Computational reconstruction of grammar in Oceanic proto-languages

Hedvig Skirgård

July 8, 2021

Contents

This is a first draft based on the chapter text. I've modified it, and I'm in the process of modifying further. Most importantly, I need to incorporate the results which derive from SCM and the whole posterior as opposed to only Max Parsimony and Max likelihood on the MCCT tree of gray et al 2009.

Abstract

Historical linguists have mainly been concerned with regular sound correspondences and the reconstruction of vocabulary, but there are also insights to be learned from reconstruction of grammar. In this paper, we reconstruct structural features of proto-languages in the Oceanic subgroup using computational methods (Maximum Parsimony, Maximum Likelihood and Stochastic Character Mapping). The computational reconstructions are compared to reconstructions derived from traditional historical linguistics and are used to measure conservatism of Oceanic languages. The results show that classical historical linguistics in this dataset is mainly an application of Maximum Parsimony, and we further discuss what this implies for the methodologies. Contrary to measurements of lexical conservatism (Blust 1981), languages in the Central Pacific subgroup are among the most innovative languages structurally. This result is discussed in relation to the social dynamics of the region and the implications of vocabulary being more salient as identity marking.

1 Introduction

Have some languages changed more than others? Are some grammatical features more stable than others? In this study, we use computational phylogenetic methods of reconstructing proto-languages of the Oceanic subgroup and measure the stability of structural features and conservatism of languages.

The aims of this study are threefold:

1. Do computational methods reconstruct the same structural features of Oceanic proto-languages as classical historical linguistics does?

2. Are some structural features more stable than others?

3. Which languages within the Oceanic subgroup have changed the most structurally?

The data for the study is taken from the Grambank-project (The Grambank Consortium 2021). The Grambank dataset consists of 195 structural features which have been coded by a large group of research assistants for over 2,000 languages. This dataset includes 234 Oceanic languages.

The computational ancestral state reconstruction is carried out using Maximum Parsimony (Sankoff 1975; Louca & Doebeli 2017) and Maximum Likelihood (Fisher 1912; Wilks 1938; Pagel 1994; Cunningham et al. 1998; Jäger & List 2018). The tree which is used in the analysis is the Maximum Clade Credibility Tree (MCCT) from Gray et al. (2009).

The Oceanic subgroup is well-studied in historical linguistics, in particular its lexicon (see the book series on the Proto-Oceanic lexicon (Ross et al. 1998, 2007, 2008, 2011, 2016), among other publications). There has also been considerable work done on reconstructing the grammar of proto-languages, in particular Proto-Polynesian. In this chapter, we test if computational methods of reconstructing structural features of proto-languages come to the same conclusions as historical linguists. We will examine the issue of case alignment in proto-Polynesian, a contested issue in Oceanic historical linguistics, specifically.

The tools of computational reconstructions are different from classical historical linguistics, and the data used in this chapter, the Grambank dataset, is different from the source material that historical linguists work with. This is further discussed in section 2.0.1

A major question in studies of language change is the stability over time of particular features. As Ross (2007:281) notes, stability matters:

> *Students of historical linguistics are often so concerned with language change that they neglect what remains stable. [...] Furthermore, there is often a consistency across languages with regard to which distinctions are lost or retained, indicating that something other than mere chance is at work.*

Ross (2007:281)

Change in language is not constant. This is one of the critiques of lexicostatistics (Blust 2000), since it assumes an even rate of change. In order to further explore language history and the dynamics of structure (as opposed to lexicon and phonology), it is necessary to examine the stability of structural features by different methods. In this chapter I will use computational phylogenetic methods to calculate the stability of 201 grammatical features on the Oceanic subgroup of languages.

There are studies showing that certain structural features are able to reveal connections between languages in deep history (Nichols 1998:143) (c.f also Evans (2019)). As Ross (2007) stated in the previous quote, certain morpho-syntactic features have been observed to be particularly stable. Ross (2004:503) notes that a particular structure of the pronominal system of Mokilese is maintained, despite the formal markers being continuously replaced. He argues that there are discourse related reasons for maintaining this system and that the interaction between this construction and the rest of the grammar is such that the distinction is maintained. When particular markers are lost in this system, new ones appear in their place. He also notes that Goddard has observed similar patterns in Algonquian languages (Goddard 1993) .

If this is the case, we may expect that features that are crucial to the organisation of the paradigm in a language are more stable than others. This would cover many of what we have been labelling as **distinction**-type variables in section **??**.

Nichols (1998:143) notices similar phenomena and proposes that structural features may be able to trace history of contact areas further back than language families:

> *[T]he relative frequencies of diagnostic structural features in large scale areally-based sets of languages can reveal fundamental affinities between some of these populations, and that these in turn point to shared geographical origins. This approach bypasses descent entirely and instead traces nongenealogical affinities between large geographically-based groupings of language families. It cannot trace the origins of individual families very well, but it can trace the settlement of continents, explain the worldwide geographic distribution of language families, and reach very far back into prehistory.*

(Nichols 1998:143)

Dediu & Cysouw (2013) conducted a meta-study which compared seven different methods of calculating the stability of structural features using data from the World Atlas of Language Structures (Dryer & Haspelmath 2013). They found that, overall, the different approaches concurred as to which features were most stable — which is encouraging. The list of the most stable features was dominated by word-order and rare features (e.g. Optative Mood or Absence of Common Consonants).

However, Thomason & Kaufman (1992), Ross (1996), and Greenhill et al. (2017) argue that there are problematic characteristics of structural data compared to lexical. In their study, (Greenhill et al. 2017) found that contrary to common perception, many grammatical features have faster rates of change than basic vocabulary. In the previous chapter (section **??**) we outlined key characteristics which make structural data different from lexical: a) smaller design space, b) cognacy necessarily implies shared descent, structural similarity does not, c) functional dependencies in structural data and d) different evolutionary constraints (cognition, complexity, contact etc). The results of the studies in chapter **??** indicated that design space is indeed a concern, that structural features correlate less with known families trees than basic vocabulary, and that dependencies are most likely not a problem for this dataset (section **??**). The issue of different evolutionary pressures was not investigated *per se*, but the fact that Northern Vanuatu was especially internally homogeneous structurally may support the proposal by François (2011) that those languages have converged more structurally than lexically because of social factors.

The aim of the second study of this chapter is to compare two different methods of calculating stability (Maximum Parsimony and Maximum Likelihood) with the aim of finding structural features that rate high in stability across both methods.

The third aim of this chapter is to measure the structural conservatism of the Oceanic languages — which languages have changed the most from the reconstructed proto-language? Sapir (1916:455) and Lynch et al. (2002:119) suggest that we should find the most genealogical diversity in the region where the migrations originate from. In the case of Oceanic, we know from archaeology that the spread occurred from the west to the east, with Aotearoa (New Zealand) being settled last (Rieth & Cochrane 2018). Lynch et al. (2002:119) also state that we should find the most conservative languages there. However, Blust (1981, as cited in Blust (2000:323)) found that the most conservative Oceanic languages are primary found in the Central Pacific linkage — to the east. Pawley (2009:523) found languages in Southeast Solomons with similarly high retention rates of basic vocabulary to Central Pacific.

Too little is known of rates of structural change to determine whether they are likely to behave similarly to Oceanic lexicon (more conservatism in the east) or in accordance with the default theory (more conservatism at the origin). The results of this chapter will reveal which is true (for this data, method and sample).

2 Reconstructing grammar

2.0.1 Historical linguistic methodology

In order to interpret the differences between the results of our computational reconstruction and classical historical linguistics, it is first necessary to clarify the different methodologies and what consequences they have for the study at hand. This section lays out the fundamental principles of Historical Linguistics and how they relate to this chapter.

The core method by which historical linguists reconstruct language history is known as the "Comparative Method". The Comparative Method is based on finding words or morphemes in different languages that have the same (or similar enough) meaning and that display non-trivial systematic phonological similarities. By investigating these sets of words, it is possible to deduce which are inherited from a common shared ancestor, i.e. are cognates. For example, Blust (2004), Greenhill & Clark (2011) and many others have reconstructed that Māori /toru/ (meaning 'three') derives from the same word in an ancestral language as Hawai'ian /kolu/ ('three'). These two words are "cognates" of each other. Furthermore, many words that mean the same/similar thing in Māori and Hawai'ian show this pattern of t/k (Māori: /mate/, Hawai'ian: /make/ 'to be dead' and Māori: /whitu/, Hawai'ian: /hiku/ 'seven' (Greenhill et al. 2008)). There is a systematic sound correspondence between these two sounds, and further research into more languages in this family shows that Hawai'ian /k/ is more likely to be an innovation and Māori /t/ a retention from an older Proto-language (for example, 'three' is /tulu/ in Amis, an Austronesian language of Taiwan.

Historical linguists use cognates and systematic sound correspondences to develop hypotheses about forms in unobserved Proto-languages and to propose sub-groupings based on shared innovations (c.f. how biological cladistics finds relationships between species based on shared derived characteristics from common ancestors (Maclaurin & Sterelny 2008:16-17)). This method provides us with sets of words which derive from the same word in an ancestor language (cognates), sequences of sounds changes from a Proto-language to the current observable daughter languages and a tree structure of the relationships between languages.

The Comparative Method in historical linguistics relies on knowledge of probable phonological shifts (/s/ is more likely to become /h/ than it is to become a /k/[1]) and on probable semantic shifts. In the above example from Māori and Hawai'ian, the words /toru/ and /kolu/ both mean 'three', but it is possible for cognates to have less similar meanings. For example, Pawley (2005) reconstructs **panua* as meaning 'land' or 'inhabited territory'. In daughter languages, this has changed to 'place', 'community', 'village', 'house', 'people', 'world' and 'weather'.

Reconstruction of words, phonemes and grammatical features of proto-languages is in historical linguistics guided by three principles (c.f. Clark (1976:17-22)):

(i) number of changes posited

(ii) plausibility of the reconstructed language as a human language

[1] Historical linguists do concede that there are instances of irregular sound change (Blust 1996; Campbell 1996) and that while they can often be explained by contact, analogy or avoidance of homophony, they sometimes remain unexplained.

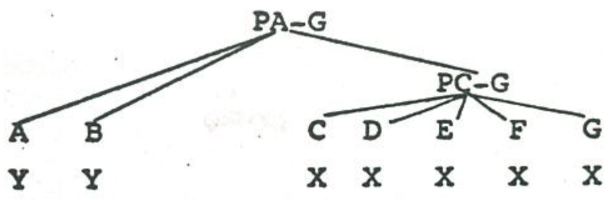


Figure 1: Tree from Clark (1976:19) illustrating Maximum Parsimony.

(iii) plausibility of the changes posited

The first of these principles is the same as what is known in phylogenetics as "Maximum Parsimony". The idea is to reconstruct states in proto-languages such that there are as few changes as possible in the entire tree. Clark (1976:17-22) explains how this works by positing an example of seven languages where there is a majority of one feature, X, and fewer of another, Y (Fig. 1). However, upon knowing the tree structure by which these languages are related and applying Maximum Parsimony, we should still reconstruct Y for the proto-node . There are three subgroups of this tree that attach directly to the root, two of which only consists of one language each and the other which has five member languages. Despite the majority of the values at the tips of the tree (the languages) being "X", the appropriate reconstruction for the root (the proto-language), according to Clark and the Maximum Parsimony approach, is "Y". This is the most parsimonious reconstruction since it would only posit one change along the tree (just before the subgroup "PC-G"). Positing "X" at the root would mean positing two changes in order to get to the known states of the tips.

It is important to note that Maximum Parsimony does not take into account branch lengths, only the changes between each node of the tree (regardless of how far apart they are). Furthermore, Maximum Parsimony makes the assumption that the slowest rate of change is the accurate one.

Reconstruction in historical linguistics also includes judgements of plausibility. This requires some assumptions about what are plausible features to co-occur in language, and which pathways of language change are more plausible than others. For example, it is rare to find a language that has a gender distinction in first person but not in third (though not impossible; c.f. Siewierska (2013)). If the most parsimonious reconstruction results in a proto-language with many rare features, it may require more investigation. Similarly, changes from certain states to others are assumed to be less plausible. For example, a language going from having no marked dual number on nouns to having trial number would be taken as unusual by most linguists (c.f. Kikusawa (2006:8)).

Plausibility is important in reconstruction, both in linguistics and in biology. However, this principle is sensitive to differing assumptions and theories. Besides debates over precise sub-groupings, many arguments in historical linguistics relate to this issue. For example, Clark (1976) disagrees with Hale (1968), Hohepa (1969), and Chung (1978) on the state of Proto-Polynesian syntax on these grounds.

Chung, Hale and Hohepa argue for a theory that is less technically parsimonious, but which they say is more plausible. They posit that Proto-Polynesian had a nominative-accusative case marking system[2]. If this was the case, given the distribution of languages that would mean positing more changes along the tree than if we assumed, as Clark (1976) does, that the Proto-Polynesian language was ergative-absolutive. Chung's plausibility critique of Clark's proposal is three-fold:

(a) the tree used is not an accurate representation of the language history (there was more interaction between Sāmoan and Tongan)

(b) it is possible that the Proto-language contained variation and was undergoing change that was only fully realised in some of the daughters[3]

(c) the morpho-syntactical historical process is less plausible

In a review of Clark (1976), Chung writes:

> *Such an approach [as Clark's] relies on the assumption that the subgroups have developed quite independently once they split off from Proto-Polynesian, so that features shared by both must be attributed to the Proto-language. But in fact, both parts of this assumption are too strong. It is well known that the two primary subgroups of Polynesian did not develop totally separately; there was long-standing contact in pre-European times between speakers of Tongic and some Samoic-Outlier languages, as Clark himself notes (p. 27). Further, and more generally, it is simply not true that every feature shared by related languages must have existed in the Proto-language uniting them. Languages are constantly undergoing change; and it is reasonable to suppose that Proto-languages were no different from real languages in this respect. But if this is so, then it is also reasonable that changes begun in a Proto-language may have continued even after its separation into daughter languages. In this way, related languages may come to share a feature which existed only in embryonic form, or not at all, in their common ancestor.*
>
> Chung (1977:539)

This debate contains more twists and turns, with each side arguing for the plausibility of their accounts. In our analysis, we will be using a tree that represents the history of the languages in a similar way to Clark, which means the results are sensitive to the same critique by Chung. We are also not able to use plausibility in our computational reconstructions since we do not have access to formalised data on what plausible language profiles or changes are. This is a key difference between computational reconstruction

[2]Hale, Hohepa and Chung actually suggest three different specific theories for this reconstruction. For a summary of the differences between the proposals, see Chung (1978:247-249).

[3]The reason that only some daughter languages exhibit the feature could be due to founder effects (my addition).

and traditional approaches to reconstruction. Knowledge of plausibility and how to weight different kinds of evidence against each other is not formalised and cannot be taken into account.

It is possible that with the added information on rates of change and branch length that comes with the Maximum Likelihood approach we are able to approximate historical linguists' knowledge of plausibility. In that case, we would expect the Maximum Likelihood results to concur more with the predictions by expert linguists. If historical linguists mainly do operate on the same principles of Maximum Parsimony, and/or Maximum Likelihood is not able to approximate plausibility, we would expect the results of Maximum Parsimony to concur more with findings in traditional historical linguistics.

The processes of subgrouping and reconstruction are done in tandem in historical linguistics. Subgroups are proposed based on shared innovations. In order to determine what is and what is not an innovation, a certain amount of reconstruction is necessary. In order to make reconstructions, some of the tree structure needs to be approximated.

Pawley (personal correspondence) notes that most of the subgrouping done in historical linguistics tends to be at the lower level. This can be seen later in this chapter in the difference between the Glottolog tree (Fig. **??**) and the Gray et al 2009-tree (Fig. 6). Most of the splits in the Glottolog tree occur close to the tips, whereas the splits are more spread out over the distance between the root and the tips in the Gray et al 2009-tree.

Besides parsimony and plausibility, it is also important to know how to weight evidence when conducting historical linguistics research, in particular when it comes to subgrouping. This is less often discussed explicitly, but it is related to issues of plausibility and is likewise a source of disagreement.

As was discussed in **??**, not all data-points are independent of each other and this may be one reason to weight them differently. It is also possible that certain data-points are more susceptible to contact-induced change than others, and should therefore carry less weight if we are trying to infer a family tree. This is why particularly stable items are used in reconstruction and subgrouping (c.f. Pawley (2009)).

For example, Wilson (2012) presents a case for Eastern Polynesia (EP) being settled from the so-called "Northern Outliers" (i.e. Polynesian languages of Micronesia and the Solomons) by demonstrating shared innovations of lexicon and grammar to the exclusion of Samoa, Pukapuka and Tokelau (which were closer to EP in previous proposals). The paper lays out 73 innovations in support for this theory, and states that there is a lack of shared innovations supporting grouping Eastern Polynesia and the Samoic group together, as had been previously suggested by Pawley (1966). Wilson (2012) proposes that a more accurate reflection of this data is to group Eastern Polynesian with the Northern Outliers. On the other hand, Pawley (1966:53, 61) presents two cases where Samoan and some of the Northern outliers shared features to the exclusion of Eastern Polynesia (sing/plural distinctions in indefinite articles and the form of the human number prefix). Besides the sheer number of data-points, it is clear that historical linguists also weight different pieces of information differently. Without an internalised in-depth knowledge of these matters, it is difficult to know how to evaluate the support for these conflicting theories of the origins of Eastern Polynesian communities. Is it as significant that the Northern Outliers and Eastern Polynesian languages shared a word for a certain kind of fish (**kamakama*) as the fact that they have also as a group added an *o-* to the Proto-Polynesian root **fia* (want) (Wilson 2012)?

In this chapter we are not proposing any new subgroupings, so the problem of weighting evidence is not present. We are, however, reconstructing grammatical features and this is another area where weighting is relevant. All languages are weighted the same for the reconstruction and the stability measurements. The tree structure and the method (Maximum Parsimony or Maximum Likelihood) determines the reconstruction. This can be compared to weighting evidence from oversampled areas/subgroups less when reconstructing. Likewise, all features contribute to the conservatism measurement per language.

The Comparative Method is most often applied to vocabulary, but it can also be applied to grammatical morphemes. Crowley (1985) for example traces the history of a common noun phrase marker **na/*a* in Oceanic languages using the Comparative Method.

The data in this chapter does not track specific forms, as is common when reconstructing Proto-languages in historical linguistics (c.f Pawley (1973); Crowley (1985); Evans (2001)). Instead we have a questionnaire that covers large parts of the grammatical domains commonly found in language descriptions in terms of structural features (see section **??**). There are some crucial differences between structural data and the kind of data that is used in historical linguistics in relation to the present study.

The kind of data used in grammatical reconstruction in historical linguistics differs from what we find in linguistic typological questionnaires such as Grambank. Crowley (1985), Clark (1976), and other scholars whose work we will compare to our results in this chapter, typically apply the comparative method to specific formal expressions of structural features (the *na* article, -*Cia* suffix, *faka-* prefix etc). They take into account fossilised forms (the common noun marker -*a* fusing to roots in Paamese (Crowley 1985:141)) and related meanings (the hypothesis of -*Cia* changing from a transitivising suffix to a marker of passive voice (Hale (1968); Hohepa (1967, 1969); Chung (1978) and Jonsson (1998)). The Grambank dataset, however, (as many other typological surveys) only considers productive patterns and does not include information on specific formal expressions of grammatical phenomena.

As has been discussed earlier in this thesis (section **??**), it is important to note that two languages can be coded identically for a structural feature in a typological questionnaire due to entirely different reasons and without being related. For example, Koasati [koas1236] of Louisiana, USA, and Mokilese [moki1238] on Mwoakilloa in the Federated States of Micronesia are both coded as having a construction for predicative possession of the type “Topic” by Stassen (2011). However, they belong to entirely different language families and different parts of the world. This is unlike cognacy data, where the fact that two

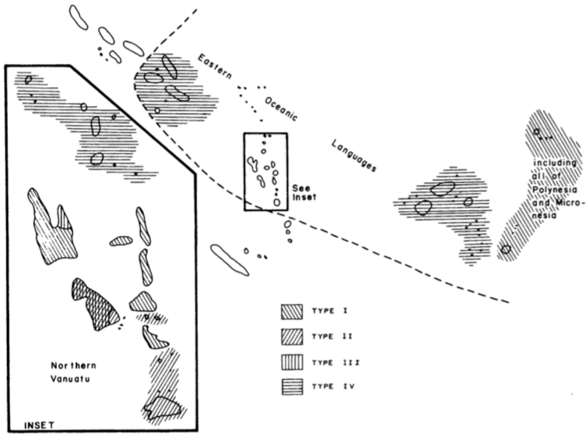


Figure 2: **Map of four different types of common noun phrase markers in Oceanic from Crowley (1985:162). Type 1: absence of common noun phrase marker or marker is not a reflex of *na /*a, type 2: non-productive system involving a reflex of *na /*a, type 3: productive marking involving *na /*a as a prefix that is regularly separable from the noun and type 4: productive marking involving *na /*a generally existing as a free-standing marker.**

languages have cognates in common is direct evidence of relatedness.

An example of the differences between the structural data used in classical historical linguistics and typological questionnaires is the definite marker in Oceanic. Crowley (1985) investigates common noun phrase markers[4] in Oceanic and finds that in many languages there is a reflex of proto-Oceanic **na/*a*, but that in some languages there is another marker with a different origin (Māori *te* for example). In Crowley's study, languages where there is no common noun phrase marking whatsoever and those with a marker which is not cognate with **na/*a* are both included in type 1 (see Fig. 2). These languages are contrasted with those that have retained some kind of reflex of **na/*a* (type 2-4 in Fig. 2). This means that we can distinguish languages which have retained the proto-form from those that haven't, but not languages which have a common noun phrase marker from those that do not.

In contrast, the corresponding feature in Grambank is 'GB022: *Are there prenominal articles?*' (see Fig. 3). Languages that have *te* (like Māori) or reflexes of **na/*a* as articles before the noun both count as "yes" (1) for GB022 and those that have no prenominal marker as a "no" (0). This Grambank feature splits Crowley's type 1 into two categories, and combines all the languages with reflexes of **na/*a*, *te* or other markers into one category. We can now distinguish those that have a pre-nominal article from those that do not, but we cannot tell apart those which have retained the proto-form from those which have not. Since many reconstructions of grammar in historical linguistics rely on explicit formal evidence, this is an important difference.

Related languages may also show similarity due to inheritance, even if the relevant state wasn't present in the ancestor, as the quote from Chung (1977) earlier continues:

[4]This term is more or less identical to a pre-nominal definite/specific article.

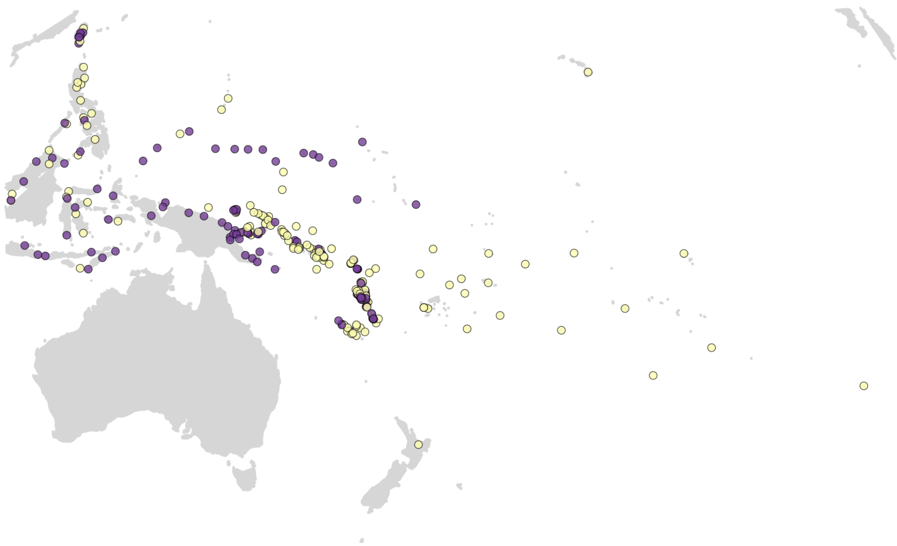


Figure 3: Map of Austronesian languages for GB022 *Are there prenominal articles?* Yellow = "yes", purple = "no".

> *[I]t is also reasonable that changes begun in a Proto-language may have continued even after its separation into daughter languages. In this way, related languages may come to share a feature which existed only in embryonic form, or not at all, in their common ancestor*[5].
>
> Chung (1977:539)

This theory proposes that the conditions for developing a certain feature may have been present in the ancestral language even if the feature itself is absent. It is therefore likely that the daughter languages would develop it even if they were isolated from each other, because they have inherited the prerequisites for developing the feature. This is similar to what is known in the literature on cultural evolution as "preadaptation" (Scott-Phillips & Kirby 2010). One might say that the seed is sown already in the parent language, and that as a result its daughter languages will likely turn out a certain way.

This ties in with the dependencies of features, insofar as dependencies are representations of features "moving as a group" through time. If some of these features are reliable "early movers", one could use them to predict certain developments in daughter languages. This theory aligns well with a view of language as a system where everything neatly fits ("La langue est *un système où tout se tient*" [language is systematic / a system where everything fits], Saussure (Koerner 1997)). This is in contrast to the view by, among others, Bloomfield (1933:328) who stated that *every word has its own history*. It is likely that systematic effects like these are more probable in structural data than in lexical data.

However, in this analysis we are investigating each feature separately so it is not possible to derive information about features co-evolving. It is also difficult for the algorithm *not* to reconstruct a certain

[5]This idea is important to a particular hypothesis about Proto-Polynesian syntax, because according to Chung (1978) Proto-Polynesian was accusative and Tongan and Sāmoan both developed ergativity semi-independently while the Eastern Polynesian languages (which are more closely related to Sāmoan than to Tongan) did not develop this feature.

state in the parent language, if the majority of daughter languages possess it. This is one instance where knowledge of plausible paths and profiles of languages from classical linguistics may contribute information that is, so far, not possible to retrieve using computational means.

In summary, the comparative method of historical linguistics involves Maximum Parsimony coupled with information on plausibility. It is not possible in computational reconstruction to take plausibility into account since it has not been formalised. The kind of data typically used in reconstruction of grammar in historical linguistics concerns specific forms, whereas the data for this chapter is structural features.

2.0.2 Findings from historical comparative linguistics on Proto-Oceanic grammar

The Proto-language of the Oceanic subbranch of the Austronesian language family is well researched in terms of its lexicon and phonology (see the book series on the Proto-Oceanic lexicon (Ross et al. 1998, 2007, 2008, 2011, 2016), among other publications). There is also substantial work done on the grammar of Proto-Oceanic using the comparative method in historical linguistics. In this chapter I have summarised several major works in the field and distilled their research into testable hypotheses given the Grambank data and our methods. This section gives an overview of the works included and examples of how they have been incorporated into the study. Table 1 lists the publications used for the reconstruction of proto-Oceanic by historical linguists in this dissertation. This aggregated reconstruction is also used in chapter ??.

Table 1: Table of historical linguistics publications used in this dissertation for Proto-Oceanic grammar

Citation	**Title**	**Proto-Languages**	**Domains**
Pawley (1970)	Grammatical reconstruction and change on Polynesia and Fiji	Proto-Central Pacific	Verbal markers and aspect particles
Pawley (1973)	Some problems in Proto-Oceanic	Proto-Oceanic and Proto-Polynesian	Possession, noun phrase marking, negation, verbal markers, clusivity, word order
Clark (1976)	Aspects of Proto-Polynesian syntax	Proto-Oceanic and Proto-Polynesian	Alignment, negation, word order, possession, noun phrase marking, voice

Citation	Title	Proto-Languages	Domains
Chung (1978)	Case marking and grammatical relations in Polynesian languages	Proto-Polynesian	Alignment, word order, voice, noun phrase marking
Crowley (1985)	Common noun phrase marking in Proto-Oceanic	Proto-Oceanic	noun phrase marking, clusivity
Jonsson (1998)	Det polynesiska verbmorfemet - *Cia*; om dess funktion i Samoanska	Proto-Polynesian	Verbal marker
Marck (2000)	Polynesian languages (in Facts About the World's Languages: An encyclopaedia of the world's major languages, past and present)	Proto-Central Pacific and Proto-Polynesian	Word order, verbal markers, possession, clusivity
Evans (2001)	A study of valency-changing devices in Proto Oceanic	Proto-Oceanic	Verbal markers
Kikusawa (2001)	Rotuman and Fijian case-marking strategies and their historical development	Proto-Oceanic	Possession, pronominal number
Kikusawa (2002)	Proto Central Pacific ergativity: Its reconstruction and development in the Fijian, Rotuman and Polynesian languages	Proto-Central Pacific	Alignment, word order
Lynch et al. (2011)	The Oceanic Languages, Chapter 4: Proto-Oceanic	Proto-Oceanic, Proto-Central Pacific and Proto-Polynesian	Negation, word order, verbal markers, clusivity, possession, pronominal number, polar interrogation, nominalisations and more

Citation	Title	Proto-Languages	Domains
Ross (2004)[6]	The morphosyntactic typology of Oceanic languages	Proto-Oceanic and Proto-Polynesian	alignment, word order, verbal markers, possession, noun phrase marking

For each of these publications, findings have been extracted that support a certain state of a Grambank feature at a certain node. For example, Marck (2000:4) writes that a causative prefix has been reconstructed for Proto-Polynesian (**faka-*). In the Grambank questionnaire we have the feature GB155 '*Are causatives formed by affixes or clitics on verbs?*'. For the ancestral node that connects all the Polynesian languages, we should expect that for GB155 the state is either wholly or overwhelmingly "yes" (presence). For simplicity, I have only considered four ancestral languages: Proto-Oceanic, Proto-Central Pacific, Proto-Polynesian and Proto-Eastern Polynesian. The choice to focus on these four in particular was because they were the most well-researched in the literature.

