement at Bourewa, southwest Viti Levu. *Archaeology in Oceania* 39:139–143.

Pawley, Andrew. 1972. On the internal relationships of Eastern Oceanic languages. In Roger C. Green and M. Kelly, eds *Studies in Oceanic culture history*, vol. 3. *Pacific Anthropological Records* No.13, 1–142. Honolulu: Bishop Museum.

—— 2003a. Locating Proto Oceanic. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic. The culture and environment of Ancestral Oceanic Society:* vol. 2. *The physical environment,* 17–34. Canberra: Pacific Linguistics.

—— 2003b. The Austronesian dispersal: languages, technologies, peoples. In P. Bellwood and C. Renfrew, eds *Examining the farming/language dispersals hypothesis*, 251–273. Cambridge: McDonald Institute of Archaeological Research, Cambridge University.

—— 2006. Explaining the aberrant Austronesian languages of Southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 116(3):213–256.

—— 2007.The origins of early Lapita culture: the testimony of historical linguistics. In Bedford, Sand and Connaughton, eds, 17–49.

—— 2008. Where and when was proto Oceanic spoken? Linguistic and archaeological evidence. In Yury Lander and Alexander Ogloblin, eds *Language and text in the Austronesian world. Studies in honour of Ulo Sirk*, 47–71. Münich: Lingcom Europa.

Roe, David. 1993. Prehistory without pots: Prehistoric settlement and economy in northwest Guadalcanal, Solomon Islands. PhD thesis, The Australian National University, Canberra.

Ross, Malcolm. 1986. Towards a classification of the Oceanic languages of Bougainville and the western Solomons. In P. Geraghty, L. Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics,* 175–200. Canberra: Pacific Linguistics.

—— 1988. *Proto Oceanic and the Oceanic languages of western Melanesia.* Canberra: Pacific Linguistics.

—— 2001. Is there an East Papuan phylum? Evidence from pronouns. In A. Pawley, M. Ross and D. Tryon, eds *The boy from Bundaberg: essays in Melanesian linguistics in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

—— 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *J. Austronesian Studies* 1(2):1–23.

Ross, Malcolm and Ashild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46.

Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds. 1998–2008. *The lexicon of Proto Oceanic. The culture and environment of ancestral Oceanic society*: vol. 1 *Material culture.* vol. 2, *The physical environment.* vol. 3. *Plants.* Canberra: Pacific Linguistics.

Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds. In prep. *The lexicon of Proto Oceanic. The culture and environment of ancestral Oceanic society*: vol. 4. *Animals.* vol. 5. *People and society.* Canberra: Pacific Linguistics.

Sand, Christophe. 2001. Evolutions in the Lapita Cultural Complex: a view from the Southern Lapita Province. *Archaeology in Oceania* 36:65–76.

—— ed. 2003. *Pacific archaeology: Assessments and prospects. Proceedings of the international conference for the 50th anniversary of the first Lapita conference (July 1952).* Les cahiers de l'archéologie en Nouvelle-Caledonie, vol. 15. Noumea: New Caledonia Museum.

Sheppard, Peter and Richard Walter. 2006. A revised model of Solomon Islands culture history. *J. Polynesian Society* 115(1):47–76.

Specht, Jim. 2005. Revisiting the Bismarcks: some alternative views. In A. Pawley, R. Attenborough, J. Golson and R. Hide, eds *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 255–288. Canberra: Pacific Linguistics.

Spriggs, Matthew. 1997. *The Island Melanesians.* Blackwell: Oxford and Cambridge, Mass.

Summerhayes, Glenn. 2000. *Lapita interaction.* Terra Australis 15. Canberra: Archaeology and Natural History Publications and the Centre for Archaeological Research, Australian National University.

—— 2001. Lapita in the far west: recent developments. *Archaeology in Oceania* 36:53–63.

Tryon, Darrell and Brian Hackman. 1983. *Solomon Islands languages: an internal classification.* Canberra: Pacific Linguistics.

Wickler, S.H. 2001. *The prehistory of Buka, a stepping stone island in the Northern Solomons.* Terra Australis 14. Canberra: Dept of Archaeology and Natural History, Research School of Pacific and Asian Studies, The Australian National University.