Another example of how information in the publications was turned into Grambank feature predictions relates to verbal markers encoding subjects and objects, as proposed by Lynch et al. (2011) among others. In their book, there is a chapter on reconstructions of grammar for Proto-Oceanic and in the section on the basic verb phrase we find the statement below:

> *Attached to the verb root were a subject proclitic and, if the verb had a non-generic object, an object enclitic.*
>
> Lynch et al. (2011:83)

This statement, together with a verb schema provided in the section, support the notion that Proto-Oceanic had subject proclitics and object enclitics. We can also infer from this publication as a whole that the authors believe Proto-Oceanic in fact did *not* have subject *en*clitics and object *pro*clitics. This second prediction relies on absence of evidence and is less strong than the first, but given that the whole chapter is void of any description of object proclitics or subject enclitics being a possibility (including the verb schema) and argument structure is well-discussed, we may dare to make this leap. This information can be translated into the Grambank questionnaire by positing absence and presence for the six relevant features that concern argument marking on the verb (where S stands for subject of intransitive, A for subject of transitive and O for object; see table 2).

[6] This paper makes statements about "canonical" Oceanic languages, which is technically different from *reconstruction* of Proto-Oceanic. However, the author does state that the "canonic type is probably also a reflection of the morphosyntax of Proto Oceanic" (Ross 2004:492) and has given personal approval for the paper to be included in this study in this manner.

Table 2: Example of predictions from historical linguistics as rendered in Grambank features.

Grambank ID	Question	Proto-language	Expert prediction	Reference
GB089	Can the S argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004:498-499), Lynch et al. (2011:83)
GB090	Can the S argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004:498-499), Lynch et al. (2011:83)
GB091	Can the A argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004:498-499), Lynch et al. (2011:83)
GB092	Can the A argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004:498-499), Lynch et al. (2011:83)
GB093	Can the P argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004:498-499), Lynch et al. (2011:83)
GB094	Can the P argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004:498-499), Lynch et al. (2011:83)

The literature suggests that Proto-Oceanic was a language with a pre-nominal definite/specific article (Crowley 1985:136), a distinction between inclusive and exclusive first person pronouns (Pawley (1973:112), Crowley (1985:184), Ross (2004:500), Lynch et al. (2011:67, 75)), no gender distinctions in pronouns (Ross 2004:498), a dual number category in pronouns (Ross (2004:498), Lynch et al. (2011:69) and Pawley (1973:173)), a distinction between alienable and inalienable possession[7] (Lynch et al. 2011:69), prepositions (Pawley (1973:167), Ross (2004:498)), subject proclitics and object enclitics on the verb (Ross (2004:498-499), Lynch et al. (2011:83)), possessive suffixes on the possessed noun (Ross (2004:495), Pawley (1973:155)) and a transitivising suffix on verbs (Pawley (1970:352), Pawley (1973:171), Lynch et al. (2011:80, 92)). More reconstructions have been made, the full table is found in the appendix G. The

[7] A distinction can be made between three different kinds of possessive classification: alienable/inalienable, direct/indirect and dominant/inactive. For the purposes of Grambank and this study, these are treated as similar enough to be included into the same category.

reconstructions regarding ergativity will be presented separately.

Most of the time, the scholars of Proto-Oceanic are in agreement in their predictions. For example, Pawley (1973:142), Ross (2007:292), Clark (1976:xiii, 125) and Lynch et al. (2011:89) all propose that the proto-language of the Polynesian subgroup had a construction marking prohibitive that was different from declarative negatives. However, in some instances there are disagreements. As discussed earlier, one such case is the alignment system of Proto-Polynesian. Clark (1976) claims that the system was ergative while Hale (1968), Hohepa (1967, 1969) and Chung (1978) argue that Proto-Polynesian was accusative and several of the daughter languages developed ergativity later. Crucial to this arugment is also the nature and development of passive voice in Polynesian languages. Because of this disagreement, the results for the computational ancestral reconstruction for Grambank features regarding passive voice and ergativity will be presented separately from the others. There are 109 features in the larger set of non-controversial findings, and 11 in the subset that concerns passives and ergativity[8].

In the Grambank project, research assistants read published grammatical descriptions and extract information such that it fits with the definitions of our typological questionnaire (see section **??**). This survey of the literature on Proto-Oceanic grammar is essentially the same task. Just as with the literature on reconstructed languages, scholars sometimes disagree on the nature of contemporary languages and how they should best be analysed. It is up to the coder, in this instance me, to make calls on which analysis to employ, what can be inferred from the literature and what should be left as unknown. It is possible to squeeze even more findings out of these publications; I have tended to be conservative in my interpretations. Out of the 201 (binarised) features in our questionnaire, 31% (63) were answerable for Proto-Oceanic given this material. The average completion per language in the whole of the dataset is 72%.

2.0.3 Computational phylogenetic methods

In this study, we will be reconstructing the presence or absence of structural features in Proto-Oceanic using Maximum Parsimony and Maximum Likelihood.

As discussed earlier, **Maximum Parsimony** finds the set of ancestral states that result in the fewest number of changes. Maximum Parsimony is intuitively simple. We saw in the previous section an example of how it can play out in a small tree in the work of Clark (1976). While Clark illustrates the principles of Maximum Parsimony well, he does not in fact use the term "Maximum Parsimony".

Part of the critique of Maximum Parsimony is that it does not take into account branch lengths in the tree (the distance between splitting events), only the change from one node to another. Furthermore, Maximum Parsimony necessarily assumes that the solution that posits the slowest possible rate of change

[8] Some of these are duplicates for different proto-languages, i.e. passive voice is examined for proto-Eastern Polynesian and for proto-Polynesian.

is the most likely one. This is not necessarily a valid assumption, some features may overall have a faster rate of change also further back in time.

Maximum Likelihood posits the most likely ancestral state distributions based on the overall probabilities given all the nodes in the tree. If the distribution of values at the tips is very scattered, with sibling pairs frequently having different profiles, Maximum Likelihood will infer that the feature has a high rate of change and use that information when positing ancestral states. The Maximum Likelihood algorithm assigns probabilities of state changes and distributions differently given branch lengths. A mutation along a shorter branch is given more weight in the likelihood calculations than if it occurred in a longer branch.

In this chapter, we compare the performance of both algorithms. More technical details of Maximum Likelihood are found later in this section.

Ancestral State Reconstruction can be done in many different ways. Three of the more common methods are: Maximum Parsimony (Occam's Razor), Maximum Likelihood Estimation and Stochastic Character Mapping (SCM). The choice of method depends on the kind of data and assumptions involved. This study utilises Maximum Parsimony (MP) and Maximum Likelihood Estimation (ML) because they are two of the most well-known and well-understood approaches. For an extensive comparison of different methods and their advantages, see Joy et al. (2016).

Jäger & List (2018) compared three different methods of ancestral state reconstruction for lexical data (cognate classes): Maximum Parsimony, Maximum Likelihood and Minimal Lateral Networks. They found that reconstructions using Maximum Likelihood performed better compared to their defined gold standard than the other two methods. The gold standard consisted of reconstructions by historical linguists, and it should be noted that it is possible that this standard may contain errors as well.

Jäger & List (2018) describe the general performance of all the computational reconstruction methods they used as "poor". They evaluated the methods with F1-scores, which is a harmonic mean of the precision and recall[9] of the result of each method as compared to the gold standard (Van Rijsbergen 1979:133). 0 is the worst possible score, and 1 the best. The highest F1-score was 0.79 (Austronesian language sample, Maximum Likelihood), and the worst was 0.44 (Indo-European, Minimal Lateral Networks). We will also assess the performance of Maximum Parsimony and Maximum Likelihood with F1-scores and compare to theirs.

For Maximum Parsimony, we are using the function `asr_max_parsimony()` from the R-package `castor` (Louca & Doebeli 2017) (which is an instantiation of the method described in Sankoff (1975)) for calculating ancestral states and stability of features. This function produces ancestral states for all nodes and reports the number of changes that was minimally required for each feature.

[9] Precision is true positives divided by True Positives + False Positives, recall is True Positives divided by False Negatives + True Positives. F1-score = 2 * ((precision*recall) / (precision + recall)) (Van Rijsbergen 1979).

For calculating the conservatism of languages a function from a different package, `acctran()` from `Phangorn` (Schliep et al. 2019), was used. ACCTRAN (ACCelerated TRANsformation) is a kind of parsimony analysis where changes are assigned along branches of a phylogenetic tree as close to the root as possible. This function rescales the branches of the tree in accordance with the amount of change the Maximum Parsimony analysis dictates. In order to make the results of language conservatism comparable to the results from the Maximum Likelihood analysis, the trees were re-rooted using Nanggu [nang1262] as an outgroup. Nanggu is a language of Temotu which is genealogically far removed from almost all other languages in the sample, making it a good candidate for outgroup-rooting[10,11]. The function `distRoot` from the R-package `adephylo` (Jombart et al. 2017) was used to calculate the amount of change from the root to each tip, i.e. the conservatism of languages. These distances were rescaled to between 0 and 1 to make them comparable to the results from the other set of analysis.

Ancestral state reconstruction using Maximum Likelihood Estimation involves computing each ancestral state from the tips up to the root taking into account branch lengths and the joint likelihood of states given all nodes in the tree (Wilks 1938; Fisher 1912; Pagel 1994; Cunningham et al. 1998). The Maximum Likelihood Estimation function takes a set of observations and computes the parameter distribution that maximises the likelihood given the observed data[12]. This means that for every split in the tree — every ancestral node — the Maximum Likelihood Estimation function computes what is the most likely distribution at that point given the nature of the entire tree. ML can be modified so that it allows for different rates of change. An Equal Rates (ER) model assumes that the chance of transition from state A to state B and from B to A are equal. However, we as linguists are aware that certain features are more likely to be lost than gained so this is not a reasonable assumption. Therefore, we allow the model to estimate different transition rates for going from A to B and from B to A given the data. This is known as "All Rates are Different" (ARD).

When estimating ancestral states with ML, it is possible to either a) find the state at each node that maximises the likelihood (integrating over all other states at all nodes, in proportion to their probability) at that particular node (marginal reconstruction), or b) find the set of character states at all nodes that (jointly) maximize the likelihood of the entire tree (joint reconstruction). We are using marginal reconstruction in this study since it is the recommended way to deal with uncertainty in reconstruction (Revell 2014). These two methods often yield similar results, but can differ, see Felsenstein (2004:259-260), Yang (2006:121-126) and Joy et al. (2016:5) for more details. For our data, a trial run of joint reconstruction did not generate drastically different outcomes.

For this study, the function `rayDISC` from the `R`-package `corHMM` (Beaulieu et al. 2017) is used for marginal

[10] Another option for outgroup rooting would have been Yapese, since it is an "Oceanic isolate". However, Yapese is not included in the Gray et al 2009-tree so it is not possible to use as an outgroup consistently over all the sets of results.

[11] Mid-point rooting is not appropriate here since that would make for an ultrametric tree where all tips have the same distance to the root.

[12] For a gentle introduction to the concept of Maximum Likelihood Estimation, see Brooks-Bartlett (2018).

reconstruction of ancestral states and rates of change per feature. Missing data in Grambank for langauges included in the analysis was converted to ambiguous, (i.e ? in Grambank → 0&1 for rayDISC()).

For the analysis of conservatism per language, the function `optim.pml()` from `Phangorn` (Schliep et al. 2019) was used. This function rescales the branches in accordance with the Maximum Likelihood Estimation of change along them. The process of re-scaling the tree in this manner unroots the tree; instead of having root, branches and tips, it becomes an acyclic graph of tips connected by lines of appropriate length. Unrooted trees can be re-rooted using midpoting rooting, outgroup rooting or other methods[13]. For this study, the trees were re-rooted using Nanggu [nang1262], Ayiwoo [ayiw1239] or Natügu [natu1246] as an outgroup. As with the Maximum Parsimony analysis, the distance from the root to each tip was carried out with `distTip` and the distances were rescaled to between 0 and 1.

Languages with missing data were pruned away in all analysis, no hidden state reconstruction of values at tips was preformed. Because Nanggu, Ayiwoo and Natügu was used as an outgroup to root the trees in the conservatism analysis, features which these three languages lacked were excluded from the conservatism analysis. This left 167 Grambank features which were included in the conservatism analysis.

Concerning missing data generally, for both methods the trees from Glottolog and Gray et al. (2009) were pruned to only tips representing Oceanic languages which are also found in Grambank. The match between Glottolog and Grambank is 226, the match between Gray et al. (2009) and Grambank is 112. For the parsimony analysis of each feature, languages with missing data were dropped from the trees in the analysis for that feature. Features which could only be assigned values for less than half of the languages in the tree were excluded from the analysis.

For both Maximum Parsimony and Maximum Likelihood it is possible for a structural feature to appear and disappear several times along a lineage. This is different from cognate data where a cognate class cannot re-appear.

2.0.4 Data coverage

For this study, we are using binarised data from the Grambank project (section **??**). We are focussing on Austronesian languages and the Oceanic sub-group in particular.

The Oceanic subgroup covers almost all languages in Remote Oceania, with the exceptions of Chamorro and Palauan. As was discussed in section **??**, Yapese is sometimes not included in the Oceanic subgroup and Nguluwan contains material from Ulithian and Yapese. For the purposes of this study, we have used the definition of Oceanic from Glottolog 4.0 (Hammarström et al. 2019) and Ross et al. (2016:2) (see Fig. 4). This results in Yapese and Nguluwan being included, and Chamorro and Palauan being excluded.

[13] For a visual introduction to outgroup rooting, see https://phylobotanist.blogspot.com/2015/01/how-to-root-phylogenetic-tree-outgroup.html by PhyloBotanist.

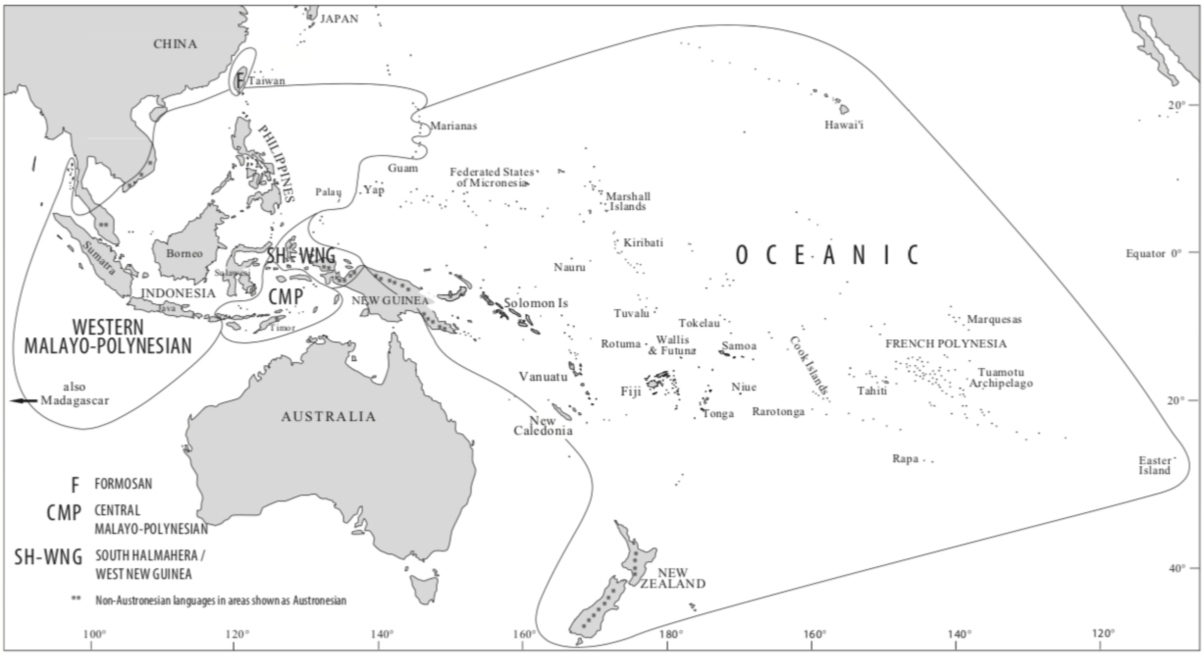


Figure 4: Map of the Austronesian language family and major subgroups, from Ross et al. (2016:2).

The Oceanic subgroup also contains languages from Near Oceania (see section **??**) and these are included in the analysis of this chapter. These languages were grouped into three geographical groups for the conservatism analysis (section 3.0.3): 1) Bismarck (including Manus), 2) Solomons and Bougainville; and 3) New Guinea Mainland/Louisiade Archipelago. See Fig. **??** in section **??** for locations of these regions.

Table 3 shows the number of languages in the Oceanic subgroup that are covered in the Grambank dataset, and which ones are possible to cover in future. Of all the languages in the subgroup, there are 261 languages that have been documented in descriptive works such that we are able to answer our questionnaire. 241 languages are currently too poorly described for us to include them in the database. As has been discussed before, it is not always possible to fill in all the features for every language. Nineteen of the Oceanic languages that are included to date are less than 50% completed. This can be due to lack of access to descriptive work, or that the content of the descriptive work doesn't cover the necessary domains in enough detail for our coders to answer enough questions. The map in Fig. 5 shows the same coverage information for the languages of the region (including a few non-Oceanic Austronesian languages as well).

Table 3: Coverage of Oceanic languages in Grambank (based on Glottolog 4.0 definition of "Oceanic".)

Coverage	**Languages**
No grammar	247
Grammar exists, but language not in Grambank (yet)	48
Less than half of the features covered in Grambank	20
More than half of the features covered in Grambank	205

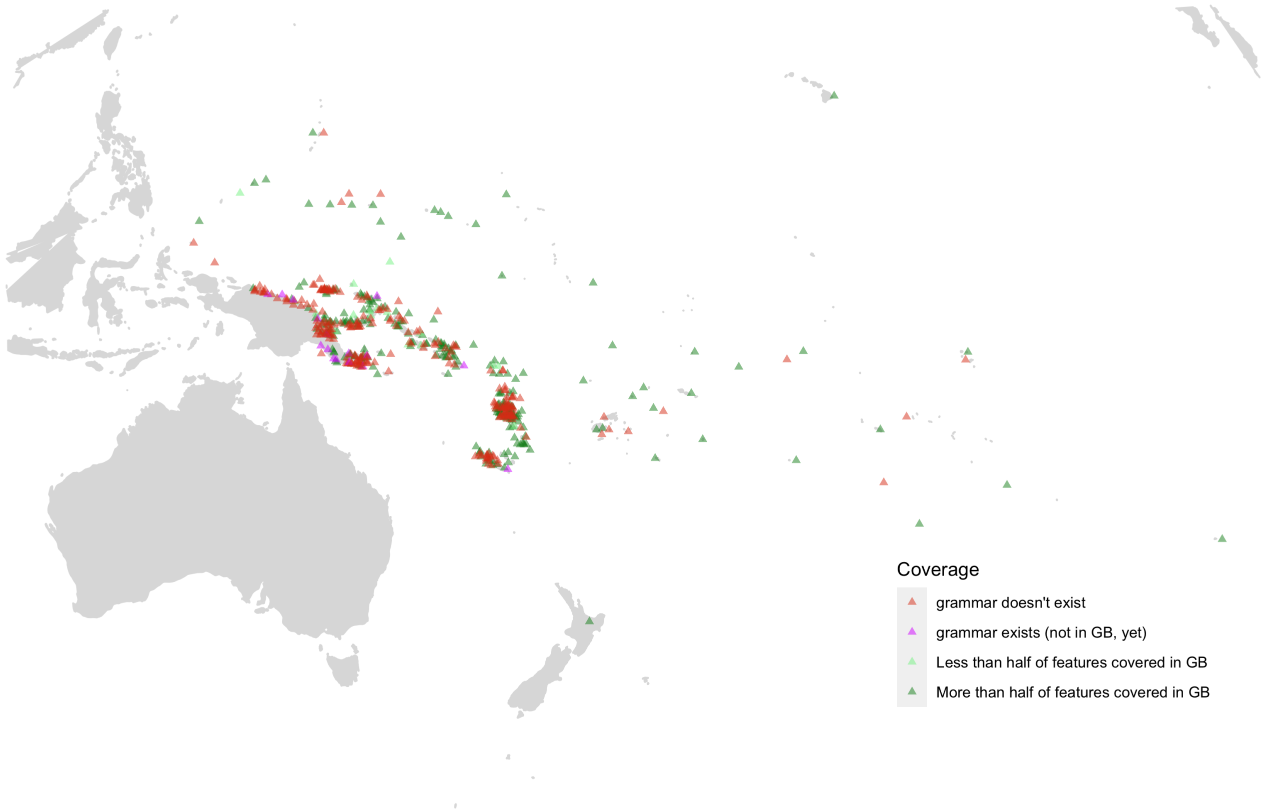


Figure 5: Map of Oceania, with Austronesian languages coloured for coverage in Grambank.

In this study, we will be using the MCCT tree from Gray et al. (2009). Figure 6 shows the Grambank coverage of languages along these tree. Grambank does cover a significant amount of languages east of the Solomons, but the coverage around the coast of New Guinea is still to be completed. Grambank is a work in progress, and these languages are within our target group for coverage in the near future. At this time, the coverage is less complete in the western languages compared to the eastern. However, since we control for genealogical relatedness through the distance control approach, this is less of a problem for our methodology than if we were using traditional probability sampling.

For both Maximum Parsimony and Maximum Likelihood the tree were first pruned down to only languages where there is data in Grambank. For the further Maximum Parsimony analysis, tips representing languages where the data for a particular feature was missing were dropped for the analysis of that feature. For the Maximum Likelihood analysis, the value at such tips was converted to ambiguous. This is necessary because otherwise the rate of change across the different features would not be comparable in the Maximum Likelihood analysis.

The tree from Gray et al. (2009) contains duplicates (see for example Nakanai). This is because it is a tree of word-lists for languages (doculects) rather than languages themselves. There are also some instances where multiple dialects of one language is included. For the analysis, only one tip per language was retained, based on which had best coverage in the underlying data for the tree (i.e. the Austronesian Basic Vocabulary Database, ABVD (Greenhill et al. 2008)). This means that duplicate languages were reduced to one, and only one dialect per language retained if there was more than one.

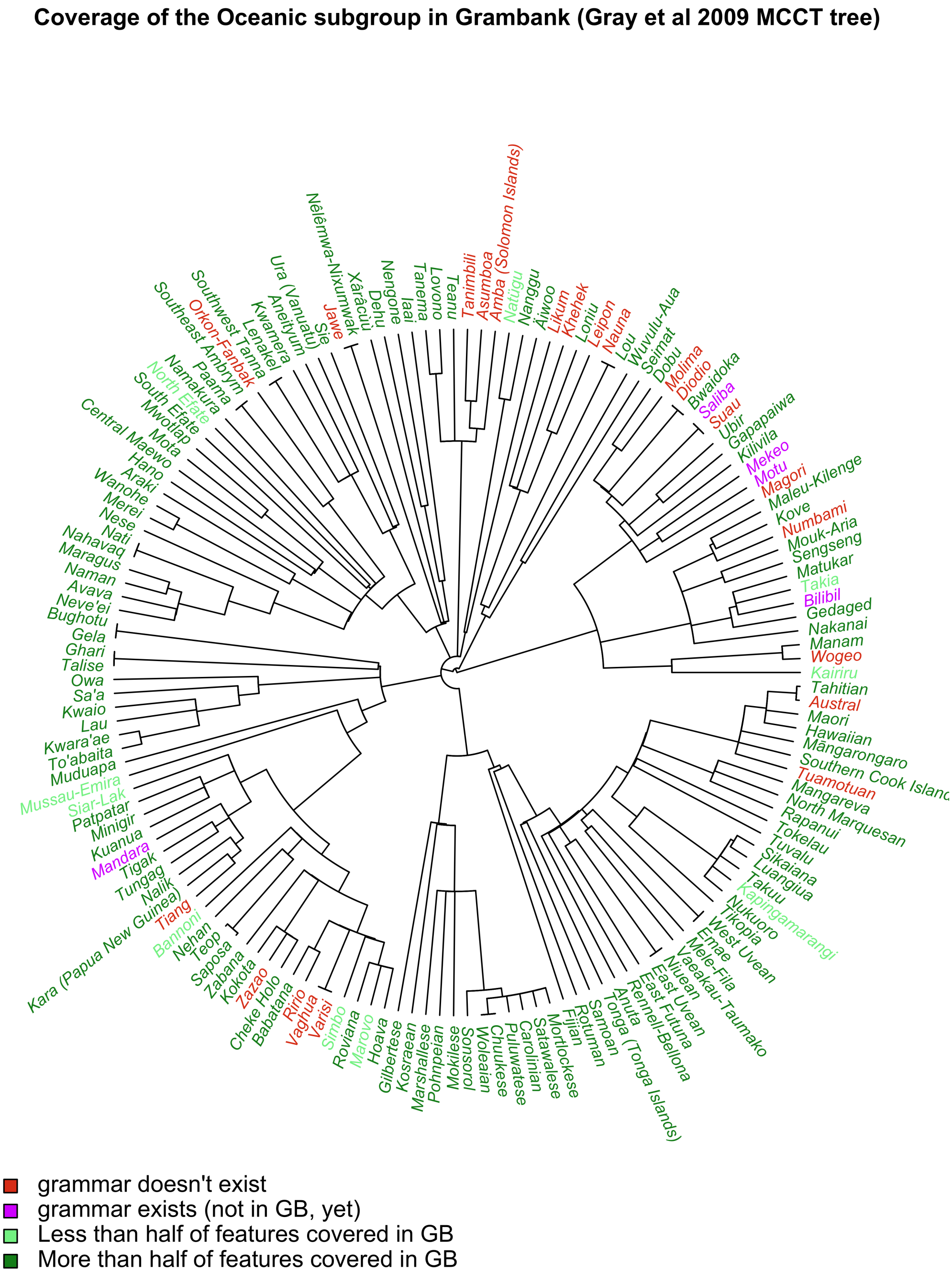


Figure 6: Tree of Oceanic from Gray et al. (2009), with languages coloured for coverage in Grambank.

3 Results

The results of this study are divided into three parts:

1. Concordance between computational reconstructions and findings in traditional historical linguists
    - the ergativity in proto-Polynesian-issue
2. Stability of features
3. Conservatism of languages

In this study we are using the MCCT tree from Gray et al. (2009) and two models (Parsimony and Maximum Likelihood Estimation), making for two sets of results. Trees with more than half of the possible tips missing are discarded. The match between languages in Grambank and the Gray et al-tree is 112, meaning that results with less than 56 tips are ignored.

All results have been calculated in `R` (R Core Team 2019) using the packages `castor` (Parsimony) (Louca & Doebeli 2017), `phangorn` Schliep et al. (2019), `adephylo` (Jombart et al. 2017), `corHMM` (Beaulieu et al. 2017), `ape` Paradis et al. (2004) and phytools (Revell 2012) and tidyverse (Wickham et al. 2019).

3.0.1 Concordance with traditional historical linguistics

In this section we compare how often the computational reconstructions are in concordance with those made by historical linguists. For each feature, the algorithm predicts a distribution of the two states (presence and absence) for every ancestral node. If the distribution is majority presence (more than 60% of the ancestral state is "1") it is registered as "Presence"; if less than 40% presence it is registered as "Absence". If the ancestral state is between 40-60% of either state, the prediction is registered as "Half/Half". If the reconstruction of a feature by experts for that ancestral node was "Presence" and the algorithm did predict presence with over 60%, it is a "True positive", and so on. Table 4 illustrates how the results are calculated.

Table 4: Table illustrating how the results of ancestral node predictions are calculated.

Finding in historical linguistics	**Prediction by MP or ML**	**Result**
Presence	>60% Presence	True Positive
Presence	>60% Absence	False Negative (type 2-error)
Absence	>60% Absence	True Negative
Absence	>60% Presence	False Positive (type 1-error)
Presence	40-60% Presence/Absence	Half/Half
Absence	40-60% Presence/Absence	Half/Half

In order to evaluate the results, we need to calculate a concordance score per method and tree. Jäger & List (2018) use the F1-score (harmonic mean between precision and recall) in their study of how computationally reconstructed lexical proto-forms compare to those reconstructed by historical linguists. For example, if for a given proto-language there are 60 features reconstructed by experts and the algorithm result is 10 True Positives, 10 False Positives, 10 True Negatives, 10 False Negatives and 20 "Half/Half" then the F1-score is 0.5 (recall = 10 / (10+10) = 0.5, precision = 10 / (10+10) = 0.5 and F1-score = 2 *((0.5*0.5) / (0.5 + 0.5) = 0.5)). Note that the F1-score disregards True Negatives and Half /Half-results.

F1-scores will be reported because they are insightful and have been used in similar studies. However, the F1-formula ignores the amount of True Negatives and Half/Half results. Therefore, in addition we will also calculate a simpler concordance score; how many concordant predictions did the algorithm make given all the predictions it made (aka "accuracy")? For example, if for a given proto-language there are 60 features predicted by experts with the same distribution of results as in the example above, then the concordance score would be (10 + 10)/40 = 0.5. We can also include the Half/Half-predictions, awarding 0.5 points for at least not strongly predicting a false value. In that case, this example has a concordance score of 0.5 ((10 + 10 + (20/2)) / 60). These scores all reflect different ways of assessing concordance and will give different perspectives on our results and how our algorithms are performing.

The full table of all predictions (excluding those relating to ergativity) and all results per feature, tree and method can be found in table G in appendix 17. The summary results for the concordance with reconstructions by experts from historical linguistics are presented in table 5.

Table 5: Comparison of how often the computational ancestral state reconstruction agrees with reconstruction from historical linguistics literature.

	Gray et al (2009)-tree	
	Parsimony	**Maximum Likelihood**
Agree	95	91
Disagree	13	10
Half / Half	7	13
Accuracy score (incl Half/Half)	0.857	0.855
Accuracy score (excl Half/Half)	0.88	0.901
False Negatives	8	8
True Negatives	42	41
False Positives	5	2
True Positives	53	50
F1-score	**0.891**	**0.909**

One of the most striking features of these results is the number of Half/Half predictions in the ML results.

The ML approach takes into account the overall likelihood of the entire data, and for our dataset and trees this leads to more Half/Half results compared to the parsimony approach. One way of interpreting this is that the ML is more careful, it makes fewer confident predictions (more than 60% either way), but it also disagrees less often with the findings from historical linguistics.

As was discussed earlier, there are several different ways of evaluating the performance of these approaches. If we consider how often they made a prediction that agreed with historical linguistics out of all the times they made a confident prediction (Accuracy score excl Half/Half), then ML. If we award half a point for Half/Half predictions, Parsimony does better. However, in both instances the numbers are very close to each other.

In a similar study, Jäger & List (2018) attempt to reconstruct cognate classes for the proto-languages of three different language families. They used various approaches: binarising versus not binarising data; Maximum Parsimony versus Maximum Likelihood Estimation versus Minimal Lateral Networks; and using a single consensus tree versus sampling several from the tree posterior. The highest F1-score they achieved in this paper was 0.79 for the Maxmimum Likelihood Estimation reconstruction of Austronesian (using either a single tree or a sample of trees). This means that all of our results above perform "better". While this may be pleasing, it is not yet entirely clear why this is.

In this study, only statements about ancestral languages that could be mapped to Grambank-features were included. It is possible that the Jäger & List (2018) study had a greater overlap between all the reconstructions made by historical linguists and the meanings that they had data for. It is also possible that the set of features that were possible to include were also somehow easier to reconstruct, and if so that would explain the higher F1-scores.

Many of the features that have been reconstructed for proto-languages of the Oceanic subgroup are also very common among Oceanic languages. For example, in our dataset 223 languages have a distinction between alienable and inalienable possession, and three do not have this feature . It is perhaps no surprise that historical linguists, Maximum Parsimony and Maximum Likelihood agree that it is likely that the proto-languages have this feature as well[14]. However, it is not always so simple. If this was all there was to it, we could just use raw distributions to reconstruct features of proto-languages. Historical linguists stress the importance of Maximum Parsimony and plausibility in their reconstructions, and as we saw in Fig. 1 (section 2.0.1), raw distributions alone can be misleading — it is essential to take into account the tree structure. For example, Pawley (1973:118) suggests that verb-final word orders may have been possible in proto-Oceanic. This is tracked by feature GB133 *'Is a pragmatically unmarked constituent order verb-final for transitive clauses?'*, and most languages are marked as 'no' for this feature. However, Maximum Likelihood and the Gray et al 2009-tree did reconstruct presence of this feature at the root.

[14] There was one exception. Maximum Likelihood on the Glottolog tree did not confidently predict presence of alienablity possession for Proto-Oceanic, the result was classified as "half". It did however predict alienablity for Proto-Polynesian.

108 of the tips of this tree were absent, and 7 present, and yet the result was presence for Proto-Oceanic. This is (partially) due to the particular tree structure, where the languages with this feature attach further up in the tree structure (see Fig. 7)[15,16].

[15]It should be noted that the Maximum Parsimony result for the same tree did not reconstruct presence at the root. This has to do with the way Maximum Likelihood takes into account the distribution across the tree in each reconstruction, which gives the ancestral node of Maleu-Kilenge [male1289] and Kove [kove1237] a higher change of presence than it would under Maximum Parsimony.

[16]The languages with a presence for this feature are also mostly on the island of New Guinea and it is possible that this is a result of contact with non-Austronesian languages, as anonymous examiner 3 kindly pointed out.

Besides the predictions made by historical linguists, we can also explore what else has strong support in our computational reconstructions that is not explicitly mentioned in the literature. For example, for Proto-Oceanic, three out of the four sets of results predicted that the order of numeral and noun is N-Num, all tests supported that "adjectives" in Proto-Central Pacific behaved like verbs when used predicatively, and all tests also supported that Proto-Polynesian had three or more distance contrasts among demonstratives.

3.0.1.1 Ergativity The nature of the alignment system of Proto-Polynesian is contested, and therefore the features that concern passive voice and ergativity are presented separately from the rest. Clark (1976) posits, primarily on the basis of parsimony, that Proto-Polynesian was ergative whereas Hale (1968), Hohepa (1967, 1969), and Chung (1978) argue that it was accusative (while they suggest different historical pathways, they agree that Proto-Polynesian was nominative-accusative).

Grambank has two features that pertain to this argument:

- GB408 *Is there any accusative alignment of flagging?*
- GB409 *Is there any ergative alignment of flagging?*

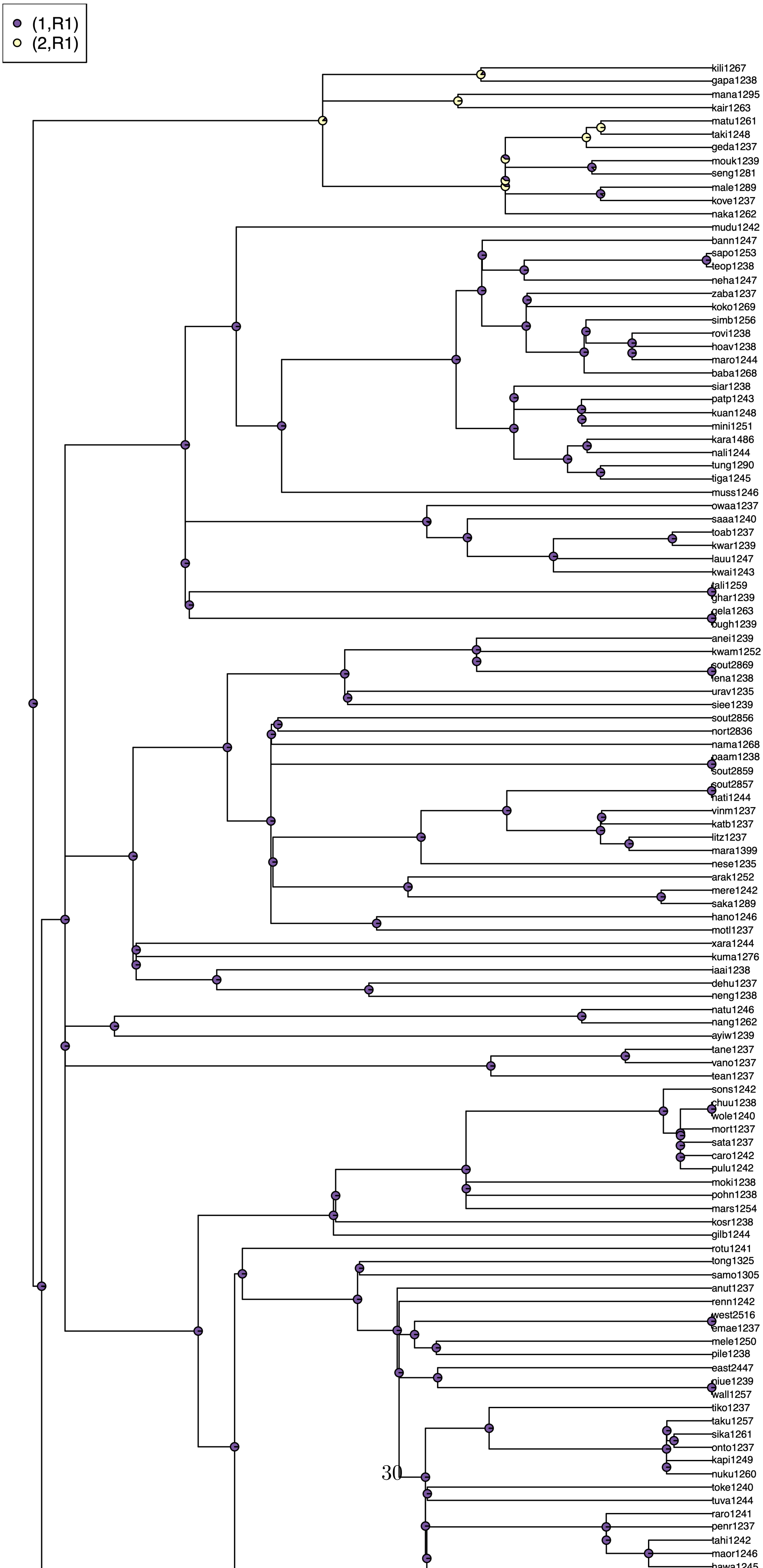
It is entirely possible for a language to be entered into the database as "yes" for several of these, i.e., from the perspective of Grambank languages aren't "ergative" or "accusative" — they can have both ergative and accusative flagging simultaneously. This makes it possible for us to prove both Chung and Clark "right", the results can come out such that Proto-Polynesian had both accusative *and* ergative alignment flagging. However, the results do come out strongly in favour of the proposal by Clark. Table 6 shows that both methods reconstruct ergative alignment for Proto-Polynesian, and only one method predicts a half/half chance of a nominative/accusative system. This lends support to Clark's suggestion that Proto-Polynesian was ergative.

Table 6: Table showing the results for the alignment system of Proto-Polynesian

Feature	**Parsimony**	**Max Likelihood**
GB408 *Is there any accusative alignment of flagging?*	Absent	half
GB409 *Is there any ergative alignment of flagging?*	Present	Present

As was noted earlier, the computational reconstructions differ from those arrived at through the comparative method primarily because the data used in this study is abstract presences or absences of structural features whereas historical linguists use specific concrete forms instead (c.f. Crowley (1985)). In the case of alignment systems, the matter of concrete markers is less of an issue. However, besides the parsimony principle (as laid out by Clark (1976:19) for example), expert historical linguists also take into account the plausibility of the proposed proto-language and the chain of changes posited (Chung 1977). It is not

Feature: GB133 (lh=−21.308, p(root)=0.96, 0.04)



possible for the computational reconstructions to take these assumptions into account without having them formally described and introduced into the model, which is not possible at this time. This may be the reason for the lack of support for Chung's theory; the crucial information that underpins it is not accounted for in this study.

Given the topology of the two trees used in this study, where the ergative flagging language Tongan is always attached to the Proto-Polynesian root at a higher level than Eastern Polynesian languages, it is very likely that GB409 would be reconstructed as present for Proto-Polynesian. As Clark pointed out, it is the most parsimonious solution. However, it could still have been the case that GB408 (accusative) or GB410 (neutral alignment) would have been reconstructed for Proto-Polynesian. The reasons for this may lie in different definitions of what counts as nominative-accusative or neutral in different descriptions, and/or in discussions of plausibility. As has been discussed earlier, it was not possible to include plausibility as a factor in this study.

The proposals of Hale (1968), Hohepa (1967, 1969), and Chung (1978) also involve reconstruction of passive voice that relate to the development of the ergative systems. They suggest different pathways by which languages can develop from a nominative-accusative system to an ergative-absolutive one that rely on changes in the specifics of the passive voice construction that we unfortunately do not track. Given our data, which simply records presence of a productive passive voice marker on the verb, we are not able to scrutinise the three precise theories in greater detail. The results largely support the hypothesis that Proto-Eastern Polynesian had a passive voice marker and that Proto-Oceanic and Proto-Polynesian did not.

3.0.2 Stability of features

We assess the stability of the features in our data by measuring the Parsimony cost (number of changes in the tree) and the rate of gains and losses in the ML results. Features where more than half of the languages of the tree were missing were excluded from these results.

Parsimony cost represents the number of changes inferred as part of the most parsimonious solution. Trees where all the tips are of the same value will have a parsimony cost of 0, since no change occurred. A parsimony cost of 1 will indicate that 1 change occurred somewhere in the tree, and so on.

The ML results produce a rate of gain and loss for each feature given the full tree. In order to make the rates comparable across features, all the trees used have the same number of tips (with tips with missing data being converted to ambiguous states). For the results, the average rate (mean of rate of gain and loss) is considered.

The large majority of the most stable features are dominated by rare phenomena. For example, all four sets of results rank the following two features as among the five most stable:

Table 7: Top 5 most stable features per method and tree, only including features with a distribution of at least 5%/95%.

Grambank Feature	Glottolog 4.0-tree		Gray et al (2009)-tree	
	Parsimony	Maximum Likelihood	Parsimony	Maximum Likelihood
GB081 VInfix		Top-5	Top-5	Top-5
GB433 POSSSfxPosd		Top-5		
GB131 TransVInitOrder		Top-5		
GB132 TransVMedOrder		Top-5		
GB024b OrderNNUM		Top-5		
GB188 AUGBound	Top-5		Top-5	
GB422 ComplThinkKnowPost	Top-5		Top-5	
GB300 VSupplGive			Top-5	
GB095 CaseSplitTAM			Top-5	
GB133 TransVFinalOrder	Top-5			
GB264 QPartMedial	Top-5			
GB330 RELCorr	Top-5			
GB091 A-ArgSfxV				Top-5
GB333 NUMDecimal				Top-5
GB148 AntipassiveBoundV				Top-5
GB335 NUMVigesimal				Top-5

1. GB110 *Is there verb suppletion for tense or aspect?*

2. GB149 *Is there a morphologically marked inverse on verbs?*

However, these are both entirely absent from the entire group. There is no Oceanic language in Grambank with verb suppletion for tense or aspect, nor a language with inverse marked on verbs. It is little wonder that these features are stable, no change is needed from the reconstructed root (absence) to all the tips (also absence). This is a characteristic of some of the results from the study by Dediu & Cysouw (2013) as well: many of the most stable features in their overview were also rare.

In order to avoid the most rare features, the results are also reported for only features where the balance of presence/absence is at least 5% one way and 95% the other. The top and bottom five for each method and tree are reported in appendix H, including a separate table for stable features with at least a distribution of 5%/95%.

Table 7 is a summary of the top-5 most stable features in each of the four sets of results, including only features with a distribution of at least 5%/95%.

These results indicate that there is little agreement between the methods as to which features are most stable. Three out of four sets of results rank verbal infixing (GB081) as among the five most stable and while it is indeed less rare than 5%/95%, it is not much rarer. Only 8% of the languages in the sample have this feature.

There are a few interesting features in this summary table. Counting systems (GB333 and GB335) are often said to be stable. There are also a few word-order related features, which is in line with previous research (c.f. Dediu & Cysouw 2013).

However, overall these results do not strongly support that certain features are more stable than others consistently. This is likely because Oceanic is not a deep enough genealogical unit for this kind of analysis. Features are either overwhelmingly present or absent with little in between. This makes phylogenetic analysis difficult. It is possible to consider other factors besides time-depth as relevant factors for stability, such as social network structure/political complexity (see section **??**), but this is outside the scope of this study. Furthermore, it is possible that the characteristics of structural data are such that it is much more difficult to apply ancestral reconstruction methods (there are several different reasons for the presence of a structural feature, and they may be unrelated, see **??** and section 2.0.1).

3.0.3 Conservatism per language

Conservatism is a measurement of how much change there has been from the reconstructed root of the tree (proto-Oceanic) down to each of the tips (languages). A high number indicates a large amount of change has occurred between the root and the tips. The change can go back and forth, i.e. flip between presence and absence of a feature several times between the root and the tip.

This is different from the direct pairwise dissimilarity between each language and a reconstructed version of Proto-Oceanic which we saw in section **??**. The analysis of pairwise dissimilarity did not take into account the intermediate nodes between Proto-Oceanic and each language, and only included the features where findings in historical linguistics indicated a particular feature in Grambank for Proto-Oceanic. The analysis here differs in two important ways: a) the state of the intermediate nodes are taken into account and b) the tips are compared to the computationally reconstructed Proto-Oceanic (which includes more features than there are findings for in historical linguistics).

For the two parsimony sets of results, conservatism was calculated by a kind of parsimony analysis known as ACCTRAN (ACCelerated TRANsformation). This lets us examine how many changes occur from the root down to each of the tips according to the most parsimonious solution of ancestral states.

For the ML analysis, the function `optim.ml()` was used to rescale the branches in accordance with the changes that the ML (marginal) reconstruction posited. The result of `optim.ml()` is an unrooted tree, which is why it is necessary to reroot. Nanggu [nang1262] was used as an outgroup to root the trees, since it is well-known as an Oceanic outlier.

In order to make the results comparable across the two methods, the ACCTRAN trees were also re-rooted with Nanggu as an outgroup. Because Nanggu was used as an outgroup, features which Nanggu lacked were excluded. This left 167 Grambank features which were included in the conservatism analysis.

The results from both the Maximum Parsimony and Maximum Likelihood were rescaled to between 0 and 1 in order to make them comparable. A score of 1 means that a great deal of change has occurred, whereas a score of 0 means that no change has occurred. Fig. 8 shows the languages of the sample coloured by conservatism in each of the four sets of results given our two methods and two trees.

Across the four sets of results the map visualisation indicates that the most conservative regions appear to be the Bismarck archipelago and Temotu (see Fig. ?? in section ?? for locations of Bismarck and Temotu).

In order to evaluate the conservatism of the different regions in our data the languages were grouped into island groups. The same groups as in chapter ?? were used with the addition of three groups in Near Oceania, making for a total of ten groups (ordered by mean conservatism averaged over all four sets of findings): Temotu, Bismarck, Solomons and Bougainville, Northern Vanuatu, New Caledonia, New Guinea Mainland/Louisiade Archipelago, Micronesia, Central Vanuatu, Central Pacific and Southern Vanuatu.

Figs. 9 and ?? show the distribution of conservativeness across the island groups over the four methods. The ridgeplots represent the distribution of conservativeness in each island group but points have also been added to show the precise locations of the data-points. The line and label on each distribution indicates the mean. For the analysis with the Gray et al 2009-tree there were unfortunate only two languages of the Northern Vanuatu group present, which is not enough to generate a ridgeplot but the points indicate the values of the languages there.

The results from the different methods and trees differ, we shall take a closer look at this later and compare the scores to each other and to time of settlement.

Temotu is the most conservative group in the sample across all the four sets of findings. If we average the conservatism of all the analyses, Temotu languages have a mean of 0.18 conservatism compared to Bismarck's 0.46. However, Nanggu was used to outgroup root the tree, which makes Nanggu the first split from proto-Oceanic in the tree which was used in the reconstruction. This is likely to have inflated the conservatism of Nanggu and therefore the conservatism of its closest relatives, which are other languages in the Temotu group. This is likely to have enhanced the conservatism of the Temotu island group. However, it should be noted that Temotu languages did also have the lowest direct distance to Proto-Oceanic as reconstructed by historical linguists (section ??).

It is possible that if certain ancestral nodes in the tree had been anchored by archaeological accounts, such as those found in Rieth & Cochrane (2018) or Kirch (2017), Temotu would have been less likely to be the most conservative. This is because the settlement order would definitely be Bismarck first and Temotu later.

Historical linguistics research indicates that the most likely homeland of Oceanic is found in the Bismarck archipelago (Lynch et al. 2002:97). The second most conservative group in our sample is indeed Bismarck, which supports this notion.

As was discussed in the previous chapter (??) certain languages of Temotu, Southern Vanuatu and New Caledonia are known to be unusually innovative lexically and phonologically (Grace 1981, 1992; Pawley

Figure 8: **Map of Oceanic languages and their average conservatism given the four sets of results. Yellow = progressive, blue = conservative. First row: Maximum Likelihood branch lengths, second row: Parsimony cost. First column: Glottolog-tree, second column: Gray et al 2009-tree.)**

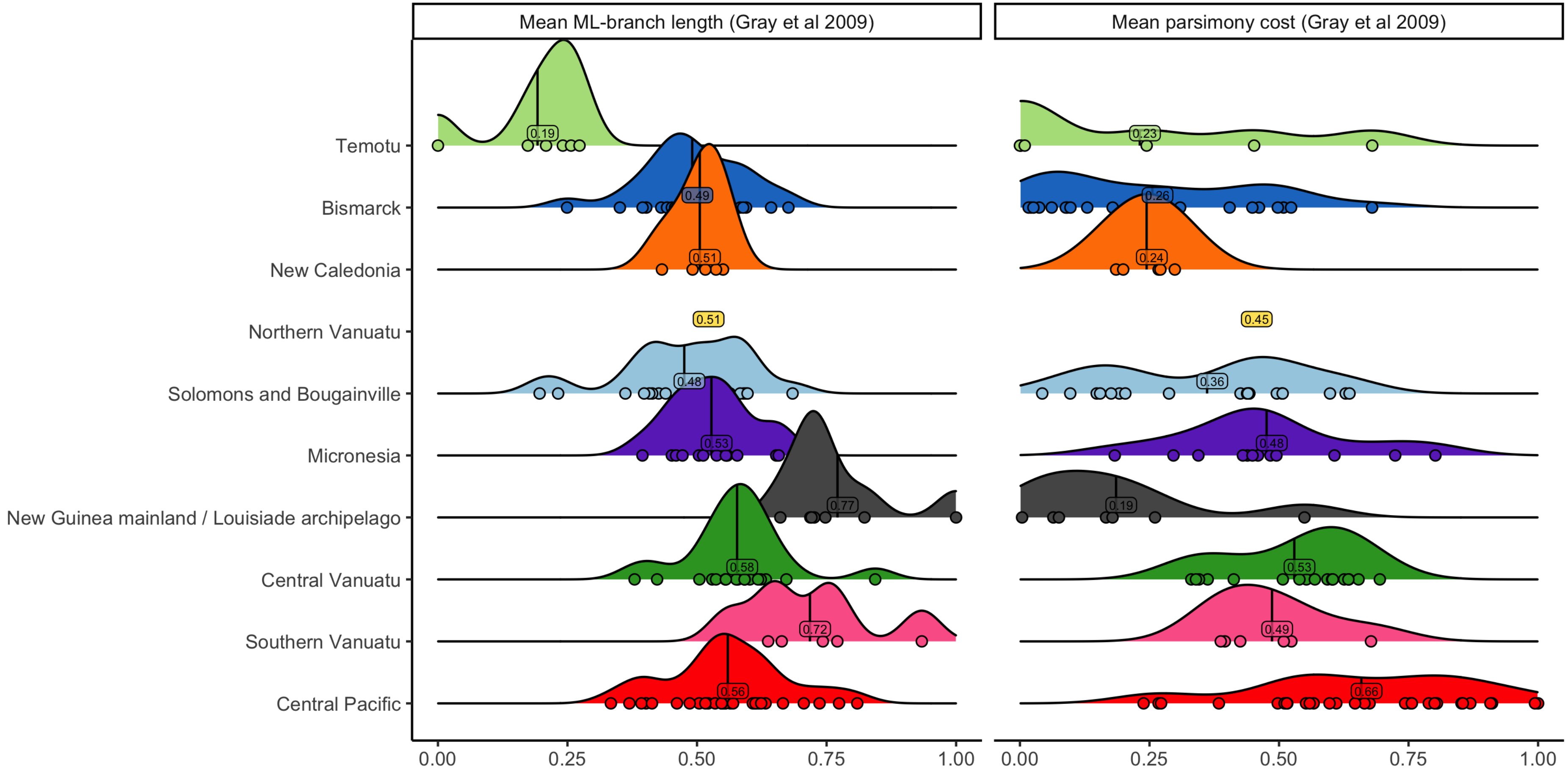


Figure 9: Ridgeplots of the distribution of conservatism over the island groups per method.

2006). In the previous chapter, we saw that Temotu and New Caledonia were more unusual lexically than Southern Vanuatu, and that structurally there was little difference. In these results however, Southern Vanuatu is the most progressive in the MP results, and in the mid-range in the ML-results, whereas Temotu is the least conservative overall and New Caledonia appears in the mid-range in all the results. It appears that it took more transitions back and forth to "get to" the structure of Southern Vanuatu languages compared to the other two "aberrant" groups Temotu and New Caledonia.

The languages of the Central Pacific are known as especially *conservative* in their lexicon (Blust (1981, 2000) and Pawley (2009)). However, this seems to not hold structurally. The Central Pacific Languages are the most progressive in the MP results (after Southern Vanuatu) and in the upper-mid range in the ML results. A few features that make the Central Pacific languages stand out are the lack of subject markers as prefixes or proclitics on verbs (GB090 & GB092) as well as the presence of ergative case marking (GB409) and passive voice (GB147).

However, there are a few methodological considerations that are important to the interpretation of these results, in particular concerning Central Pacific. The difference between the Maximum Parsimony and Maximum Likelihood results are most likely due to branch lengths and the number of nodes between each tip and the root. For the Maximum Parsimony analysis, the branch lengths are irrelevant. Maximum Parsimony only reconstructs states at each intermediate node, regardless of how close that node is to another. This means that if there are more nodes between the root and the tip, Maximum Parsimony has more "opportunity" to posit changes than if there were fewer. The Maximum Likelihood analysis however takes into account the length of the branches, which in turn means that the particular number of nodes between the root and the tips is of less importance.

The two trees we have used in this analysis, the Glottolog tree[17] and the Gray et al 2009-trees, have a structure such that there are more nodes between the languages of Central Pacific and the root (Proto-Oceanic) than there are between languages of Central Vanuatu and the root. The tree topologies are displayed in section 2.0.4, but as it is difficult to appreciate this difference in the plots there I have summarised the number of nodes between tips and root per island group in table 8. The precise number of nodes between the tips and root differ between the two trees, and also between the two methods since the Parsimony analysis drops tips with missing data. However, the rank orders are largely the same, with Temotu languages having the fewest number of nodes between them and the root and Central Pacific the most.

Since there are more nodes between Central Pacific and the root this means that the Maximum Parsimony algorithm has more "chances" to posit changes. This is not true of the Maximum Likelihood analysis which is able to posit changes in relation to the branch lengths and cares less about the precise number of nodes on the tree. Fig. 10 shows the language conservatism score per method and tree compared to

[17]Note that both trees have been binarised for the analysis in this chapter.

Table 8: Mean number of nodes between tips and root per method and tree.

	Mean number of nodes between tips and root				
		Max Parsimony		Max Likelihood	
Island group	**All**	**Glottolog**	**Gray**	**Glottolog**	**Gray**
Temotu	2.95	2.83	2.58	3.25	3.14
New Caledonia	9.85	11.19	7.73	12.08	8.40
Southern Vanuatu	10.29	10.56	9.15	11.62	9.83
New Guinea mainland / Louisiade archipelago	10.66	11.67	8.35	12.31	10.33
Micronesia	11.32	11.92	10.08	12.37	10.92
Northern Vanuatu	11.51	11.94	10.31	13.80	10.00
Bismarck	11.77	13.18	9.24	13.24	11.43
Central Vanuatu	12.47	12.90	11.21	13.15	12.62
Solomons and Bougainville	12.96	14.45	9.91	15.58	11.88
Central Pacific	14.55	15.20	12.80	16.54	13.67

the number of nodes between the languages and the root. There is a stronger correlation between the number of nodes along the route from a specific language (tip) to the root and the Maximum Parsimony scores of conservatism for that language (upper row) than there is between the Maximum Likelihood and the number of nodes (lower row). The correlation between the number of nodes and the Maximum Likelihood, though significant, is very weak.

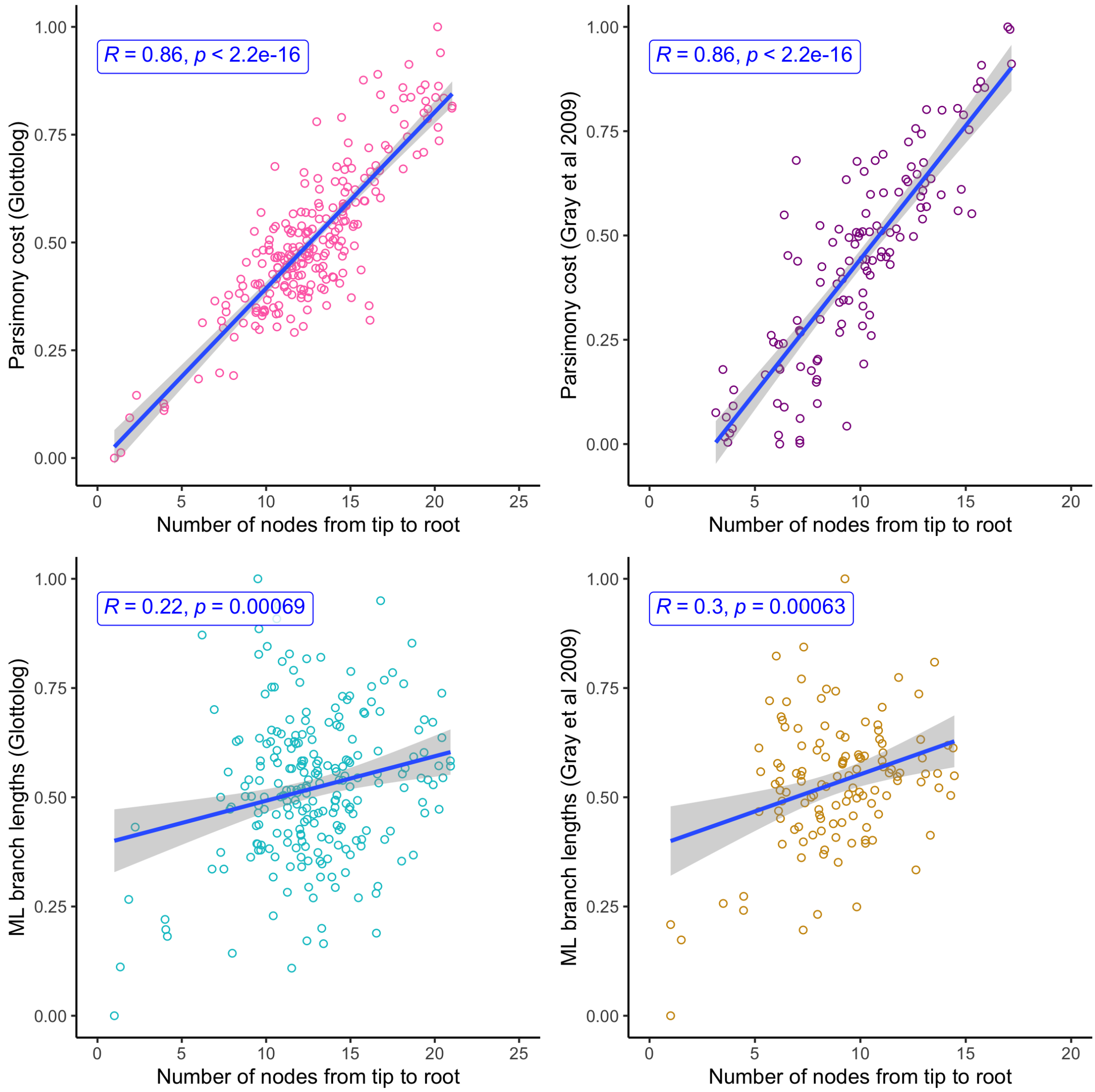


Figure 10: Scatterplots comparing number of nodes from root to tips and conservatism scores.

This means that the Maximum Parsimony conservatism score is mainly reproducing the number of splits between the language and the root — the more splits, the higher the Parsimony cost (amount of change). Conversely, fewer splits mean less change (lower Parsimony cost). Temotu languages are more conservative in the Parsimony analysis than in the Maximum Likelihood analysis, and the outgroup-rooting with Nanggu results in fewer nodes between the root and the Temotu languages. Whether or not this is good or bad depends on what we think splits in the tree represent. Do we expect languages which are more deeply nested in a family tree to have had a more tumultuous history? Or do we think that the rate of change is not correlated with the number of splits found in the tree — that "early" off-shoots like Nanggu and Yapese are equally likely to undergo substantial changes as Central Pacific language which have more splits in their lineage?

Furthermore, sub-nesting may also be a product of research efforts — areas which have been more well-researched may be better represented in the tree of Glottolog than those which have had comparatively less research. Note that this is not as relevant for the Gray et al 2009-tree which is based on lexical and archaeological data, not compilations of historical linguists' accounts like the Glottolog tree[18]. It is beyond the scope of this dissertation to reach any final conclusions on how we should interpret the nature of splits in these trees. The findings of this chapter indicate that the more splits in a lineage the less conservative the language is in the Maximum Parsimony analysis, which informs us that the Maximum Parsimony is adding little information concerning conservatism beyond the number of splits. How we should interpret the number of splits is however not clear.

Bringing this back to the specific island groups, it is still noteworthy that Central Pacific is not among the most conservative languages structurally in either of the four sets of results. This goes against what we might have expected given the lexical conservatism of these languages found in other publications.

More remarkable perhaps in light of the correlation between number of nodes and the Maximum Parsimony conservatism score is the progressiveness of Southern Vanuatu languages and the conservatism of Bismarck languages. Table 8 shows that Southern Vanuatu languages have the third least number of nodes on average between the root and the languages, and Bismark's languages appear in the mid range. And yet, the Southern Vanuatu languages are among the most progressive in the results, with Bismarck languages among the most conservative. This indicates that the progressiveness of the Southern Vanuatu languages and the conservatism of the Bismarck languages are robust findings.

We can also compare the conservatism of the languages across the methods, and to known settlement dates. Fig. 11 shows a scatterplot matrix comparing the four conservatism-per-language scores to each other and also to settlement order (see section **??** in chapter **??**). The lower triangle shows scatterplots of different combinations of data. The cell in the second row and first column displays a scatterplot of the two ML results compared against each other. The dexter diagonal shows the histograms of the various datasets. The upper triangle shows the Pearson's statistic. Stars indicate significance in the conventional manner.

[18] A total of 102 references are used in Glottolog for the structure of the Austronesian family tree, with the main references being Blust (2009, 2014) and Blust & Chen (2017).

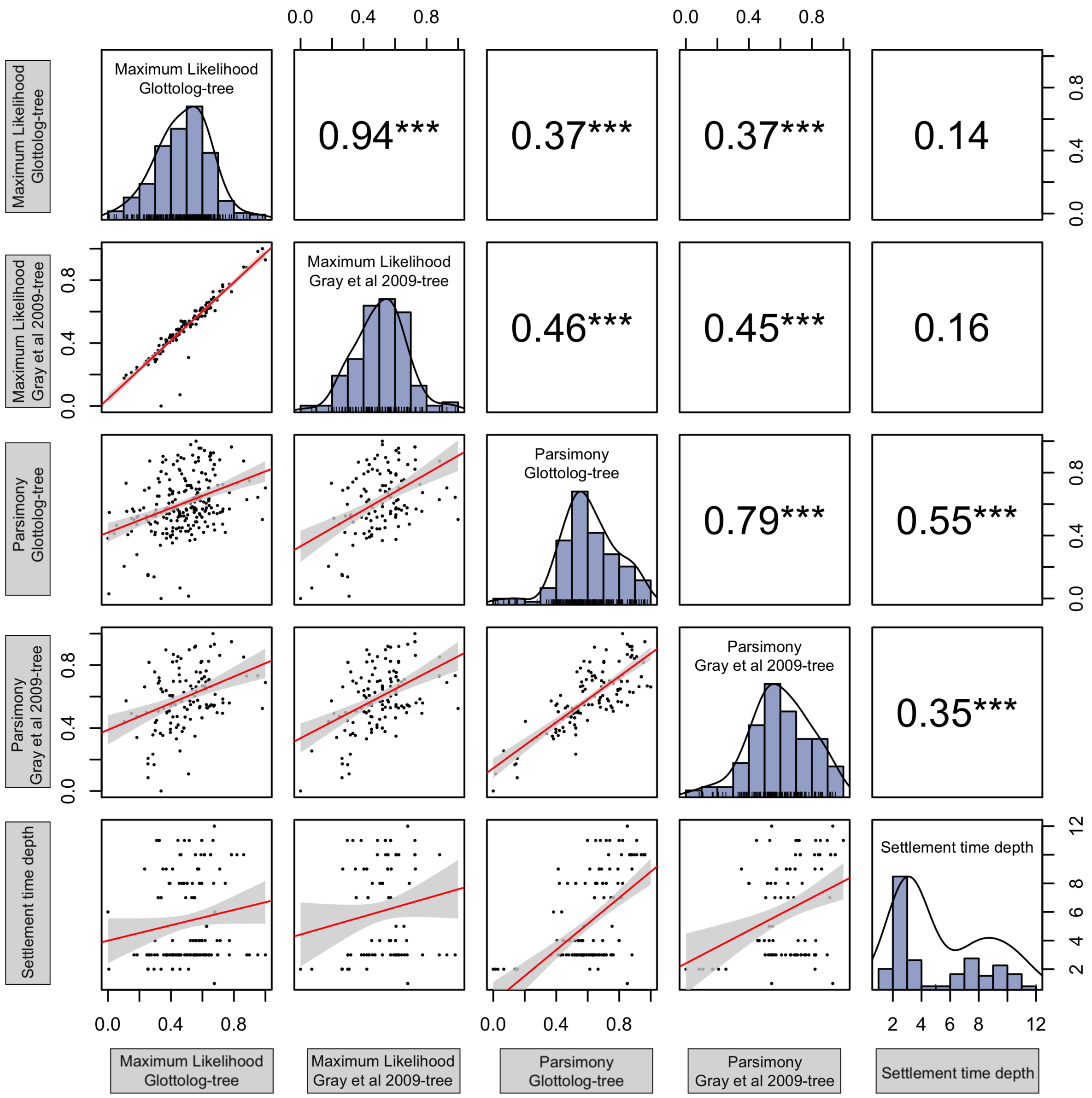


Figure 11: Scatterplot matrix of correlations between conservatism scores over the four different sets of results, and settlement order.

There is a very strong correlation between the two ML results (0.94), even though the two tree topologies are quite different. This is noticeable, we might expect that analysis using such different trees would generate more different results. The correlation between the two parsimony results is also strong (0.79), but not as strong. The parsimony results also correlate with the settlement time, the Glottolog and Maximum Parsimony results moderately (0.55) and the Gray et al 2009 and Maximum Parsimony weakly (0.35). This indicates that the higher the Parsimony cost (i.e. rate of change), the more likely a language is to be spoken on an island which has been settled more recently. This correlation, though present, is not strong and needs to be further investigated. It is possible that the correlation between the Maximum Parsimony conservatism scores and settlement time depth is related to the aforementioned relationships to degree of nesting (i.e. number of nodes between root and tip).

The fact that the two Maximum Likelihood results correlate more with each other than the two Maximum Parsimony results does suggests that it is a more robust methodology.

In summary, the results differ between the four methods, and these differences reveal some of the consequences behind the methodologies. Since Maximum Parsimony is more dependent on the number of nodes along a lineage, the findings show that it adds little information beyond that. The interpretation of the precise results depends on how we choose to interpret what it means that the lineages of certain languages involve more splitting events than others. Regardless, the findings indicate that Central Pacific languages are *not* especially conservative structurally, that the Bismarck languages are indeed particularly conservative and that the "aberrant" island group of Southern Vanuatu appears especially innovative structurally. For future studies, it is possible to investigate the drivers behind conservatism in a similar manner to how language richness is explored in chapter **??**.

4 Conclusions

In this chapter, we have investigated the history of structural features of Oceanic languages to examine how computational reconstructive methods compare to reconstructions by historical linguists (including contributing to the debate on Proto-Polynesian alignment), the stability of features and conservatism of languages of the region.

We have found that computational reconstructions show a high degree of concordance with reconstructions from expert historical linguists. Reconstructions by both Maximum Parsimony and Maximum Likelihood agreed to a very large extent with the findings from historical linguistics, but Maximum Likelihood was most likely to be "hesitant" and posit half/half states.

Within Oceanic historical linguistics, there exists a debate regarding the nature of the alignment system of Proto-Polynesian. The results of this study support the analysis that it was ergative. However, since the computational reconstructions are unable to take into account considerations of plausibility, which is the main difference between the different proposals, this cannot be taken as hard evidence.

One of the aims of this study was also to explore stability of features. The results reaffirmed previous studies which have found rare features among the most stable (Dediu & Cysouw 2013). We also expected features that pertain to paradigmatic distinctions to be more likely to be stable. However, this was not the case. The results from Maximum Parsimony and Maximum Likelihood were overall not in agreement.

More work is needed to explore the hypothesis that certain diagnostic structural features can track history at a deeper time depth than basic vocabulary (c.f. Nichols 1998). The results of this study were inconclusive, possibly because Oceanic is not deep enough for this kind of analysis.

The last part of the study investigated rates of change over languages of the region. Languages of the Bismarck archipelago are the most conservative overall[19]. The distribution of conservatism supports the theory that we are more likely to find conservative languages at the origin of the spread (Lynch et al. 2002:119). Contrary to other findings in historical linguistics concerning lexical conservatism, languages of the Central Pacific were not the most conservative and potentially even among the most progressive. Southern Vanuatu languages were also found to be particularly progressive structurally.

[19] Disregarding Temotu which most likely has a high conservatism score because of the outgroup-rooting.

Linguistic diversity can mean many things. This dissertation has explored four different ways of conceiving of diversity: number of languages (chapter ??), pairwise dissimilarity between languages (chapter ??), rate of change along a tree (chapter ??) and language internal variation (chapter ??). Together the insights from the different chapters inform our understanding of the dynamics of linguistic diversity.

Chapter ?? explored different environmental and social factors which may contribute to language diversification in Remote Oceania. The models included rainfall (mean and seasonality), temperature (mean and seasonality), isolation, size (area and shoreline), time of settlement and political complexity. Two different ways of grouping islands were explored (overnight voyage or shared language), and in both sets of findings island size, time depth and political complexity were significant factors in predicting the number of languages per island group. This confirms earlier research which suggests that political complexity correlates inversely with the number of languages in a given place (c.f. Pawley (2007) and Currie & Mace (2009)).

We need not interpret these findings as directly showing that pyramidal chiefly power in itself generates homogeneity (i.e. that chiefs enforce homogeneity explicitly). Rather, such network structures may encourage and make possible more distant interactions across space and make it more likely that community members see themselves as part of the same larger community and therefore converge to a greater extent.

This can be compared to Duhamel's study of the Raga community in North Pentecost (Vanuatu). In her thesis, she argues that the high linguistic uniformity within the community can be explained in part by the density and multiplex ties of the members of the community at large (not solely relatives and close neighbours) (Duhamel 2020). The Raga political structure is different from most Polynesian societies and falls under what is commonly labelled as "grade-taking" societies (Bonnemaison 1996). The majority of societies of this type are classified as level one in the scale of political complexity in the Ethnographic Atlas (section ??). It is clear that within Vanuatu there exist many different types of power structures. It is likely that the scale of political complexity from the Ethnographic Atlas does not capture the nuances in Vanuatu in fine enough detail to capture this, and that much can be learned by a closer study of the differences within Vanuatu between how power is enacted and the ties between members of the same community (c.f. Grace (1992)'s observations of loosely tied networks of New Caledonia).

The findings of chapter ?? need to be further substantiated, in particular in regards to the isolation metric and other ways of managing the skewed distribution of languages in the analysis. More sophisticated methods of teasing out the causality of factors in language diversification are also necessary to rule out spurious correlations (c.f. Roberts (2018) and Coelho et al. (2019)).

Chapter ?? explored the dissimilarity of languages in the region. The findings showed that structural data contains more conflicting signal than lexical data. Pairwise distances were calculated between languages based on the dissimilarity between their profiles in the Grambank database and their shared cognates in

Austronesian Basic Vocabulary Database (ABVD). The Grambank structural data consists of a set of 201 binary features (see appendix B) and the ABVD lexical dataset of 210 basic concepts coded for cognacy. A small pairwise distance indicates that the languages are similar structurally/share many cognates; a large pairwise distance indicates that they are very dissimilar. The minimum possible distance is 0 and the maximum 1. Overall, the pairwise distances of languages in Remote Oceania based on structural data ranged between 0.20 and 0.30 whereas the lexical distances ranged from 0.20 to 0.80 (sections **??** and **??**). This means that the languages overall were more similar structurally than they were lexically.

The distances were also compared to known language family trees. The lexical distances between languages showed a stronger correlation to the distances in the trees than did the structural distances (section **??**). There is a stronger phylogenetic 'signal' in the lexical data compared to the structural (c.f. Greenhill et al. (2017)). Furthermore, measurements of conflicting signal in the data (delta scores and Q-residuals in neighbour-nets) showed that the lexical data was more tree-like than the structural. It is possible that the greater amount of conflicting signal and lower correlation with family trees in the structural data is due to the restricted design space of the dataset we used. Similarity in the lexical data necessarily implies inheritance whereas structural features of languages are subject to different evolutionary pressures (section **??**). The typological questionnaire used covers 'core' grammatical domains and is able to distinguish between language families (Skirgård et al. unpublished), but may not contain enough 'rare' features for teasing out more lower level subgroups.

We also sought to test if the island groups where so called "aberrant" Oceanic languages predominate (Temotu, New Caledonia and Southern Vanuatu) do indeed stand out in terms of their structural disparity and lexical divergence (Grace (1981), Grace (1992) and Pawley (2006)). The part of the definition of "aberrant" that was tested is if the languages from these island groups are on average especially distant from their Oceanic cousins and/or from Proto-Oceanic.

None of the island groups stood out as "aberrant" in terms of structural disparity, i.e. were especially distant from other Oceanic languages or Proto-Oceanic. Given the greater conflicting signal in the structural data compared to the lexical, and the fact that structure is most likely recruited as a marker of social indexing less often, this is expected. Pawley (2006:219) also notes that most languages in New Caledonia and Southern Vanuatu are not "atypical' Oceanic languages structurally. Most of the research that has been brought to bear on "aberrant" languages concerns systematic sound correspondences and cognates.

The island groups of Temotu and New Caledonia were clearly "aberrant" lexically. Southern Vanuatu is the third most "aberrant" island group in the sample, but it should be noted that it is only slightly more distant lexically from the rest of the Oceanic languages than Micronesian languages are. This confirms Pawley (2006)'s observation that "aberrant" Oceanic languages predominate in these island groups.

Further research is needed here to complete the picture. The structural dataset used in this dissertation did not include data on phonological features. This is likely a fruitful venue for future research, both since phonology is potentially able to track deep history (c.f. Evans (2019)) and in order to explore the "aberrant" languages more fully. The languages of Southern Vanuatu and New Caledonia are less well-described than the other island groups and therefore not as well represented in the sample of this study. More research is needed into these languages, in particular their grammar.

Chapter **??** concerned ancestral state reconstruction of structural features of Oceanic languages by computational phylogenetic methods. The chapter aimed at answering three questions: do computational methods reconstruct the same structural features for proto-languages as classical historical linguists, are certain structural features more stable than others and are some island groups more conservative on average than others? Two different methods were used — Maximum Parsimony and Maximum Likelihood — and two different trees — Glottolog 4.0 (Hammarström et al. 2019) and Gray et al. (2009). The findings show that indeed, computational methods often reconstruct the same structure as traditional historical linguistics. The results concerning the stability of features across the methods and trees was not conclusive.

Conservatism of the Maximum Parsimony analysis was measured as the average number of changes from the root (proto-Oceanic) to the tip (a language) given the Maximum Parsimony solution of ancestral state. For the Maximum Likelihood analysis conservatism is measured as the average rate of change[20]. The most structurally conservative languages were found in the Bismarck archipelago and Temotu[21]. The Bismarck archipelago is also deemed the most likely location for the proto-Oceanic homeland (Lynch et al. 2002:97). Central Pacific was not, contrary to what might be expected based on research of lexical conservatism, among the most conservative languages. Instead languages in Central Pacific together with Southern Vanuatu were among the least conservative structurally.

The findings of conservatism of languages in terms of structural change in chapter **??** also revealed some key differences between the two methodologies used in the analysis: Maximum Parsimony and Maximum Likelihood. The Maximum Parsimony results were more dependent on number of splits posited along the lineages in the trees than was the Maximum Likelihood analysis. The Maximum Likelihood scores were more correlated across the two trees used, which indicates a greater robustness.

The findings from chapter **??** differ from chapter **??** primarily in the position of Central Pacific in regards to structural disparity. In chapter **??** there was overall little difference in the distances internally within each island group or their average distance to Proto-Oceanic. However, in chapter **??** the different island groups do differ in their average conservatism. Distances in chapter **??** were calculated pairwise directly

[20] For both methods it is possible for a structural feature to emerge and disappear several times along a lineage. Note that this is *not* the case for lexical cognates, unless there is borrowing involved.

[21] The conservatism of Temotu should be taken with a grain of salt. The analysis of conservatism required that the trees be re-rooted and they were rooted with Nanggu (a language of Temotu) as an outgroup. This in combination with the fact that the analysis did not include archaeological dates means that the conservatism of Temotu is most likely inflated.

between each pair of languages with no regard to the potential genealogical relationship between them. Yapese was compared directly to Proto-Oceanic just as Tongan was, even though most family trees of Austronesian have very few or no intermediate nodes between Proto-Oceanic and Yapese and many more between Proto-Oceanic and Tongan. In chapter **??** we harnessed the power of trees in our analysis and reconstructed states for the intermediate nodes and measured the change along this path, instead of direct pairwise distances. This is why the results differ. Once the reconstruction of the intermediate nodes is taken into account, the structural changes that have led to Central Pacific are greater than if we compare the raw number of changes directly between Central Pacific languages and Proto-Oceanic. One might say that more has happened along the road than would be revealed by a direct comparison of the two end-points.

In chapter **??** we took a closer look at one of the politically complex societies of Remote Oceania — Sāmoa. Polynesia is generally a place with low amounts of language splitting and the findings of chapter **??** suggests that this is related to higher levels of political complexity. This may lead us to believe that there is also less variation *within* languages in this region. Sāmoa has been described as a homogeneous society and language by scholars such as Mead (1937) and Turner (1884). The lack of variation within Sāmoan has been attributed to central governance and greater mobility. Such theories would indeed be in line with our findings in chapter **??** which suggest that political complexity retards language diversification.

However, upon closer inspection we learn that Sāmoan political history is dramatic and that central governance consisting of one high chief ruling over the entire archipelago was a rare occurrence historically. Furthermore, there *is* variation within the Sāmoan language phonologically, lexically and structurally. Even so, the variation that is found in Sāmoan is almost exclusively *social* as opposed to regional (meaning that the variants are geographically ubiquitous and are instead delimited by style, register and social setting). We can hypothesise that linguistic variation which is mostly socially conditioned does not lead to full language split since many people are likely to be knowledgeable in several of the social variants and there would be little (if any) separation of parts of the speech community.

Despite the lack of archipelago-wide rule, it appears that the government of *village districts* in Sāmoa has been historically stable. The village districts have also collaborated and had significant exchange and ties to other village districts even if they have not been continuously co-ruled. Perhaps a language community need not be fully state-like in order for language splitting to be retarded — a little political structure can go a long way?

Where does all of this leave us? Some of these findings confirm what previous literature had indicated: political complexity matters in the diversification of languages in the region, structure correlates less with known family trees of languages than lexicon does, Temotu and New Caledonia are peculiar linguistically and Bismarck is the most likely homeland of the Oceanic subgroup. However, there are certain details of these findings that warrant more discussion, in particular the languages of the Central Pacific region.

Previous research has found languages of the Central Pacific to be particularly *conservative* lexically (Blust 2000:323) and in the results of chapter **??** they have the lowest average distance to Proto-Oceanic lexically (section **??**). And yet, the analysis in chapter **??** found the languages of the region to have more than average structural change, as measured by computational ancestral reconstruction means.

Central Pacific is one of the most recently settled regions of Remote Oceania (section **??**). Many of the languages there have had less time to diversify in place than the languages in Vanuatu. And yet, chapter **??** showed that Northern Vanuatu languages are *more* similar to each other structurally than are the languages of the islands of Central Pacific.

How does this go together? Might an explanation for the fact that Central Pacific is more progressive structurally than it is lexically once again lie in networks and political complexity? François (2011) argues that the social networks of Northern Vanuatu encourage lexical divergence and that structural convergence is a result of contact and multilingualism. Due to the fact that structure may be less accessible to conscious observation of the speaker (c.f. Silverstein (1981) and Pawley (2006:237-238)), lexical items may be more likely to be recruited as markers of identity. François (2011) argues that in Northern Vanuatu words are understood as more emblematic of place, but that grammar is not and is therefore able to diffuse across networks more freely. Another possible interpretation of the findings by François (2011) is that the languages *continued* to be similar structurally, but diverged more radically lexically due to social pressures.

Ellison & Miceli (2017) have also found that individual bilingual speakers will actively avoid words that are common to the two languages they know when asked to name a certain item and instead choose a word unique to either language. If faced with a situation where there are words in language A and B that are similar, they will choose another word in language A that is not similar to language B, even if it is less common. For example, Dutch-English bilinguals chose the word “picture” in English more often than English monolinguals who instead chose “photo” when presented with a stimuli which appeared like a proto-typical photograph. The authors argue that the Dutch-English bilinguals avoid saying “photo” in English because it is “too close” to “foto” in Dutch which denotes the same meaning. The degree to which this occurs varies with the pragmatic need for the speakers to monitor for language mixing lest it hinders comprehension, but other social pressures linked to emblematic usage of language are also relevant (Ellison & Miceli 2017:277). If avoiding shared vocabulary is a common phenomenon in multilingual speakers it would spur lexical divergence also in language communities which are very much in contact.

Part of the explanation for the lack of language splitting in Central Pacific is argued to be the rise of powerful chiefs and maintenance of long distance sailing networks which encourage homogeneity (Pawley (2007) and chapter ??). Chapter ?? indicated the cultural unity of the Sāmoan islands, despite the islands' tumultuous political history. Is it possible that similar sentiments were at play across larger distances?

A speculation based on the findings in this dissertation is that there existed in the Central Pacific for a long time a sense of cultural unity across island groups. Lexical change would therefore slow down, because of its emblematic nature. However, this was not the case for structural change. By the natural drift of language differentiation through isolation, the languages of Central Pacific changed more in their grammars than in their lexicon. Structural changes were not noticed, and were not understood as signs of cultural division. In contrast, the lexicon remained intact for longer due to strong cultural affinity. The rate of structural change would not be extremely high compared to other island groups, but it would not be as low as the rate of lexical change. If words are more emblematic of cultural affinity and structural features of languages are more likely to go unnoticed by the conscious mind, then it is likely that if the pre-historic societies of the Central Pacific kept in contact over large distances and viewed themselves as culturally connected (c.f. *Hawaiki* (Kirch & Green 2001)) this would be the result — lexicon being more conservative than grammars. This is a speculation based on the findings of these studies, and definitely needs testing and further exploration.

It is also possible that the slowed-down rate of lexical change is a product of maintenance of cultural affinity within each island group only, and not necessarily wider networks over multiple island groups. This could also give a similar effect in closely related languages. We saw in the case of Sāmoan that the archipelago has a history of political division internally, and yet little of this is reflected in regional language variation. Perhaps a little hierarchical structure, "just" up to village district level, can go a long way to retard change and splitting? If this effect is scaled up, maybe it can be felt even at a macro-level?

Both of these scenarios would result in the inverse of the Northern Vanuatu case — lexical divergence is slowed but structural change goes on diversifying at a "normal rate" (Central Pacific) as opposed to lexical change accelerating and structure converging (Northern Vanuatu).

This speculation about the cause of this mismatch between lexical and structural conservatism in Central Pacific may prove to be incorrect, but it does not invalidate the point that the dynamics of change may differ depending on the part of language studied — structure is different from lexicon and we should not assume that they follow similar paths or processes.

We still have much to learn about the processes of language diversification. I hope that this dissertation has brought more light on a few specific issues with particular reference to Remote Oceania which may also prove insightful in other areas of the world and other disciplines.

References

Beaulieu, Jeremy M., Jeffrey C. Oliver & Brian O'Meara. 2017. Package 'corHMM'. URL https://CRAN.R-project.org/package=corHMM.

Bloomfield, Leonard. 1933. *Language*. Chicago and London: The University of Chicago Press.

Blust, Robert. 1981. Variation in Retention Rate among Austronesian Languages. Talk given at *The Third International Conference on Austronesian Linguistics*, Bali.

Blust, Robert. 1996. The Neogrammarian Hypothesis and Pandemic Irregularity. In Mark Durie & Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 135–156. Oxford: Oxford University Press.

Blust, Robert. 2000. Why Lexicostatistics Doesn't Work: The 'Universal Constant' Hypothesis and the Austronesian Languages. In Colin Renfrew (ed.), *Time Depth in Historical Linguistics*, vol. 2, 311–331. Cambridge, UK: The McDonald Institute for Archaeological Research.

Blust, Robert. 2014. Some Recent Proposals Concerning the Classification of the Austronesian Languages. *Oceanic Linguistics* 53(2). 300–391.

Blust, Robert A. 2004. *t to k: An Austronesian Sound Change Revisited. *Oceanic Linguistics* 43(2). 365–410.

Blust, Robert A. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Blust, Robert A. & Victoria Chen. 2017. The Pitfalls of Negative Evidence: 'Nuclear Austronesian', 'Ergative Austronesian', and Their Progeny. *Language and Linguistics* 18(4). 577–621.

Bonnemaison, Joël. 1972. Système de Grades et Différences Régionales En Aoba (Nouvelles Hébrides). *Cahiers ORSTOM. Série Sciences Humaines* 9(1). 87–108.

Bonnemaison, Joël. 1996. Graded Societies and Societies Based on Title: Forms and Rites of Traditional Power in Vanuatu. In *Arts of Vanuatu*, 200–216. University of Hawaii Press.

Brooks-Bartlett, Jonny. 2018. Probability Concepts Explained: Maximum Likelihood Estimation. *Towards Data Science* URL https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1.

Campbell, Lyle. 1996. On Sound Change and Challenges to Regularity. In Mark Durie & Malcom Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 72–89. Oxford: Oxford University Press.

Carson, Mike T. 2012. Recent Developments in Prehistory: Perspectives on Settlement Chronology, Inter-Community Relations, and Identity Formation. In Richard Feinberg & Richard Scaglion (eds.), *The Polynesian Outliers: State of the Art* (Ethnology Monographs 21), 27–48. United States of America: Department of Anthropology, University of Pittsburgh.

Chung, Sandra. 1977. Review of Clark, R: Aspects of Proto-Polynesian Syntax. *The Journal of the Polynesian Society* 86(4). 537–540.

Chung, Sandra. 1978. *Case Marking and Grammatical Relations in Polynesian Languages*. Austin: University of Texas.

Clark, D Ross. 1976. *Aspects of Proto-Polynesian Syntax*, vol. 6. Linguistic Society of New Zealand.

Coelho, Marco Túlio Pacheco, Elisa Barreto Pereira, Hannah J. Haynie, Thiago F. Rangel, Patrick H Kavanagh, Kathryn R. Kirby, Simon J. Greenhill, Claire Bowern, Russell D. Gray, Robert K. Colwell, Nicholas Evans & Michael C. Gavin. 2019. Drivers of Geographical Patterns of North American Language Diversity. *Proceedings of the Royal Society B: Biological Sciences* 286(1899). 20190242. doi: 10.1098/rspb.2019.0242.

Crowley, Terry. 1985. Common Noun Phrase Marking in Proto-Oceanic. *Oceanic Linguistics* 24(1/2). 135–193.

Cunningham, Clifford W, Kevin E Omland & Todd H Oakley. 1998. Reconstructing Ancestral Character States: A Critical Reappraisal. *Trends in Ecology & Evolution* 13(9). 361–366.

Currie, T.E. & R. Mace. 2009. Political Complexity Predicts the Spread of Ethnolinguistic Groups. *PNAS - Proceedings of the National Academy of Sciences* 106(18). 7339–7344.

Dediu, Dan & Michael Cysouw. 2013. Some Structural Aspects of Language Are More Stable than Others: A Comparison of Seven Methods. *PloS ONE* 8(1). e55009.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL http://wals.info/.

Duhamel, Marie-France. 2020. *Variation in Raga: A quantitative and qualitative study of the language of North Pentecost, Vanuatu*. Department of Culture, History and Language, College of Asia-Pacific, The Australian National University. PhD Thesis (under examination).

Ellis, S James. 2012. *Saipan Carolinian, One Chuukic Language Blended from Many*. University of Hawaii at Manoa. PhD Thesis.

Ellison, T Mark & Luisa Miceli. 2017. Language Monitoring in Bilinguals as a Mechanism for Rapid Lexical Divergence. *Language* 93(2). 255–287.

Evans, Bethwyn. 2001. *A Study of Valency-Changing Devices in Proto Oceanic*. Research School of Pacific and Asian Studies, The Australian National University. PhD Thesis.

Evans, Nicholas. 2019. Australia and New Guinea: Sundered Hemi-Continents of Sound. In Sasha Calhoun, Paul Warren, Paola Escudero & Marija Tabain (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 16–19. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Felsenstein, Joseph. 2004. *Inferring Phylogenies*, vol. 2. Sunderland, MA: Sinauer Associates.

Fisher, Ronald A. 1912. On an Absolute Criterion for Fitting Frequency Curves. *Messenger of Mathematics* 41. 155–156.

François, Alexandre. 2011. Social Ecology and Language History in the Northern Vanuatu Linkage - A Tale of Divergence and Convergence. *Journal of Historical Linguistics* 1(2). 175–246.

Goddard, Ives. 1993. Contamination in Morphological Change in Algonquian Languages. Talk given at Henk Aertsen & Robert J Jeffers (eds.), *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, New Brunswick, 14-18 August 1989*, 129–140. John Benjamins Publishing.

Grace, George W. 1981. Indirect Inheritance and the Aberrant Melanesian Languages. In J. Hollyman & A. Pawley (eds.), *Studies in Pacific Languages and Cultures*, 255–268. Auckland: Linguistic Society of New Zealand.

Grace, George W. 1992. How Do Languages Change? (More on "Aberrant" Languages). *Oceanic Linguistics* 31(1). 115–130. URL http://www.jstor.org/stable/3622968.

Grafen, Alan. 1989. The Phylogenetic Regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 326(1233). 119–157.

Gray, R.D., A.J. Drummond & S.J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* 323. 479–483.

Gray, Russell D, Quentin D Atkinson & Simon J Greenhill. 2011. Language Evolution and Human History: What a Difference a Date Makes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366(1567). 1090–1100.

Greenhill, Simon J., Robert Andrew Blust & Russell D. Gray. 2008. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics* 4. 271–283.

Greenhill, Simon J & Ross Clark. 2011. POLLEX-Online: The Polynesian Lexicon Project Online. *Oceanic Linguistics* 50(2). 551–559.

Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen Levinson & Russell Gray. 2017. The Evolutionary Dynamics of Language Systems. *PNAS - Proceedings of the National Academy of Sciences* 114(42). E8822–E8829.

Hale, Kenneth. 1968. Review of Hohepa 1967. *Journal of the Polynesian Society* 77(1). 83–99.

Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. Glottolog/Glottolog: Glottolog Database 4.0. URL https://doi.org/10.5281/zenodo.3260726.

Hohepa, Patrick W. 1967. *A Profile Generative Grammar of Maori.* Indiana University Publications in Anthropology and Linguistics, Memoir 20.

Hohepa, Patrick W. 1969. The Accusative-to-Ergative Drift in Polynesian Languages. *Journal of the Polynesian Society* 78(3). 295–329.

Intoh, Michiko. 2008. Ongoing Archaeological Research on Fais Island, Micronesia. *Asian Perspectives* 47(1). 121–138.

Intoh, Michiko & Rintaro Ono. 2007. Reconnaissance Archaeological Research on Tobi Island, Palau. *People and Culture in Oceania* 22. 53–83.

Jäger, Gerhard & Johann-Mattis List. 2018. Using Ancestral State Reconstruction Methods for Onomasiological Reconstruction in Multilingual Word Lists. *Language Dynamics and Change* 8(1). 22–54.

Jombart, Thibaut, Stéphane Dray & Anders Ellern Bilgrau. 2017. Package 'Adephylo'. URL https://cran.r-project.org/web/packages/adephylo/index.html.

Jonsson, Niklas. 1998. Det Polynesiska Verbmorfemet - Cia; Om Dess Funktion i Samoanska.

Joy, Jeffrey B, Richard H Liang, Rosemary M McCloskey, T Nguyen & Art FY Poon. 2016. Ancestral Reconstruction. *PLoS Computational Biology* 12(7). e1004763.

Kikusawa, Ritsuko. 2001. Rotuman and Fijian Case-Marking Strategies and Their Historical Development. *Oceanic Linguistics* 40(1). 85–111.

Kikusawa, Ritsuko. 2002. *Proto Central Pacific Ergativity: Its Reconstruction and Development in the Fijian, Rotuman and Polynesian Languages.* Canberra: Pacific Linguistics.

Kikusawa, Ritsuko. 2006. On the Development of Number Systems in Oceanic Pronouns. Talk given at *6th International Conference on Oceanic Linguistics (COOL6)*, Port Vila, Vanuatu.

Kirby, KR, R. D. Gray, S. J. Greenhill, F. M. Jordan, S. Gomes-Ng, H-J Bibiko, Damián E. Blasi, Carlos A. Botero, Claire Bowern, Carol R. Ember, Dan Leehr, Bobbi S. Low, Joe McCarter, William Divale & Michael C. Gavin. 2018. D-PLACE/dplace-data: D-PLACE – the Database of Places, Language, Culture and Environment (Version v2.0.1) [Data set]. URL http://doi.org/10.5281/zenodo.1466634.

Kirch, Patrick Vinton. 2017. *On the Road of the Winds: An Archaeological History of the Pacific Islands before European Contact.* Univ of California Press.

Kirch, Patrick Vinton & Roger C. Green. 2001. *Hawaiki, Ancestral Polynesia: An Essay in Historical Anthropology.* Cambridge University Press.

Kirch, PV. 2012. Baseline Prehistory. The Polynesian Outliers: Continuity, Change, and Replacement. In Richard Feinberg & Richard Scaglion (eds.), *The Polynesian Outliers: State of the Art* (Ethnology Monographs 21), 17–26. United States of America: Department of Anthropology, University of Pittsburgh.

Koerner, EFK. 1997. Notes on the History of the Concept of Language as a System 'Où Tout Se Tient'. *Linguistica atlantica* 19. 1–20.

Levin, Maureece J, Katherine Seikel & Aimee Miles. 2019. A Partial Chronological Sequence of Human Habitation for Pingelap Atoll (Pohnpei State, Federated States of Micronesia). *Radiocarbon* 61(3). 765–776. doi: 10.1017/RDC.2019.30.

Lima-Ribeiro, Matheus S., Sara Varela, Javier González-Hernández, Guilherme de Oliveira, José Alexandre F Diniz-Filho & Levo Caroma Terribile. 2015. ecoClimate: A Database of Climate Data from Multiple Models for Past, Present, and Future for Macroecologists and Biogeographers. *Biodiversity Informatics* 10. 1–21.

Louca, Stilianos & Michael Doebeli. 2017. Efficient Comparative Phylogenetics on Large Trees. *Bioinformatics* 34(6). 1053–1055.

Lynch, John, Malcolm Ross & Terry Crowley. 2002. Internal Subgrouping. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic Languages* Curzon Language Family Series, 92–120. Richmond: Curzon.

Lynch, John, Malcolm Ross & Terry Crowley. 2011. Proto Oceanic. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic Languages* Curzon Language Family Series, 54–91. Richmond: Curzon 2nd edn.

Maclaurin, James & Kim Sterelny. 2008. *What Is Biodiversity?* University of Chicago Press.

Marck, Jeffrey C. 2000. Polynesian Languages. In J. Garry & C. Rubino (eds.), *Facts About the World's Languages: An Encyclopaedia of the World's Major Languages, Past and Present*, New York: H.W. Wilson.

McMahon, April & Robert McMahon. 2006. Why Linguists Don't Do Dates: Evidence from Indo-European and Australian Languages. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic Methods and the Prehistory of Languages*, 153–160. McDonald Institute for Archaeological Research, Cambridge.

Mead, Margaret. 1937. The Samoans. In Margaret Mead (ed.), *Cooperation and Competition among Primitive Peoples*, 282–312. McGraw-Hill Book Company.

Napolitano, Matthew F., Scott M. Fitzpatrick, Geoffrey Clark & Jessica H. Stone. 2019. New Investigations of Early Prehistoric Settlement on Yap, Western Caroline Islands. *The Journal of Island and Coastal Archaeology* 14(1). 101–107. doi: 10.1080/15564894.2017.1335661.

Nichols, Johanna. 1998. The Origin and Dispersal of Languages: Linguistic Evidence. In Nina G. Jablonski & Leslie C. Aiello (eds.), *The Origin and Diversification of Language*, 127–170. California Academy of Sciences San Francisco, CA.

Pagel, Mark. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255(1342). 37–45.

Paradis, Emmanuel, Julien Claude & Korbinian Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* 20(2). 289–290.

Pawley, Andrew. 1966. Polynesian Languages: A Subgrouping Based on Shared Innovations in Morphology. *The Journal of the Polynesian Society* 75(1). 39–64.

Pawley, Andrew. 1970. Grammatical Reconstruction and Change in Polynesia and Fiji. In SA Wurm & DC Laycock (eds.), *Studies in Honour of Arthur Capell*, 301–368. Canberra: Pacific Linguistics.

Pawley, Andrew. 1973. Some Problems in Proto-Oceanic Grammar. *Oceanic Linguistics* 12(1/2). 103–188.

Pawley, Andrew. 2005. The Meaning(s) of Proto Oceanic* Panua. In D.A. Counts C. Gross, H.D. Lyons (ed.), *A Polymath Anthropologist: Essays in Honour of Ann Chowning* (Research in Anthropology and Linguistics Monograph 6), 133–145. Auckland: Department of Anthropology, University of Auckland.

Pawley, Andrew. 2006. Explaining the Aberrant Austronesian Languages of Southeast Melanesia: 150 Years of Debate. *The Journal of the Polynesian Society* 115(3). 215–258.

Pawley, Andrew. 2007. Why Do Polynesian Island Groups Have One Language and Melanesian Island Groups Have Many? Patterns of Interaction and Diversification in the Austronesian Colonisation in Remote Oceania. Talk given at *Workshop on Migration*, Ile de Porquerolles, France.

Pawley, Andrew. 2009. The Role of the Solomon Islands in the First Settlement of Remote Oceania: Bringing Linguistic Evidence to an Archaeological Debate. In Alexander Adelaar & Andrew Pawley (eds.), *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, 515–540. Canberra: Pacific Linguistics.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. URL https://www.R-project.org/.

Revell, Liam J. 2012. Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things). *Methods in Ecology and Evolution* 3(2). 217–223.

Revell, Liam J. 2014. Ancestral State Reconstruction. Talk given at *AnthroTree Workshop 2014*, Duke University, Durham, NC. URL http://www.phytools.org/anthrotree/ancestral-states.pdf.

Rieth, Timothy M. & Ethan E. Cochrane. 2018. The Chronology of Colonization in Remote Oceania. In Ethan E. Cochrane & Terry L. Hunt (eds.), *The Oxford Handbook of Prehistoric Oceania*, Oxford: Oxford University Press. URL https://doi.org/10.1093/oxfordhb/9780199925070.013.010.

Roberts, Seán. 2018. Robust, Causal, and Incremental Approaches to Investigating Linguistic Adaptation. *Frontiers in Psychology* 9. doi: 10.3389/fpsyg.2018.00166.

Ross, Malcolm. 1996. Contact-Induced Change and the Comparative. In Mark Durie & Malcolm Ross (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, 180–217. Oxford: Oxford University Press on Demand.

Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 1998. *Material Culture*, vol. 1 The Lexicon of Proto Oceanic: The Culture and Environment of Ancestral Oceanic Society. Canberra: ANU Electronic Press.

Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2007. *The Physical Environment*, vol. 2 The Lexicon of Proto Oceanic: The Culture and Environment of Ancestral Oceanic Society. Canberra: ANU Electronic Press 2nd edn.

Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2008. *Plants*, vol. 3 The Lexicon of Proto Oceanic: The Culture and Environment of Ancestral Oceanic Society. Canberra: ANU Electronic Press.

Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2011. *Animals*, vol. 4 The Lexicon of Proto Oceanic: The Culture and Environment of Ancestral Oceanic Society. Canberra: ANU Electronic Press.

Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2016. *People: Body and Mind*, vol. 5 The Lexicon of Proto Oceanic: The Culture and Environment of Ancestral Oceanic Society. Canberra: ANU Electronic Press.

Ross, Malcolm D. 2004. The Morphosyntactic Typology of Oceanic Languages. *Language and Linguistics* 5(2). 491–541.

Ross, Malcolm D. 2007. Two Kinds of Locative Construction in Oceanic Languages: A Robust Distinction. In Jeff Siegel, John Lynch & Diana Eades (eds.), *Language Description, History and Development: Linguistic Indulgence in Memory of Terry Crowley*, 281–295. John Benjamins Publishing Company.

Sankoff, David. 1975. Minimal Mutation Trees of Sequences. *SIAM Journal on Applied Mathematics* 28(1). 35–42.

Sapir, Edward. 1916. *Time Perspective in Aboriginal American Culture: A Study in Method*, vol. 90. Government Printing Bureau.

Schliep, Klaus, Alastair Potts, Tim W. White, Cyrill Stachniss, Michelle Kendall, Keren Halabi, Richel Bilderbeek & Kristin Winchell. 2019. Package 'Phangorn'. URL https://CRAN.R-project.org/package=phangorn.

Scott-Phillips, Thomas C & Simon Kirby. 2010. Language Evolution in the Laboratory. *Trends in Cognitive Sciences* 14(9). 411–417.

Sheehan, Oliver, Joseph Watts, Russell D Gray & Quentin D Atkinson. 2018. Coevolution of Landesque Capital Intensive Agriculture and Sociopolitical Hierarchy. *Proceedings of the National Academy of Sciences* 115(14). 3628–3633.

Siewierska, Anna. 2013. Gender Distinctions in Independent Personal Pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. URL https://wals.info/chapter/44.

Silverstein, Michael. 1981. *The Limits of Awareness.* Southwest Educational Development Laboratory Austin, TX.

Skirgård, Hedvig, Hannah J. Haynie, Harald Hammarström, Martin Haspelmath, Damián E. Blasi, Jeremy Collins, Alena Witzlack-Makarevich, Jakob Lesage, Jay J. Latarche, Ger Reesink, Michael Dunn, Ruth Singer, Claire Bowern, Patience Epps, Jane Hill, Karolin Abbas, Nancy Bakker, Giulia Barbos, Robert Borges, Swintha Danielsen, Hilário de Sousa, Hugo de Vos, Luise Dorenbusch, Giada Falcone, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima Goodall, Andrew Harvey, Rebekah H. Hayes, Roberto Herrera, Nataliia Hübler, Biu Huntington-Rainey, Marilen Johns, Erika Just, Eri Kashima, Aarifah Khoodoruth, Carolina Kipf, Janina Klingenberg, Katerina Koti, Richard Kowalik, Olga Krasnoukhova, Nora Lindvall, Mandy Lorenzen, Tânia Martins, Saliha Muradoglu, Johanna Nickel, Cheryl Akinyi Oluoch, Jesse Peacock, India Pearey, Naomi Peck, Stephanie F. E. Petit, Sören Pieper, Daniel Prestipino, Linda Raabe, Amna Raja, Eloisa Ruppert, Rhiannon Schembri, Frederick Schmidt, Kristin Sverredal, Suzanne van der Meer, Tobias Weber, Jana Winkler, Tim Witte, Henry Wu, Stephanie Yam, Maisie Yong, Tessa Yuditha, Nicholas Evans, Robert Forkel, Stephen C. Levinson, Simon J. Greenhill, Quentin D. Atkinson & Russell D. Gray. unpublished. Grambank reveals global patterns in the structural diversity of the world's languages. Unpublished.

Stassen, Leon. 2011. Predicative Possession. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Munich: Max Planck Digital Library. URL http://wals.info/chapter/117.

The Grambank Consortium (ed.). 2021. *Grambank.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://grambank.clld.org, Accessed on 2021-07-01.

Thomason, Sarah Grey & Terrence Kaufman. 1992. *Language Contact, Creolization, and Genetic Linguistics.* Univ of California Press.

Turner, George. 1884. *Samoa: A Hundred Years Ago and Long before; Together with Notes on the Cults and Customs of Twenty-Three Other Islands in the Pacific.* London: Macmillan and Co.

Van Rijsbergen, Cornelis Joost. 1979. *Information Retrieval.* Butterworths.

Walworth, Mary. 2015. *The Language of Rapa Iti: Description of a Language in Change.* University of Hawaii at Mānoa. PhD Thesis.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686. doi: 10.21105/joss.01686.

Wilks, Samuel S. 1938. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* 9(1). 60–62.

Wilson, William H. 1982. *Proto-Polynesian Possessive Marking.* Canberra: Pacific Linguistics.

Wilson, William H. 2012. Whence the East Polynesians?: Further Linguistic Evidence for a Northern Outlier Source. *Oceanic Linguistics* 51(2). 289–359.

Yang, Ziheng. 2006. *Computational Molecular Evolution.* Oxford University Press.

Appendices

A Roboustness control with Glottolog-tree

We will be using two different trees of Austronesian languages, one from Glottolog (Hammarström et al. 2019)[22] and one from Gray et al. (2009). In order for the method to function, the trees need to have branch lengths and all splits need to be binary. The tree from Glottolog does not contain information about branch lengths (all branches are the same length) and is not binary, whereas the Gray et al. (2009) tree has both branch length information and is binary. A part of the Glottolog tree can be seen in Fig. 12a.

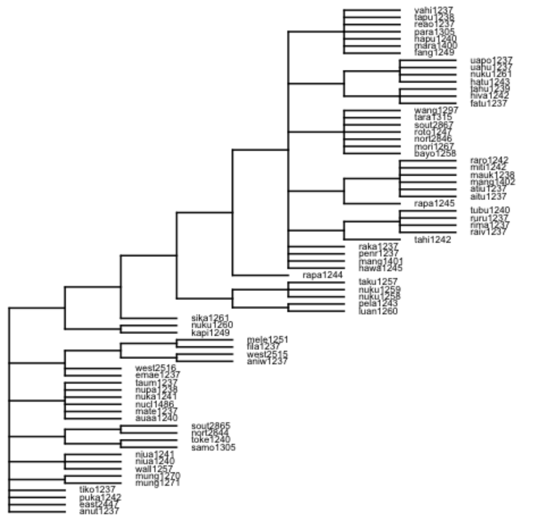
We can modify the Glottolog tree so that it is more appropriate for our analysis. I used the functions `multi2di` and `compute.brlen` in the R-package `ape` by Paradis et al. (2004) to binarise and compute branch lengths respectively. Fig. 12 illustrates the process using languages of the Nuclear Polynesian subgroup in Glottolog. The tree as stored in Glottolog only contains information about subgrouping; there is no data on tree depth and therefore no branch lengths (Fig. 12a.). The trees also contain non-binary splits. Fig. 12a shows a part of this tree . The function `multi2di` randomly resolves non-binary splits until only binary ones remain, this can be seen in (b). We use Grafen's transform (Grafen 1989) as implemented in `ape` to "pull down" all tips so that they all have the same length to the root (this is known as an 'ultrametric' tree), this can be seen in (c). The final result is accomplished by first binarising the tree and then computing branch lengths (d).

Family trees in historical linguistics are often under-specified for branch lengths. For example, the publications that underlie the Glottolog tree (Blust (2009, 2014) and Blust & Chen (2017)) do not contain information on the relative branch lengths, only which languages are in which subgroup. It has been said that "linguist don't do dates" (McMahon & McMahon (2006) and Gray et al. (2011)). This reflects the disappointment with early lexicostatistics which included failures at predicting language splitting events in time. However, it is possible to do relative branch lengths without dates. Given that all languages we now observe exist in the present, it seems fair to assume that an equal amount of time has passed from each of them to the proto-language. This supports the transformation as shown in Fig 12 from the subgrouping-only tree in Glottolog (a) to an ultrametric tree (c and d).

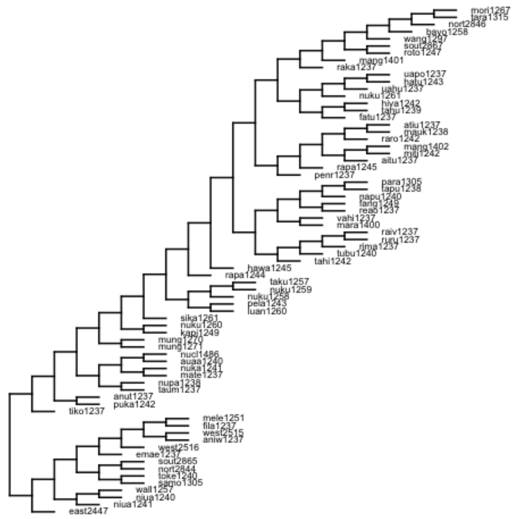
The Gray et al 2009-tree incorporates information from archaeology (c.f. Lynch et al. (2002:92)), this is one of the reasons why it has branch lengths.

B Grambank features

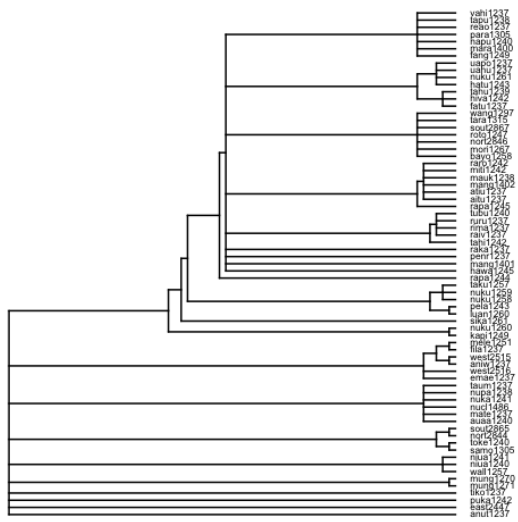
[22] The tree of Glottolog 4.0 is based on work by Blust (2009, 2014) and Blust & Chen (2017).



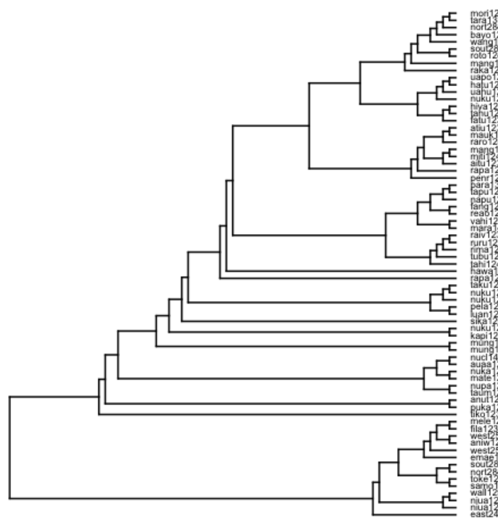
(a) Non-binary splits occur, all branches have the same length.



(b) Non-binary splits have been resolved to binary, all branches same length



(c) Non-binary splits occur, all branches have been scaled so that the length from the tip to the root is the same



(d) First, non-binary splits have been resolved to binary; second, all branches have been scaled so that the length from the tip to the root is the same.

Figure 12: Four trees of Nuclear Polynesian languages from Glottolog, illustrating the process of binarization and imputing branch lengths.

Table 9: Grambank features

Grambank ID	Abbreviation	Feature
GB020	ARTDef	Are there definite or specific articles?
GB021	ARTIndef	Do indefinite nominals commonly have indefinite articles?
GB022	ARTPre	Are there prenominal articles?
GB023	ARTPost	Are there postnominal articles?
GB024a	OrderNUMN	Is the order of numeral noun NUM-N?
GB024b	OrderNNUM	Is the order of numeral noun N-NUM?
GB025a	OrderDEMN	Is the order of adnominal demonstratives and nouns Dem-N?
GB025b	OrderNDEM	Is the order of adnominal demonstratives and nouns N-Dem?
GB026	ADJDiscont	Can adnominal property words occur discontinuously?
GB027	ComitConjDifferent	Are nominal conjunction and comitative expressed by different elements?
GB028	Clusivity	Is there a distinction between inclusive and exclusive?
GB030	PRO3PGender	Is there a gender distinction in independent 3rd person pronouns?
GB031	PRODualAug	Is there a dual or unit augmented form (in addition to plural or augmented) for all person categories in the pronoun system?
GB035	DEMDistContrast	Are there three or more distance contrasts in demonstratives?
GB036	DEMElevation	Do demonstratives show an elevation distinction?
GB037	DEMVisNonvis	Do demonstratives show a visible-nonvisible distinction?
GB038	DEMClassifier	Are there demonstrative classifiers?
GB039	NounNUMAllomorph	Is there nonphonological allomorphy of noun number markers?
GB041	NUMSupplNoun	Are there several nouns (more than three) which are suppletive for number?
GB042	SingularNoun	Is there productive overt morphological singular marking on nouns?
GB043	DualBound	Is there productive morphological dual marking on nouns?
GB044	PluralBound	Is there productive morphological plural marking on nouns?
GB046	AssocPlural	Is there an associative plural marker for nouns?
GB047	NMZActionState	Is there a productive morphological pattern for deriving an action/state noun from a verb?

Grambank ID	Abbreviation	Feature
GB048	NMZAgent	Is there a productive morphological pattern for deriving an agent noun from a verb?
GB049	NMZObject	Is there a productive morphological pattern for deriving an object noun from a verb?
GB051	GenderSex	Is there a gender/noun class system where sex is a factor in class assignment?
GB052	GenderShape	Is there a gender/noun class system where shape is a factor in class assignment?
GB053	GenderAnimacy	Is there a gender/noun class system where animacy is a factor in class assignment?
GB054	GenderPlants	Is there a gender/noun class system where plant status is a factor in class assignment?
GB057	NUMClassif	Are there numeral classifiers?
GB058	POSSClassifier	Are there possessive classifiers?
GB059	POSSAlienability	Is the adnominal possessive construction different for alienable and inalienable nouns?
GB065a	OrderPosrPosd	Is the order of possessor noun and possessed noun possessor-possessed?
GB065b	OrderPosdPosr	Is the order of possessor noun and possessed noun possessed-possessor?
GB068	PredAdjLikeV	Do core adjectives (defined semantically as property concepts such as value, shape, age, dimension) act like verbs in predicative position?
GB069	AttrAdjLikeV	Do core adjectives (defined semantically as property concepts; value, shape, age, dimension) used attributively require the same morphological treatment as verbs?
GB070	CoreCaseNoun	Are there morphological cases for non-pronominal core arguments (i.e. S/A/P)?
GB071	CoreCasePRO	Are there morphological cases for pronominal core arguments (i.e. S/A/P)?
GB072	ObliqueCaseNoun	Are there morphological cases for oblique non-pronominal NPs (i.e. not S/A/P)?
GB073	ObliqueCasePRO	Are there morphological cases for independent oblique personal pronominal arguments (i.e. not S/A/P)?
GB074	Prepositions	Are there prepositions?
GB075	Postpositions	Are there postpositions?

Grambank ID	**Abbreviation**	**Feature**
GB079	VPrefixing	Do verbs have prefixes/proclitics, other than those that only mark A, S or P (do include portmanteau: A and S + TAM)?
GB080	VSuffixing	Do verbs have suffixes/enclitics, other than those that only mark A, S or P (do include portmanteau: A and S + TAM)?
GB081	VInfix	Is there productive infixation in verbs?
GB082	PresentBoundV	Is there overt morphological marking of present tense on verbs?
GB083	PastBoundV	Is there overt morphological marking on the verb dedicated to past tense?
GB084	FutureBoundV	Is there overt morphological marking on the verb dedicated to future tense?
GB086	AspectBoundV	Is a morphological distinction between perfective and imperfective aspect available on verbs?
GB089	S-ArgSfxV	Can the S argument be indexed by a suffix/enclitic on the verb in the simple main clause?
GB090	S-ArgPfxV	Can the S argument be indexed by a prefix/proclitic on the verb in the simple main clause?
GB091	A-ArgSfxV	Can the A argument be indexed by a suffix/enclitic on the verb in the simple main clause?
GB092	A-ArgPfxV	Can the A argument be indexed by a prefix/proclitic on the verb in the simple main clause?
GB093	P-ArgSfxV	Can the P argument be indexed by a suffix/enclitic on the verb in the simple main clause?
GB094	P-ArgPfxV	Can the P argument be indexed by a prefix/proclitic on the verb in the simple main clause?
GB095	CaseSplitTAM	Are variations in marking strategies of core participants based on TAM distinctions?
GB096	CaseSplitVerbClass	Are variations in marking strategies of core participants based on verb classes?
GB098	CaseSplitPerson	Are variations in marking strategies of core participants based on person distinctions?
GB099	VSupplPerson	Can verb stems alter according to the person of a core participant?
GB103	BenefApplBoundV	Is there a benefactive applicative marker on the verb (including indexing)?
GB104	AppInstrBoundV	Is there an instrumental applicative marker on the verb (including indexing)?

Grambank ID	Abbreviation	Feature
GB105	CaseRecipientObj	Can the recipient in a ditransitive construction be marked like the monotransitive patient?
GB107	NEGBoundV	Can standard negation be marked by an affix, clitic or modification of the verb?
GB108	DirLocBoundV	Is there directional or locative morphological marking on verbs?
GB109	VSupplNUM	Is there verb suppletion for participant number?
GB110	VSupplTA	Is there verb suppletion for tense or aspect?
GB111	VerbClass	Are there conjugation classes?
GB113	TransitivizingBound	Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?
GB114	ReflBoundV	Is there a phonologically bound reflexive marker on the verb?
GB115	RecipBoundV	Is there a phonologically bound reciprocal marker on the verb?
GB116	VClassifiers	Do verbs classify the shape, size or consistency of absolutive arguments by means of incorporated nouns, verbal affixes or suppletive verb stems?
GB117	CopulaPredNom	Is there a copula for predicate nominals?
GB118	SerialV	Are there serial verb constructions?
GB119	MoodAUX	Can mood be marked by an inflecting word ("auxiliary verb")?
GB120	AspectAUX	Can aspect be marked by an inflecting word ("auxiliary verb")?
GB121	AUXTense	Can tense be marked by an inflecting word ("auxiliary verb")?
GB122	VCompounding	Is verb compounding a regular process?
GB123	LightVerb	Are there verb-adjunct (aka light-verb) constructions?
GB124	NounIncorpIntrans	Is incorporation of nouns into verbs a productive intransitivizing process?
GB126	ExistentialV	Is there an existential verb?
GB127	PostureVerbs	Are different posture verbs used obligatorily depending on an inanimate locatum's shape or position (e.g. 'to lie' vs. 'to stand')?
GB129	FewVerbs	Is there a notably small number, i.e. about 100 or less, of verb roots in the language?
GB130a	OrderSV	Is the order of S and V in intranstive clauses SV?
GB130b	OrderVS	Is the order of S and V in intranstive clauses VS?

Grambank ID	Abbreviation	Feature
GB131	TransVInitOrder	Is a pragmatically unmarked constituent order verb-initial for transitive clauses?
GB132	TransVMedOrder	Is a pragmatically unmarked constituent order verb-medial for transitive clauses?
GB133	TransVFinalOrder	Is a pragmatically unmarked constituent order verb-final for transitive clauses?
GB1134	MainSubSameOrder	Is the order of constituents the same in main and subordinate clauses?
GB135	SUBClauseInOPosition	Do clausal objects usually occur in the same position as nominal objects?
GB136	WordOrderFixed	Is the order of core argument (i.e. S/A/P) constituents fixed?
GB137	NEGFinal	Can standard negation be marked clause-finally?
GB138	NEGInitial	Can standard negation be marked clause-initially?
GB139	NEGProhibitive	Is there a difference between imperative (prohibitive) and declarative negation constructions?
GB140	SameNEGLoc ExistNom	Is verbal predication marked by the same negator as all of the following types of predication: locational, existential and nominal?
GB146	VControl	Is there a morpho-syntactic distinction between predicates expressing controlled versus uncontrolled events or states?
GB147	PassiveBoundV	Is there a morphological passive marked on the lexical verb?
GB148	AntipassiveBoundV	Is there a morphological antipassive marked on the lexical verb?
GB149	InverseBoundV	Is there a morphologically marked inverse on verbs?
GB150	ClauseChain	Is there clause chaining?
GB151	SwitchReference	Is there an overt verb marker dedicated to signalling coreference or noncoreference between the subject of one clause and an argument of an adjacent clause ("switch reference")?
GB152	SimulSeqBound	Is there a morphologically marked distinction between simultaneous and sequential clauses?
GB155	CAUSBound	Are causatives formed by affixes or clitics on verbs?
GB156	CAUSSay	Is there a causative construction involving an element that is unmistakably grammaticalized from a verb for 'to say'?
GB158	RedupV	Are verbs reduplicated?

Grambank ID	Abbreviation	Feature
GB159	RedupNoun	Are nouns reduplicated?
GB160	RedupOther	Are elements apart from verbs or nouns reduplicated?
GB165	TrialBound	Is there productive morphological trial marking on nouns?
GB166	PaucalBound	Is there productive morphological paucal marking on nouns?
GB167	PROLogophore	Is there a logophoric pronoun?
GB170	ADJGender	Can an adnominal property word agree with the noun in gender/noun class?
GB171	DEMGender	Can an adnominal demonstrative agree with the noun in gender/noun class?
GB172	ARTGender	Can an article agree with the noun in gender/noun class?
GB177	AnimacyBoundV	Can the verb carry a marker of animacy of argument, unrelated to any gender/noun class of the argument visible in the NP domain?
GB184	AdjNUM	Can an adnominal property word agree with the noun in number?
GB185	DEMNum	Can an adnominal demonstrative agree with the noun in number?
GB186	ARTNum	Can an article agree with the noun in number?
GB187	DIMBound	Is there any productive diminutive marking on the noun (exclude marking by system of nominal classification only)?
GB188	AUGBound	Is there any productive augmentative marking on the noun (exclude marking by system of nominal classification only)?
GB192	GenderPhono	Is there a gender system where a noun's phonological properties are a factor in class assignment?
GB193a	OrderANMN	Is the order of the adnominal property and the noun ANM-N?
GB193b	OrderNANM	Is the order of the adnominal property and the noun N-ANM?
GB196	PRO2PMascFem	Is there a male/female distinction in 2nd person independent pronouns?
GB197	PRO1PMascFem	Is there a male/female distinction in 1st person independent pronouns?
GB198	NUMGender	Can an adnominal numeral agree with the noun in gender/noun class?
GB203a	OrderQuantUQN	Is the order of the adnominal collective universal quantifier ('all') and the noun UQ-N?
GB203b	OrderQuantNUQ	Is the order of the adnominal collective universal quantifier ('all') and the noun N-UQ?

Grambank ID	Abbreviation	Feature
GB204	QUANTUniversal	Do collective ('all') and distributive ('every') universal quantifiers differ in their forms or their syntactic positions?
GB250	PredPOSSHabeo	Can predicative possession be expressed with a transitive 'habeo' verb?
GB252	PredPOSSLoc	Can predicative possession be expressed with an S-like possessum and a locative-coded possessor?
GB253	PredPOSSDat	Can predicative possession be expressed with an S-like possessum and a dative-coded possessor?
GB254	PredPOSSAdnom	Can predicative possession be expressed with an S-like possessum and a possessor that is coded like an adnominal possessor?
GB256	PredPOSS Comitative	Can predicative possession be expressed with an S-like possessor and a possessum that is coded like a comitative argument?
GB257	QIntonation	Can polar interrogation be marked by intonation only?
GB260	QWordOrder	Can polar interrogation be indicated by a special word order?
GB262	QPartInitial	Is there a clause-initial polar interrogative particle?
GB263	QPartFinal	Is there a clause-final polar interrogative particle?
GB264	QPartMedial	Is there a polar interrogative particle that most commonly occurs neither clause-initially nor clause-finally?
GB265	COMPARExceed	Is there a comparative construction that includes a form that elsewhere means 'surpass, exceed'?
GB266	COMPARLoc	Is there a comparative construction that employs a marker of the standard which elsewhere has a locational meaning?
GB270	COMPARConjoin	Can comparatives be expressed using two conjoined clauses?
GB273	COMPAR OtherMarker	Is there a comparative construction with a standard marker that elsewhere has neither a locational meaning nor a 'surpass/exceed' meaning?
GB275	COMPDegreeBound	Is there a bound comparative degree marker on the property word in a comparative construction?
GB276	COMPDegreeFree	Is there a non-bound comparative degree marker modifying the property word in a comparative construction?
GB285	QPartVMorph	Can polar interrogation be marked by a question particle and verbal morphology?

Grambank ID	Abbreviation	Feature
GB286	QVMorph	Can polar interrogation be indicated by overt verbal morphology only?
GB291	QTone	Can polar interrogation be marked by tone?
GB296	Ideophones	Is there a phonologically or morphosyntactically definable class of ideophones that includes ideophones depicting imagery beyond sound?
GB297	QVNotV	Can polar interrogation be indicated by a V-not-V construction?
GB298	NEGAux	Can standard negation be marked by an inflecting word ("auxiliary verb")?
GB299	NEGPart	Can standard negation be marked by a non-inflecting word ("auxiliary particle")?
GB300	VSupplGive	Does the verb for 'give' have suppletive verb forms?
GB301	Inclusory	Is there an inclusory construction?
GB302	PassiveFree	Is there a phonologically free passive marker ("particle" or "auxiliary")?
GB303	AntipassiveFree	Is there a phonologically free antipassive marker ("particle" or "auxiliary")?
GB304	PassiveA-ArgOvert	Can the agent be expressed overtly in a passive clause?
GB305	PRORefl	Is there a phonologically independent reflexive pronoun?
GB306	PROReciproc	Is there a phonologically independent non-bipartite reciprocal pronoun?
GB309	MultiplePastFuture	Are there multiple past or multiple future tenses, distinguishing distance from Time of Reference?
GB312	MoodBoundV	Is there overt morphological marking on the verb dedicated to mood?
GB313	PROPoss	Are there special adnominal possessive pronouns that are not formed by an otherwise regular process?
GB314	AUGgender	Can augmentative meaning be expressed productively by a shift of gender/noun class?
GB315	DIMGender	Can diminutive meaning be expressed productively by a shift of gender/noun class?
GB316	SingularFree	Is singular number regularly marked in the noun phrase by a dedicated phonologically free element?
GB317	DualFree	Is dual number regularly marked in the noun phrase by a dedicated phonologically free element?

Grambank ID	Abbreviation	Feature
GB318	PluralFree	Is plural number regularly marked in the noun phrase by a dedicated phonologically free element?
GB319	TrialFree	Is trial number regularly marked in the noun phrase by a dedicated phonologically free element?
GB320	PaucalFree	Is paucal number regularly marked in the noun phrase by a dedicated phonologically free element?
GB321	GenderUnpredict	Is there a large class of nouns whose gender/noun class is not phonologically or semantically predictable?
GB322	EvidSense	Is there grammatical marking of direct evidence (perceived with the senses)?
GB323	EvidIndirect	Is there grammatical marking of indirect evidence (hearsay, inference, etc.)?
GB324	QV	Is there an interrogative verb for content interrogatives (who?, what?, etc.)?
GB325	QCountMass	Is there a count/mass distinction in interrogative quantifiers?
GB326	QInSitu	Do (nominal) content interrogatives normally or frequently occur in situ?
GB327	RELPost	Can the relative clause follow the noun?
GB328	RELPre	Can the relative clause precede the noun?
GB329	RELInternalHead	Are there internally-headed relative clauses?
GB330	RELCorr	Are there correlative relative clauses?
GB331	RELAdjoined	Are there non-adjacent relative clauses?
GB333	NUMDecimal	Is there a decimal numeral system?
GB334	NUMQuinary	Is there synchronic evidence for any element of a quinary numeral system?
GB335	NUMVigesimal	Is there synchronic evidence for any element of a vigesimal numeral system?
GB336	NUMBodyTally	Is there a body-part tallying system?
GB400	PersonNeutralized	Are all person categories neutralized in some voice, tense, aspect, mood and/or negation?
GB401	PatientLabile	Is there a class of patient-labile verbs?
GB402	VSupplSee	Does the verb for 'see' have suppletive verb forms?
GB403	VSupplCome	Does the verb for 'come' have suppletive verb forms?
GB408	Accusative	Is there any accusative alignment of flagging?

Grambank ID	Abbreviation	Feature
GB409	Ergative	Is there any ergative alignment of flagging?
GB410	NeutralAlign	Is there any neutral alignment of flagging?
GB415	PRO2PPoliteness	Is there a politeness distinction in 2nd person forms?
GB421	ComplThink KnowPre	Is there a preposed complementizer in complements of verbs of thinking and/or knowing?
GB422	ComplThink KnowPost	Is there a postposed complementizer in complements of verbs of thinking and/or knowing?
GB430	POSSPfxPosr	Can adnominal possession be marked by a prefix on the possessor?
GB431	POSSPfxPosd	Can adnominal possession be marked by a prefix on the possessed noun?
GB432	POSSSfxPosr	Can adnominal possession be marked by a suffix on the possessor?
GB433	POSSSfxPosd	Can adnominal possession be marked by a suffix on the possessed noun?
GB519	MoodAuxPart	Can mood be marked by a non-inflecting word ("auxiliary particle")?
GB520	AspectAuxPart	Can aspect be marked by a non-inflecting word ("auxiliary particle")?
GB521	TenseAuxPart	Can tense be marked by a non-inflecting word ("auxiliary particle")?
GB522	PRODrop	Can the S or A argument be omitted from a pragmatically unmarked clause when the referent is inferrable from context ("pro-drop" or "null anaphora")?

C Table of Political complexity scores per society

The political complexity scores are based on Sheehan et al. (2018), the Ethnographic Atlas (Kirby et al. 2018),Bonnemaison (1972) and Bonnemaison (1996). The scores from Sheehan et al. (2018) and the Ethnographic Atlas are displayed along with the score used in chapter **??**. I re-evaluated the scores and inspected the references again to come to an independent decision, this is why the scores sometimes differ.

In the study by Sheehan et al. (2018) they make a distinction which between 0 = local communities are associations of households (or other sub-local groups, such as village wards) with no overarching system of authority and 1 = autonomous local communities which each had a system of authority, e.g. a village council (Sheehan personal correspondence). These two levels are merged in my coding and the coding from the Ethnographic Atlas. *NA* stands for missing data.

Table 10: Table of Political Complexity score per society

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
East Ambae	east2443	1	*NA*	1	Bonnemaison, J. (1972). Système de grades et différences régionales en Aoba (Nouvelles Hébrides). Cahiers ORSTOM. Série Sciences Humaines, 9(1), 87-108.
West Ambae	west2513	1	*NA*	1	Bonnemaison, J. (1972). Système de grades et différences régionales en Aoba (Nouvelles Hébrides). Cahiers ORSTOM. Série Sciences Humaines, 9(1), 87-108.
Southeast Ambrym	sout2859	1	1	*NA*	Tonkinson R (1981) Church and Kastom in Southeast Ambrym. Vanuatu: Politics, Economics and Ritual in Island Melanesia, ed Allen M (Academic Press, Sydney, Australia), pp 237-267.
Aneityum	anei1239	2	2	*NA*	Humphreys CB (1926) Southern New Hebrides: An Ethnological Record (Cambridge Univ Press, Cambridge, UK); Spriggs M (1982) Taro Cropping Systems in the Southeast Asian-Pacific Region: Archaeological Evidence. Archaeol Ocean 17(1):7-15; Spriggs M (1986) Landscape, Land Use, and Political Transformation in Southern Melanesia. Island Societies: Archaeological Approaches to Evolution and Transformation, ed Kirch PV (Cambridge Univ Press, New York, NY), pp 6-19.
Anuta	anut1237	1	1	*NA*	Feinberg R (1988) Socio-Spatial Symbolism and the Logic of Rank on Two Polynesian Outliers. Ethnology 27(3):291-310; Feinberg R (1991) Anuta. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 13-16.; Kirch PV (2002) Te Kai Paka-Anuta: Food in a Polynesian Outlier Society. Le Journal de la Société des Océanistes 114-115:71-89.
Aore	aore1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Rennell-Bellona/Mu-Ngava-Mu-Hgiki	renn1242	2	2	2	Birket-Smith K (1969) An Ethnological Sketch of Rennell Island, a Polynesian Outlier in Melanesia (2nd Ed) (Bianco Lunos Bogtrykkeri, Copenhagen, Denmark); Monberg T (1991) Bellona Island Beliefs and Rituals (Univ Hawaii Press, Honolulu, HI).
Chuukese	chuu1238	2	2	1	Goodenough WH (1991) Truk. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 351-354; Goodenough WH (2002) Under Heaven's Brow: Pre-Christian Religious Tradition in Chuuk (American Philosopical Society, Philadelphia, PA); (1960) Taro cultivation in Truk. Taro Cultivation Practices and Beliefs: Part II. The Eastern Carolines and the Marshall Islands, ed Young JE (Office of the Staff Anthropologist, Guam, GU), pp 70-98.
East Futuna	east2447	2	2	2	Kirch PV (1994) The Wet and the Dry: Irrigation and Agricultural Intensification in Polynesia (Univ Chicago Press, Chicago, IL);
North Efate	nort2836	2	2	*NA*	Facey EE (1981) Hereditary chiefship in Nguna. Vanuatu: Politics, Economics and Ritual in Island Melanesia, ed Allen M (Academic Press, Sydney, Australia), pp 295-314. Facey EE (1991) Nguna. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 242-244.
Sie	siee1239	2	2	*NA*	Humphreys CB (1926) Southern New Hebrides: An Ethnological Record (Cambridge Univ Press, Cambridge, UK). Spriggs M, Wickler S (1989) Archaeological Research on Erromango: Recent Data on Southern Melanesian Prehistory. Bulletin of the Indo-Pacific Prehistory Association 9:68-91.
Futuna-Aniwa	futu1245	2	2	*NA*	Capell A (1958) Culture and Language of Futuna and Aniwa, New Hebrides (Univ Sydney, Sydney, Australia).

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Chamorro	cham1312	1	1	2	Cordy R (1983) Social stratification in the Mariana Islands. Oceania 53(3):272-276; Thompson L (1971) The Native Culture of the Marianas Islands (Bernice P Bishop Museum Bulletin, Honolulu, HI) (Originally published 1945).
Hawaiian	hawa1245	4	4	3	Kirch PV (1994) The Wet and the Dry: Irrigation and Agricultural Intensification in Polynesia (Univ Chicago Press, Chicago, IL). Kirch PV (2010) How Chiefs Became Kings: Divine Kingship and the Rise of Archaic States in Ancient Hawai'i (Univ California Press, Oakland, CA).
Hiw	hiww1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Fijian	fiji1243	3	3	*NA*	Kuhlken R (2002) Intensive Agricultural Landscapes of Oceania. Journal of Cultural Geography 19(2):161-195. 80) Scarr D (1984) Fiji: A Short History (George Allen and Unwin, Sydney, Australia). 81) Walter MAHB (1978) An examination of hierarchical notions in Fijian society: A test case for the applicability of the term 'chief'. Oceania 49(1):1-19.
Ajië	ajie1238	1	*NA*	2	Winslow, Don (1991) Ajie. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 9.
Xârâcùù	xara1244	3	3	*NA*	Young MW (1991) Goodenough Island. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 85-88.
Kapingamarangi	kapi1249	1	1	2	Buck PH (1950) Material Culture of Kapingamarangi (Bernice P. Bishop Museum, Honolulu, HI). Emory KP (1965) Kapingamarangi: Social and Religious Life of a Polynesian Atoll (The Museum, Honolulu, HI).

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Kosraean	kosr1238	3	3	3	Athens JS (2007) Prehistoric Population Growth on Kosrae, Eastern Caroline Islands. The Growth and Collapse of Pacific Island Societies, eds Kirch PV, Rallu J (Univ Hawaii Press, Honolulu, HI), pp 257-277. Graves MW (1986) Late Prehistoric Complexity on Lelū: Alternatives to Cordy's Model. J Polyn Soc 95(4), 479-489. Peoples JG (1991) Kosrae. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 128-131.
Lauan	laua1243	3	0	3	Hocart, A. M. 1929. Lau Islands, Fiji. (Bull. Bishop Mus., 62.) 1-240pp. Quain, Buell H. 1948. Fijian village. Chicago: University of Chicago Press. Thompson, L. 1940. Southern Lau, Fiji. (Bull. Bishop Mus., 162.) Thompson, Laura. 1940. Fijian frontier. (Studies of the Pacific.) San Francisco: Institute of Pacific Relations.
Dehu	dehu1237	2	*NA*	2	Hadfield, E. 1920. Among the Natives of the Loyalty Group., Ray, S. 1917. The People and Language of Lifu, Loyalty Islands. Journ. Roy. Anth. Inst. 47. 239-322.
Äiwoo	ayiw1239	1	0	*NA*	Davenport WH (1969) Social organization notes on the Northern Santa Cruz Islands: the Main Reef Islands. Baessler-Archiv, Neue Folge 17(1):151-243.
Luangiua	onto1237	1	1	1	Sahlins MD (1958) Social Stratification in Polynesia (Univ Washington Press, Seattle, WA). Bayliss-Smith T (1974) Constraints on population growth: The case of the Polynesian Outlier atolls in the precontact period. Hum Ecol 2(4):259-295. Donner WW (1991) Ontong Java. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 253-255.
Baetora	baet1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
Central Maewo	cent2058	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Sunwadia	mari1426	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Aulua	aulu1238	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Avok	avok1244	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Axamb	axam1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Big Nambas	bign1238	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Burmbar	burm1263	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Bwenelang	bwen1239	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Dixon Reef	dixo1238	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Avava	katb1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Ninde	labo1244	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Larevat	lare1249	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Letemboi-Repanbitip	lete1241	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Neverver	ling1265	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Naman	litz1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Tirax	maee1241	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Na'ahai	malf1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Malua Bay	malu1245	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Maragus	mara1399	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Maskelynes	mask1242	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Mpotovoro	mpot1241	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Nese	nese1235	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Nisvai	nisv1234	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Nitita	niti1249	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Port Sandwich	port1285	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Rerep	rere1240	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Unua	unua1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Uripiv-Wala-Rano-Atchin	urip1239	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Neve'ei	vinm1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Vivti	vivt1234	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Nahavaq	sout2857	1	*NA*	1	Deacon, A. B. 1934. Malekula.
Mangareva	mang1401	2	2	3	Buck PH (1971) Ethnology of Mangareva (Bernice P. Bishop Museum, Honolulu, HI) (Originally published 1938). Conte E, Kirch PV (2004) Archaeological Investigations in the Mangareva Islands (Gambier Archipelago), French Polynesia (Univ California, Berkeley, CA). Green RC and Weisler ML (2000) Mangarevan Archaeology: Interpretations using new data and 40 year old excavations to establish a sequence from 1200 to 1900 AD (Univ Otago, Dunedin, New Zealand).
Māngarongaro	penr1237	2	2	2	Buck PH (1932) Ethnology of Tongareva (Bernice P. Bishop Museum, Honolulu, HI). Roscoe PB (1991) Tongareva. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 339-342.
Mota	mota1237	1	*NA*	1	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Mwotlap	motl1237	1	0	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Rimatara-Rurutu-Tupua'i-Ra'ivavae	aust1304	2	2	*NA*	Aitken RT (1971) Ethnology of Tubuai (Bernice P Bishop Museum, Honolulu, HI) (Originally published 1930). Bollt R (2008) Excavations in Peva Valley, Rurutu, Austral Islands (East Polynesia). Asian Perspect 47(1):158-187. Edwards E (2003) Archaeological Survey of Ra'ivavae (Bearsville Press, Los Osos, CA).

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
Māori	maor1246	2	2	2	Sahlins MD (1958) Social Stratification in Polynesia (Univ Washington Press, Seattle, WA). Buck PH (1952) The Coming of the Maori (Whitcombe and Tombs: Wellington, New Zealand). Kirch PV (1984) The Evolution of the Polynesian Chiefdoms (Cambridge Univ Press, Cambridge, UK). Van Meijl T (1995) Maori Socio-Political Organization in Pre- and Proto-History: On the Evolution of Post-Colonial Constructs. Oceania 65(4):304-322.
Tuamotuan	tuam1242	2	2	2	Emory KP (1975) Material Culture of the Tuamotu Archipelago (Bernice P Bishop Museum, Honolulu, HI).
Nengone	neng1238	3	3	*NA*	Dubois M (1984) Gens de Maré (Éditions Anthropos, Paris, France). Guiart J (1952) L'Organisation Sociale et Politique Traditionelle à Maré. (Institut Francçais d'Océanie, Nouméa, New Caledonia) .
Pohnpeian	pohn1238	3	2	3	Hanlon D (1988) Upon a Stone Altar: A History of the Island of Pohnpei to 1890 (Univ Hawaii Press, Honolulu HI). Haun AD (1984) Prehistoric Subsistence, Population, and Sociopolitical Evolution on Ponape, Micronesia. PhD thesis (Univ Oregon, Eugene, OR). Raynor WC, Fownes JH (1991) Indigenous agroforestry of Pohnpei. Agroforestry Systems 16:139-157. Riesenberg S (1968) The Native Polity of Ponape (Smithsonian Institution Press, Washington, DC). ; Hanlon, D. L. (2019). Upon a stone altar: A history of the island of Pohnpei to 1890. University of Hawaii Press.
Niuean	niue1239	2	2	2	Loeb EM (1978) History and Traditions of Niue (Bernice P Bishop Museum, Honolulu, HI). Smith SP (1983) Niue: The Island and Its People (The Polynesian Society, Suva, Fiji) (Originally published 1902-1903). Walter R, Anderson A (1995) Archaeology of Niue island: Initial Results. J Polyn Soc 104(4):471-481.

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
North Marquesan	nort2845	1	1	2	Sahlins MD (1958) Social Stratification in Polynesia (Univ Washington Press, Seattle, WA).
Nukuoro	nuku1260	1	1	*NA*	Carroll V (1966) Nukuoro Kinship. PhD thesis (Univ Chicago, Chicago, IL). Carroll V (1975) Demographic concepts and techniques for the study of small populations. Pacific Atoll Populations, ed Carrol V (Univ Hawaii Press, Honolulu, HI), pp 344-416. Eilers A (1934) Islands around Ponape : Kapingamarangi, Nukuoro, Ngatik, Mokil, Pingelap (Friederichsen, De Gruyter and Co, Hamburg, Germany).
Paama	paam1238	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Palauan	pala1344	2	1	3	Force RW (1960) Leadership and Cultural Change in Palau (Chicago tural History Museum, Chicago, IL).
Sa	saaa1241	1	*NA*	1	Lane, R. B. 1956. The Heathen Communities of Southeast Pentecost. Journal de la Soci,te des Oceanistes 12. 139-180., Lane, R. B. 1965. The Melanesians of South Pentecost. In P. Lawrence and M. G. Meggitt (eds.), Gods, Ghosts and Men in Melanesia, 250-279. Lane, R. B., and B. S. Lane. 1957. Unpublished field notes.
Pukapuka	puka1242	2	*NA*	2	Beaglehole, E., and P. Beaglehole. 1938. Ethnology of Pukapuka. Bull. Bishop Mus. 110. 1-419. Macgregor, G. 1935. Notes on the Ethnology of Pukapuka. Bishop Mus. Occas. Pap. 11. vi, 1-52.
Rakahanga-Manihiki	raka1237	2	*NA*	2	Buck, P. H. 1932. Ethnology of Manihiki and Rakahanga. Bull. Bishop Mus. 99. 1-238.

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
Marshallese	mars1254	3	3	*NA*	Carucci LM (1991) Marshall Islands. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 191-194. Erdland A (1961) The Marshall Islanders: Life and Customs, Thought and Religion of a South Seas People (R. Neuse, Trans) (Human Relations Area Files, New Haven, CT) (Originally published 1914). Williamson I, Sabath MD (1982) Island Population, Land Area, and Climate: a Case Study of the Marshall Islands. Hum Ecol 10(1):71-84.
Rapanui	rapa1244	2	2	2	Sahlins MD (1958) Social Stratification in Polynesia (Univ Washington Press, Seattle, WA). Kirch PV (1984) The Evolution of the Polynesian Chiefdoms (Cambridge Univ Press, Cambridge, UK). 209) Métraux A (1971) Ethnology of Easter Island (Bernice P Bishop Museum, Honolulu, HI).
Tahitian	tahi1242	3	4	3	Oliver DL (1974) Ancient Tahitian Society (Volume 2: Social Relations) (Univ Hawaii Press: Honolulu, HI). Pages: 970-973
Māori o te Pae Tonga	raro1241	2	2	2	Bellwood PS (1971) Varieties of Ecological Adaptation in the Southern Cook Islands. Archaeol Ocean 6(2):145-169. Buck PH (1934) Mangaian Society (Bernice P. Bishop Museum, Honolulu, HI). 240) Crocombe RG (1967) Ascendancy to dependency: the politics of Atiu. J Pac Hist 2(1):97-111. Gilson R, Crocombe R (1980) The Cook Islands 1820-1950 (Victoria Univ Press, Wellington, New Zealand). 242) Walter R (1996) Settlement pattern archaeology in the Southern Cook Islands: a review. J Polyn Soc 105(1):63-99.

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
Rotuman	rotu1241	3	3	2	Gardiner JS (1898) The natives of Rotuma. The Journal of the Anthropological Institute of Great Britain and Ireland 27:396-435. Howard A (1963) Conservatism and non-traditional leadership in Rotuma. J Polyn Soc 72(2):65-77. Howard A (1991) Rotuma. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 280-283.
Saipan Carolinian	caro1242	1	*NA*	1	Joseph, A., and V. F. Murray. 1951. Chamorros and Carolinians of Saipan: personality tests with an analysis of the Bender Gestalt test by Lauretta Bender. Cambridge: Harvard University Press. Spehr, A. 1954. Saipan. Fieldiana: Anth. 41. 1-383.
Samoan	samo1305	3	3	*NA*	Sahlins MD (1958) Social Stratification in Polynesia (Univ Washington Press, Seattle, WA). Buck PH (1930) Samoan Material Culture (Bernice P. Bishop Museum, Honolulu, HI). Keesing FM (1934) Modern Samoa: Its Government and Changing Life (Allen and Unwin Ltd, London, UK). 226) Watters RF (1958) Cultivation in Old Samoa. Economic Geography 43(4):338-351.
Farafi	butm1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Polonombauk	polo1242	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Akei	akei1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Amblong	ambl1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Kiai	fort1240	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Nethalp	lore1244	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Merei	mere1242	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Morouas	moro1286	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Nokuku	noku1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Piamatsina	piam1242	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Mores	rori1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Wanohe	saka1289	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Ngen	shar1244	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Tambotalo	tamb1253	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Tasmate	tasm1246	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Tiale	tial1239	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Tolomako	tolo1255	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Valpei	valp1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Ale	wail1242	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	**Glottocode**	**Political Complexity (EA033)**			**References**
		This thesis	**Sheehan et al 2008**	**Ethnographic Atlas (D-PLACE)**	
Kula (Vanuatu)	wusi1237	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Movono	tang1347	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Southwest Tanna	sout2869	1	*NA*	2	Lindström, Lamont (1991) Ajie. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 314.
Lo-Toga	loto1240	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Tikopia	tiko1237	2	2	2	Kirch PV (1994) The Wet and the Dry: Irrigation and Agricultural Intensification in Polynesia (Univ Chicago Press, Chicago, IL). Sahlins MD (1958) Social Stratification in Polynesia (Univ Washington Press, Seattle, WA). Firth R (1939) Primitive Polynesian Economy (George Routledge and Sons, London, UK). Firth R (1959) Social Change in Tikopia: Re-Study of a Polynesian Community after a Generation (Allen and Unwin, London, UK). Firth R (1991) Tikopia. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 324-327.
Tokelau	toke1240	2	2	2	Hooper A, Huntsman J (1973) A demographic history of the Tokelau Islands. J Polyn Soc 84(4):366-411. MacGregor G (1937) Ethnology of Tokelau Islands (Bernice P Bishop Museum, Honolulu, HI).
Tonga (Tonga Islands)	tong1325	3	3	3	Kirch PV (1984) The Evolution of the Polynesian Chiefdoms (Cambridge Univ Press, Cambridge, UK). Cummins HG (1977) Tongan Society at the Time of European Contact. Friendly Islands: A History of Tonga, ed Rutherford N (John Sands Ltd, Melbourne, Australia), pp 63-89. Ferdon EN (1987) Early Tonga (Univ Arizona Press, Tucson, AZ).

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
Gilbertese	gilb1244	2	2	*NA*	Lambert B (1966) The Economic Activities of a Gilbertese Chief. Political Anthropology, ed Schwartz MJ, Turner VW, Tuden A (Transaction Publishers, New Brunswick, NJ), pp 155-172. Lambert B (1975) Makin and the Outside World. Pacific Atoll Populations, ed Carroll V (Univ Hawaii Press, Honolulu, HI), pp 212-285. Lambert B (1991) Kiribati. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 120-124. Macdonald B (1982) Cinderellas of the Empire: Towards a History of Kiribati and Tuvalu (ANU Press, Canberra, Australia).
Tuvalu	tuva1244	2	2	2	Macdonald B (1982) Cinderellas of the Empire: Towards a History of Kiribati and Tuvalu (ANU Press, Canberra, Australia). Goldsmith M (1991) Tuvalu. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 354-357.
Ulithian	ulit1238	2	*NA*	2	Lessa, W. A. 1950. The Ethnography of Ulithi Atoll. Unpublished Manuscript Ulithi (Micronesia). Lessa, William Armand. 1966. Ulithi: A Micronesian design for living. New York: Holt, Rinehart and Winston.
Lehali	leha1243	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Lehalurup	leha1244	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
East Uvean	wall1257	2	2	3	Burrows EG (1971) Ethnology of Uvea (Wallis Island) (The Museum, Honolulu, HI) (Originally published 1937) . Pollock NJ (1995) The Power of Kava in Futuna and 'Uvea/Wallis. Canberra Anthropology 18(1-2):136-165.
Lemerig	leme1238	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.

Language name	Glottocode	Political Complexity (EA033)			References
		This thesis	Sheehan et al 2008	Ethnographic Atlas (D-PLACE)	
Vera'a	vera1241	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Vurës	vure1239	1	*NA*	*NA*	Bonnemaison, J (1996) The Art of Power. In Bonnemaison (eds) Arts of Vanuatu. University of Hawaii Press.
Woleaian	wole1240	2	2	*NA*	Alkire WH (1991) Woleai. Encyclopaedia of World Cultures (Vol II: Oceania) (G.K. Hall and Co, New York, NY), pp 383-384. Burrows EG, Spiro ME (1953) An Atoll Culture: Ethnography of Ifaluk in the Central Carolines (Human Relations Area Files, New Haven, CT).
Yapese	yape1248	3	*NA*	2	Hunt, E. E., Jr., D. M. Schneider, N. R. Kidder, and W. D. Stevens. 1949. The Micronesians of Yap and Their Depopulation. Muller, W. 1917. Yap. (Ergebnisse der Südsee-Expedition 1908-1910, 2, B, iii.) In G. Thilenius (ed.) 1-380pp. Murdock, G. P., C. S. Ford, and J. W. M. Whiting. 1944. West Caroline Islands. 1-222pp. Salesius. 1906. Die Karolineninsel Jap. Schneider, David M. 1953. Yap Kinship Terminology and Kin Groups. American Anthropologist 55. 215-236 . Yapese Schneider. 1957. Political Organization, Supernatural Sanctions and the Punishment for Incest on Yap. American Anthropologist 59. 791-800. Schneider, D. M. 1962. Double Descent on Yap. Journal of the Polynesian Society 71. 1-24. Tetens, A. 1958. Among the Savages of the South Seas. (Trans. F. M. Spoehr). Tetens, A., and J. Kubary. 1873. Die Carolineninsel Yap. Journal des Museum Godeffroy 1. 84-120.

D Table of settlement date per island group based on archaeology

Table 11: Table of settlement date per island group based on archaeology

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Mariana Islands	Guam	1	3500, 2950	3500	Carson (2014) and Athens et al (2004) as cited in Rieth & Cochrane (2018)	
Mariana Islands	Saipan	1	3500, 2950	3500	Carson (2014) and Athens et al (2004) as cited in Rieth & Cochrane (2018)	
Vanuatu	Tangoa	2	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Reef Islands and Santa Cruz Islands	Reef Islands	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	
Reef Islands and Santa Cruz Islands	Lomlom	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	
Reef Islands and Santa Cruz Islands	Nendö	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Reef Islands and Santa Cruz Islands	Te Anu	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	
Reef Islands and Santa Cruz Islands	Utupua	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	
Reef Islands and Santa Cruz Islands	Vanikoro	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Reef Islands and Santa Cruz Islands	Tegua	2	3185, 2639	3185	Green (1991), Green et al (2008) and Sheppard et al (2015) as cited in Rieth & Cochrane (2018)	
Vanuatu	Ambae	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Ambrym	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Aneityum	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Vanuatu	Aore	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Araki	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Efate	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Epi	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Erromango	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Vanuatu	Gaua	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Hiu	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Loh-Toga	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Maewo	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Mafea	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Vanuatu	Malakula	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Merelava group	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Mota	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Mota Lava	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Paama	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Vanuatu	Pentecost	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Santo	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Tamambo	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Tanna	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Tongoa	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Vanuatu	Tutuba	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Ureparapara	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Vanua Lava	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Vao	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Etarik	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Vanuatu	Lopevi	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Vanuatu	Mataso	3	3200, 3000	3200	Bedford et al (2006) as cited in Rieth & Cochrane (2018)	
Fiji	Vanua Levu	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Fiji	Yasawa (greater)	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	
Fiji	Taveuni	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Fiji	Lau	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	
Fiji	Viti Levu	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Fiji	Lomaviti	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	
Fiji	Leveuka (Lomaviti)	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Fiji	Vatulele	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	
Fiji	Kadavu	4	3130, 2870	3130	Hope et al (2009), Denham et al (2012) and Nunn and Petchey (2013) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Palau	Palau	4	3100, 2900	3100	Athens and Ward (2001), Clark et al (2006) and Liston (2005, 2013) as cited in Rieth & Cochrane (2018)	
New Caledonia	Belep	4	3050, 3000	3050	Sand (2001) as cited in Rieth & Cochrane (2018)	
New Caledonia	Kanaky (New Caledonia main island)	4	3050, 3000	3050	Sand (2001) as cited in Rieth & Cochrane (2018)	
New Caledonia	Lifou	4	3050, 3000	3050	Sand (2001) as cited in Rieth & Cochrane (2018)	
New Caledonia	Nengone	4	3050, 3000	3050	Sand (2001) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
New Caledonia	Ouvea	4	3050, 3000	3050	Sand (2001) as cited in Rieth & Cochrane (2018)	
Ouvea	Ouvea (Iaai)	4	3050, 3000	3050	Sand (2001) as cited in Rieth & Cochrane (2018)	
Tikopia	Tikopia	4	3000	3000	Carson (2012) 0	
Tonga	Tonga	5	2846, 2750	2846	Burley and Connaughton (2007) and Burley et al (1999, 2001, 2012) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Samoa	Samoa	5	2800, 2400	2800	Cochrane et al (2013), Kirch and Hunt (1993) and Petchey (2001) as cited in Rieth & Cochrane (2018)	
Nguluwan	Nguluwan	6	2400, 2100	2400	Napolitano et al. (2019) 0	Yes
Yap	Yap	6	2400, 2100	2400	Napolitano et al. (2019) 0	
Chuuk	Chuuk	7	2300, 1750	2300	Shutler (1984) Rieth & Cochrane (2018)	
Chuuk	Mortlock	7	2300, 1750	2300	Shutler (1984) as cited in Rieth & Cochrane (2018)	Yes

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Chuuk	Namonuito	7	2300, 1750	2300	Shutler (1984) as cited in Rieth & Cochrane (2018)	Yes
Chuuk	Paafang	7	2300, 1750	2300	Shutler (1984) as cited in Rieth & Cochrane (2018)	Yes
Chuuk	Pollap	7	2300, 1750	2300	Shutler (1984) as cited in Rieth & Cochrane (2018)	Yes
Chuuk	Puluwat	7	2300, 1750	2300	Shutler (1984) as cited in Rieth & Cochrane (2018)	Yes
Chuuk	Oroluk	7	2300, 1750	2300	Shutler (1984) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Futuna	East Futuna	7	2300, 2200	2300	Kirch (1981) as cited in Rieth & Cochrane (2018)	
Futuna	Uvea (Wallis)	7	2300, 2200	2300	Kirch (1981) as cited in Rieth & Cochrane (2018)	Yes
Niuafoou	Niuafoou	7	2300, 2200	2300	Kirch (1981) as cited in Rieth & Cochrane (2018)	Yes
Niuatoputapu	Niuatoputapu	7	2300, 2200	2300	Kirch (1981) as cited in Rieth & Cochrane (2018)	Yes
Pohnpei	Mwoakilloa	8	1700	1700	Poteate et al (2016) as cited in Levin et al. (2019)	Yes

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Tungaru	Tungaru	8	2150, 1750	2150	DiPazza (1999) as cited in Rieth & Cochrane (2018)	
Tungaru	Nauru	8	2150, 1750	2150	DiPazza (1999) as cited in Rieth & Cochrane (2018)	Yes
Tungaru	Banaba	8	2150, 1750	2150	DiPazza (1999) as cited in Rieth & Cochrane (2018)	Yes
Tuvalu	Niu	8	2150, 1750	2150	DiPazza (1999) as cited in Rieth & Cochrane (2018)	Yes
Kosrae	Kosrae	8	2100, 1750	2100	Athens (1995) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Marshall Islands	Ratak	8	2000, 1600	2000	Beardsley (1994), Riley (1987), weisler (1999, 2001) and Weisler et al (2012) as cited in Rieth & Cochrane (2018)	
Marshall Islands	Ralik	8	2000, 1600	2000	Beardsley (1994), Riley (1987), weisler (1999, 2001) and Weisler et al (2012) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Marshall Islands	Anewetak	8	2000, 1600	2000	Beardsley (1994), Riley (1987), weisler (1999, 2001) and Weisler et al (2012) as cited in Rieth & Cochrane (2018)	Yes
Niue	Niue	8	2000, 1600	2000	Walter and Anderson (2002) as cited in Rieth & Cochrane (2018)	
Ulithi	Ulithi	8	2000, 1700	2000	Intoh and Leach (1985) and Takayama (1982) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Pohnpei	Pohnpei	8	1900, 1700	1900	Athens (1990) and Galipaud (2000) as cited in Rieth & Cochrane (2018)	
Pohnpei	Ngatik	8	1900, 1700	1900	Athens (1990) and Galipaud (2000) as cited in Rieth & Cochrane (2018)	Yes
Ulithi	Sorol	8	2000, 1700	2000	Intoh and Leach (1985) and Takayama (1982) as cited in 0	Yes
Pohnpei	Satawal	8	1900, 1700	1900	Napolitano et al. (2019) 0	Yes
Pohnpei	Wolei	8	1900, 1700	1900	Napolitano et al. (2019) 0	Yes

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Pohnpei	Satawal (Woleai speaking)	8	1900, 1700	1900	Napolitano et al. (2019) 0	Yes
Pohnpei	Pingelap	8	1700, 1550	1700	Levin et al. (2019) 0	Yes
Rotuma	Rotuma	9	1300, 1050	1300	Ladefoged et al (1998) as cited in Rieth & Cochrane (2018)	
Tuamotu	Hereheretue	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuamotu	Morane	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Tuamotu	Nukutaveke	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuamotu	Tuamotu	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuamotu	Puka-puka	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Tuamotu	Tatakoto	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuamotu	Tureia	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuamotu	Marutea	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Tuamotu	Napuka	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuamotu	Reao	9	1100, 770	1100	Chazine (1985) and Hatanaka et al (1978) as cited in Rieth & Cochrane (2018)	
Tuvalu	Tuvalu	9	1070, 770	1070	Dickinson et al (1990) as cited in Rieth & Cochrane (2018)	
Emae	Emae	9	1000	1000	Kirch (2012) and Carson (2012) 0	Yes
Rennell-Bellona	Rennelle/Mu Ngava	9	1000	1000	Carson (2012) 0	
Rennell-Bellona	Bellona/Mi Ngiki	9	1000	1000	Carson (2012) 0	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Reef Islands and Santa Cruz Islands	Duff Islands	9	1000	1000	Carson (2012) 0	Yes
Kapinga- marangi	Kapinga- marangi	9	1000, 700	1000	Carson (2012) 0	
Futuna and Aniwa	Futuna and Aniwa	9	1000, 900	1000	Carson (2012) 0	
Society Islands	Raro Matai	10	960, 800	960	Anderson and Sinoto (2002), Kahn (2012), Wilmshurst et al (2011) and Parkes (1997) as cited in Rieth & Cochrane (2018)	
Society Islands	Nia Matai	10	960, 800	960	Anderson and Sinoto (2002), Kahn (2012), Wilmshurst et al (2011) and Parkes (1997) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Marquesas	South Marquesas/Te Fenua 'Enata	10	950, 750	950	Allen (2004) and Allen and McAllister (2010, 2013) as cited in Rieth & Cochrane (2018)	
Marquesas	North Marquesas/Te Henua 'Enana	10	950, 750	950	Allen (2004) and Allen and McAllister (2010, 2013) as cited in Rieth & Cochrane (2018)	
Hawaii	Hawaii	10	950, 850	950	Athens et al (2014) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Mangareva	Mangareva	10	920, 660	920	Anderson et al (2003), Green and Weisler (2002) and Kirch et al (2010) as cited in Rieth & Cochrane (2018)	
Southern Cook Islands	Rarotonga	10	900, 680	900	Allen and Morrison (2013), Allen and Wallace (2007), Kirch et al (1995) and Parkes (1997) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Southern Cook Islands	Nga Pu Toru	10	900, 680	900	Allen and Morrison (2013), Allen and Wallace (2007), Kirch et al (1995) and Parkes (1997) as cited in Rieth & Cochrane (2018)	
Austral islands	Rimatara	10	850, 750	850	Bollt (2008), Kennett et al (2012) and Prebble and Anderson (2012) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Austral islands	Tupuai	10	850, 750	850	Bollt (2008), Kennett et al (2012) and Prebble and Anderson (2012) as cited in Rieth & Cochrane (2018)	
Austral islands	Raivavae	10	850, 750	850	Bollt (2008), Kennett et al (2012) and Prebble and Anderson (2012) as cited in Rieth & Cochrane (2018)	
Tokelau	Tokelau	10	750, 550	750	Petchey et al (2010) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Tokelau	Pukapuka	10	750, 550	750	Petchey et al (2010) as cited in Rieth & Cochrane (2018)	Yes
New Zealand, Auckland Islands	South Island (NZ)	10	750, 670	750	Higham et al (1999), McGlone and Wilmshurst (1999) and Wilmshurst et al (2008) as cited in Rieth & Cochrane (2018)	
New Zealand, Auckland Islands	North Island (NZ)	10	750, 670	750	Higham et al (1999), McGlone and Wilmshurst (1999) and Wilmshurst et al (2008) as cited in Rieth & Cochrane (2018)	

Name in source	**Island group (smallest)**	**Time depth settlement group**	**Date ranges**	**Oldest date**	**Sources**	**Based on inference from neighbouring island?**
Rapa Nui	Rapa Nui	10	750, 700	750	Hunt and Lipo (2006) and Mann et al (2008) as cited in Rieth & Cochrane (2018)	
Southern Cook Islands	Mangaia	10	960, 780	960	Walter and Reilly 2010 as cited in Walworth (2015)	
Rapa Iti	Rapa Iti	10	800	800	Kennett et al. 2006, 2012:196, 201 as cited in Walworth (2015)	
Austral islands	Rurutu	11	670, 550	670	Bollt (2008), Kennett et al (2012), Prebble and Anderson (2012) as cited in Rieth & Cochrane (2018)	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Northern Cook Islands	Māngarongaro	11	550, 300	550	Chikamori (1998) and Chikamori and Yoshida (1988) as cited in Rieth & Cochrane (2018)	
Northern Cook Islands	Rakahanga-Manihiki	11	550, 300	550	Chikamori (1998) and Chikamori and Yoshida (1988) as cited in Rieth & Cochrane (2018)	
Aotearoa	Rekohou	11	450, 400	450	McFadgen (1994) as cited in Rieth & Cochrane (2018)	
Mapia	Mapia	11	500, 400	600	Intoh & Ono (2007) 0	Yes
Tobi	Tobi	11	500, 400	600	Intoh & Ono (2007) 0	
Sonsorol	Sonsorol	11	600, 450	600	Intoh (2008) 0	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Luangiua	Luangiua	11	500	500	Kirch (2012) and Carson (2012) 0	Yes
Nukumanu	Nukumanu	11	500	500	Kirch (2012) and Carson (2012) 0	Yes
Nukuoro	Nukuoro	11	500	500	Kirch (2012) and Carson (2012) 0	
Nukuria	Nukuria	11	500	500	Kirch (2012) and Carson (2012) 0	Yes
Sikaiana	Sikaiana	11	500	500	Kirch (2012) and Carson (2012) 0	Yes
Takuu	Takuu	11	500	500	Kirch (2012) and Carson (2012) 0	Yes
None	Ouvea (West Uvean)	11	500	500	Kirch (2012) and Carson (2012) 0	Yes
Anuta	Anuta	11	500, 400	500	Carson (2012) 0	

Name in source	Island group (smallest)	Time depth settlement group	Date ranges	Oldest date	Sources	Based on inference from neighbouring island?
Mariana Islands	Northern Marianas	12	180	180	Fritz (1911), Spoehr (1954), Bowers (1953) and Quackenbush (1968) as cited in Ellis (2012)	

E Languages per island group

This table contains information on languages per island group, listed by the smallest possible island group. Additional columns also provide information which shared language-island group and overnight sailing distance-island group each smallest island group belongs to.

Table 12: Table on languages per island group

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Ambae	west2513, east2443	Ambae	Vanuatu and Temotu	167.85	-15.40
Ambrym	port1286, sout2859, daka1243, orko1234, lonw1238, nort2839	Ambrym	Vanuatu and Temotu	168.12	-16.25
Aneityum	anei1239	Aneityum	Vanuatu and Temotu	169.73	-20.22
Anewetak	mars1254	Marshall Islands (greater)	Anewetak	162.28	11.54
Anuta	anut1237	Anuta	Anuta-Tikopia	169.85	-11.58
Aore	aore1237	Aore	Vanuatu and Temotu	167.17	-15.61
Araki	arak1252	Araki	Vanuatu and Temotu	166.94	-15.64
Banaba	gilb1244	Tungaru and Tuvalu	Banaba	169.54	-0.89
Belep	nyal1254	Belep	Kanaky	163.64	-19.76
Bellona	renn1242	Mo-ava-mo-iki	Mo-ava-mo-iki	159.80	-11.30
Chuuk	chuu1238	Chuuk	Chuuk	151.82	7.36
Duff Islands	pile1238	Duff and Reef Islands	Vanuatu and Temotu	167.14	-9.84
East Futuna	east2447	East Futuna	East Futuna	-178.11	-14.30
Efate	mele1250, eton1255, sout2856, lele1267, nort2836	Efate	Vanuatu and Temotu	168.32	-17.63
Emae	emae1237	Emae	Vanuatu and Temotu	168.40	-17.11
Epi	bier1246, lewo1242, maii1238, baki1244, bier1244, lame1260	Epi	Vanuatu and Temotu	168.24	-16.71

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Erromango	siee1239, ifoo1237, urav1235	Erromango	Vanuatu and Temotu	169.20	-18.76
Etarik	nort2836	Efate	Vanuatu and Temotu	168.46	-17.27
Futuna and Aniwa	futu1245	Futuna and Aniwa	Vanuatu and Temotu	169.88	-19.40
Gaua	weta1242, lako1245, koro1308, nume1241	Gaua	Vanuatu and Temotu	167.51	-14.26
Guam	cham1312	Laguas yan g ani	Laguas yan g ani	144.72	13.37
Hawaii	hawa1245	Hawaii	Hawai'i	-157.34	21.15
Hereheretue	tuam1242	Tuamotu	Hereheretue	-144.95	-19.86
Hiu	hiww1237	Hiu	Vanuatu and Temotu	166.56	-13.14
Kadavu	kada1285	Kadavu	Fiji	178.34	-19.01
Kanaky (New Caledonia main island)	nume1242, dumb1241, xara1243, tiri1258, xara1244, zire1240, neku1237, orow1242, bwat1240, arho1237, ajie1238, arha1237, paic1239, haek1239, have1241, hmwa1243, cemu1238, pwaa1237, pwap1237, pije1237, vama1243, fwai1237, nemi1240, jawe1237, yuag1237, waam1236, caac1237, kuma1276, bala1316	Kanaky (New Caledonia main island)	Kanaky	165.64	-21.55
Kapinga-marangi	kapi1249	Kapinga-marangi	Kapinga-marangi	154.79	1.07
Kosrae	kosr1238	Kosrae	Kosrae	163.01	5.33
Lau	laua1243	Lau	Fiji	-171.68	-18.03
Leveuka (Lomaviti)	loma1261	Lomaviti	Fiji	178.77	-17.74

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Lifou	dehu1237	Lifou	Kanaky	167.12	-20.80
Loh-Toga	loto1240	Loh-Toga	Vanuatu and Temotu	166.63	-13.30
Lomaviti	loma1261	Lomaviti	Fiji	109.67	-18.31
Lomlom	ayiw1239	Duff and Reef Islands	Vanuatu and Temotu	166.26	-10.31
Lopevi	paam1238	Paama	Vanuatu and Temotu	168.32	-16.51
Luangiua	onto1237	Luangiua	Nukumanu-Luangiua	159.48	-5.34
Maewo	baet1237, cent2058, mari1426	Maewo	Vanuatu and Temotu	168.18	-15.13
Mafea	mafe1237	Mafea	Vanuatu and Temotu	167.23	-15.42
Malakula	navw1234, nasv1234, nisv1234, malf1237, sout2857, mask1242, port1285, axam1237, nati1244, lete1241, labo1244, burm1263, aulu1238, nasa1240, bwen1239, dixo1238, rere1240, vivt1234, katb1237, unua1237, vinm1237, niti1249, ling1265, lare1249, litz1237, bign1238, mara1399, urip1239, maee1241, malu1245, nese1235, mpot1241, avok1244	Malakula	Vanuatu and Temotu	167.64	-16.40
Mangaia	raro1241	Southern Cook Islands	Mangaia	-157.93	-21.91
Mangareva	mang1401	Mangareva	Mangareva	-134.93	-23.14
Māngarongaro	penr1237	Māngarongaro	Māngarongaro	-157.97	-9.03
Mapia	mapi1250	Mapia	Mapia	134.30	0.90
Marutea	tuam1242	Tuamotu	Marutea	-136.06	-21.52

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Mataso	nort2836	Efate	Vanuatu and Temotu	168.42	-17.26
Merelava group	merl1237	Merelava group	Vanuatu and Temotu	168.05	-14.42
Morane	tuam1242	Tuamotu	Morane	-137.12	-23.15
Mortlock	mort1237	Mortlock (greater)	Chuuk	153.56	5.53
Mota	mota1237	Mota	Vanuatu and Temotu	167.69	-13.85
Mota Lava	motl1237	Mota Lava	Vanuatu and Temotu	167.65	-13.69
Mwoakilloa	moki1238	Mwoakilloa	Pohnpei	159.78	6.67
Namonuito	namo1247	Namonuito	Chuuk	150.22	8.76
Napuka	tuam1242	Tuamotu	Napuka	-141.27	-14.16
Nauru	naur1243	Nauru	Nauru	166.93	-0.52
Nendö	nalo1235, nang1262, natu1246	Nendö	Vanuatu and Temotu	165.97	-10.75
Nengone	neng1238	Nengone	Kanaky	167.87	-21.48
Nga Pu Toru	raro1241	Southern Cook Islands	Nga Pu Toru	-159.31	-19.13
Ngatik	pohn1238	Pohnpei	Pohnpei	157.31	5.84
Nguluwan	ngul1236	Nguluwan	Waab	137.49	8.40
Nia Matai	tahi1242	Tahiti	Nia Matai	-150.80	-16.91
Niu	gilb1244	Tungaru and Tuvalu	Tuvalu	177.17	-7.27
Niuafoou	niua1240	Niuafoou	Niuafoou	-175.64	-15.60
Niuatoputapu	niua1241	Niuatoputapu	Niuatoputapu	-173.74	-15.92

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Niue	niue1239	Niue	Niue	-169.86	-19.05
North Island (NZ)	maor1246	Aotearoa	Aotearoa	175.17	-36.71
North Marquesas/Te Henua 'Enana	nort2845	North Marquesas/Te Henua 'Enana	Marquesas	-140.08	-8.81
Northern Marianas	caro1242	Laguas yan g ani	Laguas yan g ani	145.58	17.98
Nukumanu	nuku1258	Nukumanu	Nukumanu-Luangiua	159.41	-4.52
Nukuoro	nuku1260	Nukuoro	Nukuoro	154.98	3.86
Nukuria	nuku1259	Nukuria	Nukuria (New Guinea Polynesian outlier special case)	154.61	-3.25
Nukutaveke	tuam1242	Tuamotu	Nukutaveke	-138.82	-21.61
Oroluk	mort1237	Mortlock (greater)	Oroluk	154.38	7.35
Ouvea	iaai1238, west2516	Ouvea	Kanaky	166.58	-20.56
Ouvea (Iaai)	iaai1238	Ouvea	Kanaky	166.66	-20.48
Ouvea (West Uvean)	west2516	Ouvea	Kanaky	166.49	-20.60
Paafang	paaf1237	Paafang	Chuuk	151.96	8.61
Paama	paam1238	Paama	Vanuatu and Temotu	168.22	-16.47
Palau	pala1344	Palau	Palau	134.53	7.40
Paumotu	tuam1242	Tuamotu	Paumotu	-144.42	-16.38

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Pentecost	saaa1241, seke1241, apma1240, sowa1244, hano1246	Pentecost	Vanuatu and Temotu	168.19	-15.67
Pingelap	ping1243	Pingelap	Pohnpei	160.69	6.22
Pohnpei	pohn1238	Pohnpei	Pohnpei	158.22	6.87
Pollap	poll1238	Pollap	Chuuk	149.41	7.59
Puka-puka	tuam1242	Tuamotu	Puka-puka	-138.82	-14.82
Pukapuka	puka1242	Pukapuka	Pukapuka	-165.76	-11.05
Puluwat	pulu1242	Puluwat	Chuuk	149.21	7.25
Raivavae	raiv1237	Raivavae	Raivavae	-147.63	-23.87
Rakahanga-Manihiki	raka1237	Rakahanga-Manihiki	Rakahanga-Manihiki	-161.01	-10.40
Ralik	mars1254	Marshall Islands (greater)	Ralik	167.70	8.69
Rapa Iti	rapa1245	Rapa Iti	Rapa Iti	-144.14	-27.68
Rapa Nui	rapa1244	Rapa Nui	Rapa Nui	-109.36	-27.12
Raro Matai	tahi1242	Tahiti	Raro Matai	-154.36	-16.44
Rarotonga	raro1241	Southern Cook Islands	Rarotonga	-159.75	-21.24
Ratak	mars1254	Marshall Islands (greater)	Ratak	170.85	8.27
Reao	tuam1242	Tuamotu	Reao	-136.72	-18.42
Reef Islands	pile1238	Duff and Reef Islands	Vanuatu and Temotu	166.05	-10.18
Rekohou	mori1267	Rekohou	Rekohou	-176.27	-44.14
Rennelle	renn1242	Mo-ava-mo-iki	Mo-ava-mo-iki	160.25	-11.64
Rimatara	rima1237	Rimatara	Rimatara	-154.71	-21.81
Rotuma	rotu1241	Rotuma	Rotuma	177.03	-12.49

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Rurutu	ruru1237	Rurutu	Rurutu	-152.32	-22.60
Saipan	tana1281, cham1312, caro1242	Laguas yan g ani	Laguas yan g ani	145.75	15.19
Samoa	samo1305	Samoa	Samoa	-171.07	-14.09
Santo	akei1237, wail1242, tang1347, nara1263, ambl1237, tamb1253, polo1242, fort1240, rori1237, moro1286, butm1237, lore1244, navu1237, mere1242, wusi1237, shar1244, tial1239, tolo1255, tasm1246, saka1289, piam1242, noku1237, vuna1239, valp1237	Santo	Vanuatu and Temotu	166.96	-15.24
Satawal	sata1237	Satawal	Satawal	147.04	7.36
Satawal (Woleai speaking)	wole1240	Woleai	Satawal	146.16	7.51
Sikaiana	sika1261	Sikaiana	Sikaiana	162.87	-8.37
Sonsorol	sons1242	Sonsorol	Sonsorol	132.27	4.84
Sorol	ulit1238	Ulithi (greater)	Sorol	140.38	8.15
South Island (NZ)	maor1246	Aotearoa	Aotearoa	169.16	-44.73
South Marquesas/Te Fenua 'Enata	sout2866	South Marquesas/Te Fenua 'Enata	Marquesas	-138.92	-9.90
Takuu	taku1257	Takuu	Takuu	157.02	-4.74
Tamambo	malo1243	Tamambo	Vanuatu and Temotu	167.21	-15.70
Tangoa	tang1347	Santo	Vanuatu and Temotu	166.98	-15.60
Tanna	kwam1252, sout2869, whit1269, lena1238, nort2847	Tanna	Vanuatu and Temotu	169.33	-19.49
Tatakoto	tuam1242	Tuamotu	Tatakoto	-138.39	-17.35

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Taveuni	gone1237	Vanua Levu	Fiji	179.96	-16.93
Te Anu	tean1237	Te Anu	Vanuatu and Temotu	166.96	-11.63
Tegua	loto1240	Loh-Toga	Vanuatu and Temotu	166.58	-13.26
Tikopia	tiko1237	Tikopia	Anuta-Tikopia	168.81	-12.28
Tinian	cham1312	Laguas yan g ani	Laguas yan g ani	145.57	14.90
Tobi	tobi1238	Tobi	Tobi	131.12	3.01
Tokelau	toke1240	Tokelau	Tokelau	-171.66	-9.17
Tonga	tong1325	Tonga	Tonga	-174.48	-19.66
Tongoa	nama1268	Tongoa	Vanuatu and Temotu	168.59	-16.98
Tungaru	gilb1244	Tungaru and Tuvalu	Tungaru	173.68	0.83
Tupuai	tubu1240	Tupuai	Tupuai	-149.43	-23.35
Tureia	tuam1242	Tuamotu	Tureia	-140.63	-21.69
Tutuba	tutu1241	Tutuba	Vanuatu and Temotu	167.27	-15.58
Tuvalu	tuva1244	Tungaru and Tuvalu	Tuvalu	178.86	-8.38
Ulithi	ulit1238	Ulithi (greater)	Waab	139.73	9.97
Ureparapara	leha1244, leha1243	Ureparapara	Vanuatu and Temotu	167.32	-13.53
Utupua	tani1255, asum1237, amba1266	Utupua	Vanuatu and Temotu	166.45	-11.25
Uvea (Wallis)	wall1257	Uvea (Wallis)	Uvea (Wallis)	-176.18	-13.29

Island group (smallest)	**glottocodes**	**Island group (shared languages)**	**Island group (overnight distance)**	**Longitude (mean)**	**Latitude (mean)**
Vanikoro	tane1237, vano1237	Vanikoro	Vanuatu and Temotu	166.90	-11.66
Vanua Lava	vure1239, vera1241, leme1238	Vanua Lava	Vanuatu and Temotu	167.55	-13.86
Vanua Levu	gone1237	Vanua Levu	Fiji	167.82	-16.54
Vao	vaoo1237	Vao	Vanuatu and Temotu	167.30	-15.91
Vatulele	west2519	Viti Levu + Yasawa	Fiji	177.65	-18.50
Viti Levu	sout2864, nort2842, west2519, namo1248	Viti Levu + Yasawa	Fiji	178.13	-17.83
Wolei	wole1240	Woleai	Woleai	143.99	7.50
Yap	yape1248	Yap	Waab	138.13	9.52
Yasawa (greater)	west2519	Viti Levu + Yasawa	Fiji	177.30	-17.18

F Data tables for island groups (shared language and overnight distance

This section contains the necessary data for the analysis in chapter ??. The data is divided up into the two kinds of island groups: shared language (section .1 and overnight sailing distances (section .2). Each section contains two tables: a) data on political complexity, time depth, area, shoreline, isolation and latitude and b) the environmental variables from the ecoClimate datbase.

Tables 13 and 15 (part 1) contain data on:

- language count (response variable)
- EA033 (aka political complexity) - mean score per island group (see section ?? for more details)
- Settlement wave order - oldest per group (see section ?? for more details)
- Area (log10) (see section ?? for more details)
- Shoreline (log10) (see section ?? for more details)
- Ratio shoreline to ratio (see section ?? for more details)
- Isolation (see section ?? for more details)
- Latitude (absolute mean)

Tables 14 and 15 (part 2) contain data on the variables related to temperature and rainfall from the ecoClimate database (Lima-Ribeiro et al. 2015):

- Bio1: Annual mean temperature
- Bio4: Temperature seasonality (standard deviation *100) - mean for island group
- Bio12: Annual precipitation/rainfall (mm/m2)- mean for island group
- Bio15: Precipitation/rainfall seasonality (coefficient of variation)- mean for island group

All measurements, except the response variable, were scaled and centred to make the coeffecients easier to interpret and compare. Area, shoreline and Isolation were also log-10-transformed. For reasons of space, this data is displayed in two separate tables for each island grouping and the numbers are rounded to 4 digits.

.1 Tables of data on shared language island groups

Table 13: Data on shared language island groups - part 1

Island group	**lg count**	**EA033**	**Settlement wave order**	**Area (log10)**	**Shoreline (log10)**	**Ratio coastline to area**	**Isolation (log10)**	**Latitude (abs)**
Ambae	2	-1.153	1.0728	0.412	-0.1928	-0.4309	-1.0546	-0.0077
Ambrym	6	-1.153	1.0728	0.6118	-0.0111	-0.4461	-1.3899	0.1077
Aneityum	1	0.162	1.0728	0.0039	-0.5892	-0.3998	-0.4694	0.6416
Anuta	1	-1.153	-1.3847	-2.326	-2.4838	-3.4044	-0.1573	-0.5209
Aore	1	-1.153	1.0728	-0.3852	-0.7321	-0.2546	-2.0755	0.0206
Aotearoa	1	0.162	-1.0775	3.0738	3.038	-0.421	1.5649	3.4026
Chuuk	1	0.162	-0.156	-0.1816	0.3867	0.0656	-0.0556	-1.0899
Duff and Reef Islands	2	-1.153	1.38	-0.9389	-0.8752	0.1603	0.0175	-0.7194
East Futuna	1	0.162	-0.156	-0.2069	-0.4176	-0.231	0.2604	-0.1558
Efate	5	0.162	1.0728	0.7573	0.4874	-0.361	-0.6821	0.2602
Erromango	3	0.162	1.0728	0.7191	0.0765	-0.456	-0.6713	0.446
Futuna and Aniwa	1	0.162	-0.7703	-0.6944	-0.8409	-0.0788	-1.1846	0.532
Hawaii	1	2.792	-1.0775	1.9315	1.5919	-0.4346	2.2335	0.7677
Hiu	1	-1.153	1.0728	-0.4744	-0.815	-0.2343	-1.8819	-0.3121
Kadavu	1	1.477	0.7656	0.4747	0.5695	-0.2331	-0.1528	0.479
Kanaky (New Caledonia main island)	29	0.162	0.7656	2.2133	2.2754	-0.3746	0.0914	0.8208
Kapingamarangi	1	-1.153	-0.7703	-1.7805	-1.1191	3.5905	0.4727	-1.9369
Kosrae	1	1.477	-0.4631	-0.1311	-0.527	-0.3139	0.365	-1.3631
Laguas yan gani	3	-1.153	1.6871	0.7411	0.8054	-0.2727	0.9891	0.008
Lau	1	1.477	0.7656	0.4553	1.0877	-0.0749	-0.2701	0.3471
Lifou	1	0.162	0.7656	0.8239	0.3187	-0.4253	-0.5629	0.7196

Island group (shared language)	lg count	EA033 (mean)	Settlement wave order	Area (log10)	Shoreline (log10)	Ratio coastline to area	Isolation (log10)	Latitude (abs)
Loh-Toga	1	-1.153	1.38	-0.3376	-0.3824	-0.136	-1.8819	-0.2926
Luangiua	1	-1.153	-1.3847	-0.8461	-0.1797	0.4634	-0.4079	-1.3619
Maewo	3	-1.153	1.0728	0.2758	-0.0214	-0.327	-1.0546	-0.044
Malakula	33	-1.153	1.0728	1.0682	0.7173	-0.4004	-1.0065	0.1274
Mangareva	1	0.162	-1.0775	-0.654	-0.0856	0.2747	1.4029	1.0353
Mangarongaro	1	0.162	-1.3847	-0.9719	-0.1571	0.6778	0.5951	-0.8645
Marshall Islands (greater)	1	1.477	-0.4631	0.2237	1.6928	0.2413	0.9547	-0.8017
Mo-ava-mo-iki	1	0.162	-0.7703	0.6281	0.3012	-0.3665	0.7762	-0.5364
Mota	1	-1.153	1.0728	-1.0306	-1.497	-0.1457	-1.6094	-0.2155
Mota Lava	1	-1.153	1.0728	-0.695	-0.9476	-0.1347	-1.6094	-0.2368
Nengone	1	1.477	0.7656	0.5962	0.0308	-0.4296	-0.3928	0.8112
Niue	1	0.162	-0.4631	0.2188	-0.47	-0.4462	0.8819	0.4847
North Marquesas	1	-1.153	-1.0775	0.5651	0.5808	-0.2666	-0.0389	-0.8949
Nukuoro	1	-1.153	-1.3847	-1.7351	-1.0785	3.0451	0.4727	-1.5604
Paama	1	-1.153	1.0728	-0.3785	-0.714	-0.2509	-1.3899	0.1398
Palau	1	0.162	0.7656	0.3899	0.3744	-0.2545	0.5216	-1.0847
Pentecost	5	-1.153	1.0728	0.49	0.0174	-0.3973	-0.8451	0.0297
Pohnpei	1	1.477	-0.4631	0.3559	0.3325	-0.2524	0.0283	-1.225
Pukapuka	1	0.162	-1.0775	-1.2583	-1.0468	0.5725	0.8724	-0.5931
Raivavae	1	0.162	-1.0775	-0.8584	-0.5644	0.2506	0.1093	1.1339
Rakahanga-Manihiki	1	0.162	-1.3847	-0.8995	-0.4663	0.3652	0.5951	-0.6809
Rapa Nui	1	0.162	-1.0775	0.0319	-0.4237	-0.3542	2.0827	1.571
Rekohou	1	0.162	-1.3847	0.7581	0.6345	-0.3232	2.1367	3.8625

Island group (shared language)	**lg count**	**EA033 (mean)**	**Settlement wave order**	**Area (log10)**	**Shoreline (log10)**	**Ratio coastline to area**	**Isolation (log10)**	**Latitude (abs)**
Rimatara	1	0.162	-1.0775	-1.9055	-1.6653	4.74	0.5654	0.8561
Rotuma	1	1.477	-0.7703	-0.486	-0.5881	-0.1222	0.8469	-0.3994
Rurutu	1	0.162	-1.3847	-0.5315	-0.8	-0.1879	0.1907	0.962
Samoa	1	1.477	0.4584	1.23	0.9719	-0.388	0.4631	-0.1831
Santo	24	-1.153	1.38	1.3461	0.8771	-0.4394	-0.9964	-0.0045
Southern Cook Islands	1	0.162	-1.0775	0.1161	0.2018	-0.1815	0.8724	0.7149
Tahiti	1	1.477	-1.0775	0.9622	1.2296	-0.2452	0.9177	0.1644
Tanna	5	-1.153	1.0728	0.5294	-0.0726	-0.4363	-1.1846	0.5442
Tikopia	1	0.162	0.7656	-1.3358	-1.6566	0.1686	-0.1573	-0.4275
Tokelau	1	0.162	-1.0775	-0.9703	-0.0859	0.7219	0.9694	-0.8461
Tonga	1	1.477	0.4584	0.6086	1.1281	-0.1338	0.9672	0.5668
Tuamotu	1	0.162	-0.7703	0.6767	2.1038	0.0973	0.6982	0.4638
Tungaru and Tuvalu	2	0.162	-0.4631	0.3814	1.4521	0.0707	0.8311	-1.552
Tupuai	1	0.162	-1.0775	-0.5	-0.9097	-0.261	0.1093	1.0632
Ulithi (greater)	1	0.162	-0.4631	-1.1221	-0.4171	0.7879	0.3421	-0.8608
Ureparapara	2	-1.153	1.0728	-0.5782	-0.9352	-0.2193	-1.1644	-0.2583
Uvea (Wallis)	1	0.162	-0.156	-0.2559	-0.3707	-0.1834	0.2604	-0.2908
Vanua Lava	3	-1.153	1.0728	0.3082	-0.1657	-0.3847	-1.4419	-0.2145
Viti Levu + Yasawa	4	1.477	0.7656	1.7558	1.4764	-0.4169	-0.1528	0.3207
Woleai	1	0.162	-0.4631	-1.0943	-0.3779	0.7585	0.5461	-1.0703
Yap	1	1.477	0.1512	-0.1703	-0.0766	-0.1205	-0.0619	-0.7989

Table 14: Data on shared language island groups - part 2

Island group (shared language)	**Bio1**	**Bio4**	**Bio12**	**Bio15**
Ambae	0.2317	0.2234	-0.04	0.6743
Ambrym	0.1008	0.4736	-0.295	0.9325
Aneityum	-0.4594	1.2994	-1.1962	1.2691
Anuta	0.6134	-0.6679	1.1671	-1.0279
Aore	0.1562	0.3759	-0.2043	0.9318
Aotearoa	-4.5187	2.4015	-1.2678	-1.5234
Chuuk	0.7055	-1.3196	1.5424	-0.9215
Duff and Reef Islands	0.6832	-0.9096	1.5898	-1.269
East Futuna	0.4487	-0.358	0.4804	-0.5337
Efate	-0.0348	0.7559	-0.6132	1.2509
Erromango	-0.2223	1.0252	-0.9229	1.3657
Futuna and Aniwa	-0.3366	1.1552	-1.0523	1.2789
Hawaii	-0.5574	0.0795	-1.7603	-0.9972
Hiu	0.4771	-0.2631	0.521	-0.0866
Kadavu	-0.2495	0.8779	-0.7783	0.6145
Kanaky (New Caledonia main island)	-0.7698	1.6115	-1.4999	1.5655
Kapingamarangi	0.7358	-1.4459	-0.1607	-0.6397
Kosrae	0.7072	-1.4114	1.9197	-0.6973
Laguas yan gani	0.4742	-0.3963	-0.6537	1.779
Lau	-0.0726	0.5688	-0.5432	0.3829
Lifou	-0.5877	1.4496	-1.3099	1.5315

Island group	**Bio1**	**Bio4**	**Bio12**	**Bio15**
Loh-Toga	0.4771	-0.2631	0.521	-0.0866
Luangiua	0.7529	-1.22	1.5697	-0.945
Maewo	0.2491	0.1899	-0.0005	0.5776
Malakula	0.0537	0.5781	-0.4094	1.1292
Mangareva	-0.8035	0.8928	-0.5258	-1.4485
Mangarongaro	0.655	-1.2584	0.537	-0.1419
Marshall Islands (greater)	0.6182	-1.0049	0.4464	0.4457
Mo-ava-mo-iki	0.5358	-0.4607	0.819	-0.3542
Mota	0.4323	-0.1693	0.3934	0.0496
Mota Lava	0.4323	-0.1693	0.3934	0.0496
Nengone	-0.7361	1.5631	-1.4122	1.3606
Niue	-0.2497	0.6178	-0.8222	0.2197
North Marquesas	0.4106	-0.9246	-0.5088	3.372
Nukuoro	0.7277	-1.4902	1.2671	-0.2987
Paama	0.0712	0.5365	-0.3725	1.0192
Palau	0.7285	-1.3326	1.438	-0.9203
Pentecost	0.1867	0.308	-0.1245	0.7321
Pohnpei	0.7117	-1.3677	1.87	-0.7674
Pukapuka	0.6769	-1.1194	0.965	-0.931
Raivavae	-1.0083	1.2524	-0.9721	-0.6001
Rakahanga-Manihiki	0.6766	-1.1918	0.7244	-0.6963
Rapa Nui	-1.6529	0.8816	-1.6697	1.0276
Rekohou	-4.7121	1.7901	-1.6218	-1.1949
Rimatara	-0.6668	0.9622	-1.0044	-0.2696
Rotuma	0.5943	-0.6832	1.0353	-1.0745

Island group	**Bio1**	**Bio4**	**Bio12**	**Bio15**
Rurutu	-0.8107	1.0986	-1.0221	-0.4214
Samoa	0.468	-0.5139	0.4714	-0.606
Santo	0.2004	0.3128	-0.1617	0.9115
Southern Cook Islands	-0.4885	0.7497	-0.9644	-0.0844
Tahiti	0.2166	-0.167	-0.1862	-0.3595
Tanna	-0.302	1.1278	-1.0206	1.3737
Tikopia	0.5726	-0.5438	0.958	-0.7527
Tokelau	0.724	-1.2963	1.782	-1.3956
Tonga	-0.3201	0.8327	-0.9117	0.3454
Tuamotu	-0.1446	0.1743	-0.4247	-0.8314
Tungaru and Tuvalu	0.6869	-1.3212	0.5521	-0.9004
Tupuai	-0.9186	1.1854	-0.9839	-0.5412
Ulithi (greater)	0.697	-1.2277	1.0142	-0.971
Ureparapara	0.4299	-0.1553	0.3713	0.1061
Uvea (Wallis)	0.5494	-0.601	0.7755	-0.8339
Vanua Lava	0.4318	-0.1665	0.389	0.0609
Viti Levu + Yasawa	-0.0866	0.6029	-0.4932	0.7092
Woleai	0.7172	-1.3396	1.4332	-0.9165
Yap	0.6912	-1.1947	0.9631	-1.027

.2 Table of data on overnight distance island groups

Table 15: Data on overnight distance island groups - part 1

Island group (overnight sailing)	**lg count**	**EA033 (mean)**	**Settlement wave order**	**Area (log10)**	**Shoreline (log10)**	**Ratio coastline to area**	**Isolation (log10)**	**Latitude (abs)**
Anewetak	1	1.464	0.1596	-0.7316	-0.3102	0.33	0.8506	-0.4186
Anuta-Tikopia	2	-1.058	1.9227	-0.8831	-1.3429	-0.0523	0.4121	-0.3717
Aotearoa	1	-0.217	-0.7219	2.9933	2.7539	-0.4749	1.5977	3.0406
Banaba	1	-0.217	0.1596	-0.8198	-1.4356	-0.19	-0.4898	-1.6807
Chuuk	6	-0.217	0.6004	0.2109	0.7392	-0.1196	-0.5209	-0.8948
East Futuna	1	-0.217	0.6004	0.0855	-0.3638	-0.3544	-0.8047	-0.0914
Fiji	8	1.464	1.9227	2.025	2.0128	-0.4185	0.2679	0.3222
Hawai'i	1	3.145	-0.7219	1.9809	1.4492	-0.4835	3.3248	0.7215
Hereheretue	1	-0.217	-0.2812	-0.9789	-0.9329	0.4162	0.3483	0.5685
Kapingamarangi	1	-1.898	-0.2812	-1.3092	-0.9967	2.0696	-0.4027	-1.6591
Kosrae	1	1.464	0.1596	0.1527	-0.4625	-0.4069	0.5216	-1.154
Laguas yan gani	3	-1.898	3.245	0.9258	0.7396	-0.3808	0.8062	0.0528
Mangaia	1	-0.217	-0.7219	-0.1381	-0.8989	-0.4189	-1.0424	0.8114
Mangareva	1	-0.217	-0.7219	-0.3107	-0.0643	-0.0336	-1.1215	0.957
Mangarongaro	1	-0.217	-1.1627	-0.5925	-0.1288	0.2221	-0.1216	-0.7152
Marquesas	2	-1.898	-0.7219	0.989	0.8634	-0.3748	0.5102	-0.6774
Marutea	1	-0.217	-0.2812	-0.4065	0.0496	0.0926	-1.1124	0.7648
Mo-ava-mo-iki	1	-0.217	-0.2812	0.8256	0.2847	-0.4403	0.2581	-0.4264
Morane	1	-0.217	-0.2812	-1.7354	-1.6663	-4.541	-0.9491	0.9581
Napuka	1	-0.217	-0.2812	-0.7624	-0.7294	0.1465	-0.5886	-0.1072
New Caledonia (incl loyalties)	34	0.203	1.9227	2.2524	2.1357	-0.4404	1.0143	0.685
Nga Pu Toru	1	-0.217	-0.7219	0.0444	-0.0546	-0.2498	-0.6983	0.482
Nia Matai	1	1.464	-0.7219	1.119	1.0811	-0.3697	0.2388	0.2181

Island group	**lg count**	**EA033**	**Settlement wave order**	**Area (log10)**	**Shoreline (log10)**	**Ratio coastline to area**	**Isolation (log10)**	**Latitude (abs)**
Niue	1	-0.217	0.1596	0.4629	-0.4111	-0.4908	0.4605	0.4723
Nukumanu-Luangiua	2	-1.898	-1.1627	-0.4042	0.0608	0.0948	-0.6368	-1.2014
Nukuoro	1	-1.898	-1.1627	-1.269	-0.9601	1.7236	-0.3736	-1.3277
Nukutaveke	1	-0.217	-0.2812	-0.2491	0.297	0.0445	-1.4	0.7752
Palau	1	-0.217	1.9227	0.6146	0.3508	-0.3692	0.0982	-0.909
Paumotu	1	-0.217	-0.2812	0.8024	1.7961	-0.1435	-0.34	0.1558
Pohnpei	3	1.464	0.1596	0.5883	0.3681	-0.3575	0.2463	-1.0271
Puka-puka	1	-0.217	-0.2812	-1.1088	-0.9813	0.7682	-0.6484	-0.0292
Pukapuka	1	-0.217	-0.7219	-0.8463	-0.9315	0.1553	0.4712	-0.4763
Raivavae	1	-0.217	-0.7219	-0.4919	-0.4962	-0.0489	-1.103	1.0438
Rakahanga-Manihiki	1	-0.217	-1.1627	-0.5283	-0.4078	0.0238	-0.1294	-0.5536
Ralik	1	1.464	0.1596	0.2021	1.1464	-0.0103	0.3421	-0.7553
Rapa Nui	1	-0.217	-0.7219	0.2972	-0.3693	-0.4325	2.9514	1.4285
Raro Matai	1	1.464	-0.7219	-0.6646	-0.4141	0.1782	0.537	0.1628
Rarotonga	1	-0.217	-0.7219	-0.0201	-0.7218	-0.4135	-1.0879	0.7315
Ratak	1	1.464	0.1596	0.1965	1.0886	-0.022	1.024	-0.8058
Reao	1	-0.217	-0.2812	-0.2171	-0.0008	-0.081	-0.8064	0.3971
Rekohou	1	-0.217	-1.1627	0.9408	0.5854	-0.4128	3.0747	3.4454
Rimatara	1	-0.217	-0.7219	-1.42	-1.4895	2.7986	-0.3	0.7992
Rotuma	1	1.464	-0.2812	-0.1619	-0.5177	-0.2853	0.3505	-0.3059
Rurutu	1	-0.217	-1.1627	-0.2021	-0.7088	-0.327	-0.9579	0.8925
Samoa	1	1.464	1.4819	1.3592	0.8898	-0.454	-0.0509	-0.1154
Satawal	2	-0.217	0.1596	-1.0619	-0.9888	0.5989	-0.7546	-0.9046
Sorol	1	-0.217	0.1596	-1.4348	-1.2832	3.8314	-0.4943	-0.8201

Island group	**lg count**	**EA033**	**Settlement wave order**	**Area (log10)**	**Shoreline (log10)**	**Ratio coastline to area**	**Isolation (log10)**	**Latitude (abs)**
Tatakoto	1	-0.217	-0.2812	-0.8742	-0.6363	0.3937	-0.7672	0.2706
Tokelau	1	-0.217	-0.7219	-0.5911	-0.0646	0.25	0.3738	-0.699
Tonga	1	1.464	1.4819	0.8084	1.0307	-0.2927	0.3245	0.5446
Tungaru	1	-0.217	0.1596	0.5582	1.2193	-0.1654	0.6501	-1.6876
Tupuai	1	-0.217	-0.7219	-0.1742	-0.8078	-0.3734	-1.1395	0.9815
Tureia	1	-0.217	-0.2812	-1.1205	-0.8803	0.9146	-1.1931	0.7854
Tuvalu	2	-0.217	0.1596	-0.2199	0.258	0.0082	0.2333	-0.8583
Uvea (Wallis)	1	-0.217	0.6004	0.0421	-0.3215	-0.3242	-0.8592	-0.2102
Vanuatu + Temotu	128	-1.495	2.8042	1.8961	1.8641	-0.4152	0.2991	-0.0042
Waab	3	0.623	1.0411	0.1517	0.2073	-0.2291	-0.4886	-0.6838
Woleai	1	-0.217	0.1596	-0.7967	-0.4919	0.3371	-0.2046	-0.8972

Table 16: Data on overnight distance island groups - part 2

Island group (shared language)	**Bio1**	**Bio4**	**Bio12**	**Bio15**
group	**Bio1**	**Bio4**	**Bio12**	**Bio15**
Anewetak	0.5446	-0.5433	-0.3363	1.9468
Anuta-Tikopia	0.5647	-0.3731	0.9917	-0.5428
Aotearoa	-4.225	2.6003	-1.2214	-1.2779
Banaba	0.6201	-1.1146	-1.2755	-0.0835
Chuuk	0.6678	-1.0585	1.3105	-0.4682
East Futuna	0.4295	-0.128	0.4389	-0.1287
Fiji	-0.0455	0.7863	-0.4603	1.1273
Hawai'i	-0.5133	0.3046	-1.689	-0.667
Hereheretue	-0.1957	0.6124	-0.4328	-0.3884
Kapingamarangi	0.6984	-1.2036	-0.17	-0.2518
Kosrae	0.6716	-1.1695	1.8057	-0.3187
Laguas yan gani	0.4533	-0.1659	-0.6382	2.5567
Mangaia	-0.6338	1.1814	-1.0551	0.2982
Mangareva	-0.7438	1.1086	-0.5167	-1.191
Mangarongaro	0.6228	-1.0182	0.4926	0.3262
Marquesas	0.3983	-0.6487	-0.5994	4.0035
Marutea	-0.4532	0.8051	-0.4644	-1.2268
Mo-ava-mo-iki	0.511	-0.2295	0.7604	0.0797
Morane	-0.7565	1.1723	-0.5508	-1.0911
Napuka	0.3986	-0.3594	-0.3774	0.3367
New Caledonia (incl loyalties)	-0.5708	1.6947	-1.3215	2.3875
Nga Pu Toru	-0.1595	0.6216	-0.6946	0.475

Island group (overnight sailing)	**Bio1**	**Bio4**	**Bio12**	**Bio15**
Nia Matai	0.1862	0.099	-0.2064	0.0461
Niue	-0.225	0.8368	-0.7982	0.7461
Nukumanu-Luangiua	0.7203	-0.9959	1.2818	-0.6187
Nukuoro	0.6908	-1.2474	1.186	0.1441
Nukutaveke	-0.467	0.8699	-0.4806	-1.0572
Palau	0.6916	-1.0916	1.3483	-0.5776
Paumotu	0.2447	-0.0448	-0.1774	0.0035
Pohnpei	0.6742	-1.1171	1.7348	-0.4482
Puka-puka	0.3464	-0.3014	-0.4706	0.2851
Pukapuka	0.6433	-0.8808	0.8991	-0.5901
Raivavae	-0.9358	1.4642	-0.9405	-0.2059
Rakahanga-Manihiki	0.643	-0.9524	0.6706	-0.3176
Ralik	0.6077	-0.8523	0.6772	0.6463
Rapa Nui	-1.5398	1.0975	-1.603	1.6842
Raro Matai	0.2378	0.0226	-0.1819	0.101
Rarotonga	-0.5529	1.0986	-1.0501	0.4056
Ratak	0.6124	-0.907	0.8788	0.4323
Reao	-0.0057	0.2275	-0.3859	-0.5152
Rekohou	-4.4062	1.9958	-1.5575	-0.8965
Rimatara	-0.6158	1.1773	-0.9712	0.178
Rotuma	0.5659	-0.4495	0.9658	-0.7566
Rurutu	-0.7506	1.3121	-0.988	0.0016
Samoa	0.4475	-0.2821	0.4303	-0.2127
Satawal	0.6791	-1.1001	1.4016	-0.5817

Island group (overnight sailing)	**Bio1**	**Bio4**	**Bio12**	**Bio15**
Sorol	0.6778	-1.058	1.1661	-0.763
Tatakoto	0.1381	0.0459	-0.33	-0.1765
Tokelau	0.6874	-1.0557	1.675	-1.1295
Tonga	-0.2909	1.0492	-0.8832	0.8921
Tungaru	0.5813	-1.0937	-0.9395	0.3468
Tupuai	-0.8517	1.3979	-0.9518	-0.1374
Tureia	-0.5146	0.9556	-0.5366	-0.9137
Tuvalu	0.7045	-1.0565	2.1213	-1.2406
Uvea (Wallis)	0.5238	-0.3683	0.7191	-0.4773
Vanuatu + Temotu	0.2247	0.406	0.0398	1.0235
Waab	0.6606	-0.9729	0.9372	-0.6737
Woleai	0.6836	-1.1032	1.3233	-0.5486

G Table of comparison of computational ancestral state reconstruction to findings in historical linguistics

Table 17: Ancestral State Reconstruction per feature for Oceanic languages

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB028	Is there a distinction between inclusive and exclusive?	Proto-Polynesian	Present	Marck (2000:4)	True Positive	True Positive	True Positive	True Positive
GB031	Is there a dual or unit augmented form (in addition to plural or augmented) for all person categories in the pronoun system?	Proto-Polynesian	Present	Marck (2000:4)	True Positive	True Positive	True Positive	True Positive
GB035	Are there three or more distance contrasts in demonstratives?	Proto-Polynesian	Present	Clark (1976:77)	True Positive	True Positive	True Positive	True Positive
GB059	Is the adnominal possessive construction different for alienable and inalienable nouns?	Proto-Polynesian	Present	Marck (2000:4), Clark (1976:47)	True Positive	True Positive	True Positive	True Positive
GB070	Are there morphological cases for non-pronominal core arguments (i.e. S/A/P)?	Proto-Polynesian	Absent	(Chung 1978:261-261)	True Negative	True Negative	True Negative	True Negative

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB071	Are there morphological cases for pronominal core arguments (i.e. S/A/P)?	Proto-Central Pacific	Absent	Kikusawa (2002:1)	True Negative	True Negative	True Negative	True Negative
GB074	Are there prepositions?	Proto-Polynesian	Present	Chung (1978:26), Clark (1976:24, 43)	True Positive	True Positive	True Positive	True Positive
GB079	Do verbs have prefixes/proclitics, other than those that only mark A, S or P (do include portmanteau: A and S + TAM)?	Proto-Central Pacific	Present	Pawley (1970:352)	True Positive	True Positive	True Positive	True Positive
GB113	Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?	Proto-Central Pacific	Present	Pawley (1970:352), Marck (2000:4)	True Positive	True Positive	True Positive	True Positive
GB113	Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?	Proto-Polynesian	Present	Marck (2000:4), Jonsson (1998:57), Pawley (1970:352)	True Positive	True Positive	True Positive	True Positive

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB133	Is a pragmatically unmarked constituent order verb-final for transitive clauses?	Proto-Central Pacific	Absent	Kikusawa (2002:122)	True Negative	True Negative	True Negative	True Negative
GB133	Is a pragmatically unmarked constituent order verb-final for transitive clauses?	Proto-Polynesian	Absent	(Chung 1978:15), Pawley (1973:118)	True Negative	True Negative	True Negative	True Negative
GB139	Is there a difference between imperative (prohibitive) and declarative negation constructions?	Proto-Polynesian	Present	Pawley (1973:142), Ross (2007:292), Clark (1976:xiii, 125), Lynch et al. (2011:89)	True Positive	True Positive	True Positive	True Positive
GB155	Are causatives formed by affixes or clitics on verbs?	Proto-Central Pacific	Present	Pawley (1970:352)	True Positive	True Positive	True Positive	True Positive
GB193a	Is the order of the adnominal property and the noun ANM-N?	Proto-Polynesian	Absent	Clark (1976:xiv)	True Negative	True Negative	True Negative	True Negative
GB193b	Is the order of the adnominal property and the noun N-ANM?	Proto-Polynesian	Present	Ross (2004:495)	True Positive	True Positive	True Positive	True Positive

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB299	Can standard negation be marked by a non-inflecting word ("auxiliary particle")?	Proto-Polynesian	Present	Clark (1976:130)	True Positive	True Positive	True Positive	True Positive
GB327	Can the relative clause follow the noun?	Proto-Polynesian	Present	Clark (1976:74)	True Positive	True Positive	True Positive	True Positive
GB328	Can the relative clause precede the noun?	Proto-Polynesian	Absent	Clark (1976:74)	True Negative	True Negative	True Negative	True Negative
GB519	Can mood be marked by a non-inflecting word ("auxiliary particle")?	Proto-Polynesian	Present	Chung (1978:20)	True Positive	True Positive	True Positive	True Positive
GB520	Can aspect be marked by a non-inflecting word ("auxiliary particle")?	Proto-Central Pacific	Present	Pawley (1970:347)	True Positive	True Positive	True Positive	True Positive
GB520	Can aspect be marked by a non-inflecting word ("auxiliary particle")?	Proto-Polynesian	Present	Chung (1978:20)	True Positive	True Positive	True Positive	True Positive

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB522	Can the S or A argument be omitted from a pragmatically unmarked clause when the referent is inferrable from context ("pro-drop" or "null anaphora")?	Proto-Polynesian	Present	Chung (1978:30)	True Positive	True Positive	True Positive	True Positive
GB133	Is a pragmatically unmarked constituent order verb-final for transitive clauses?	Proto-Polynesian	Absent	Kikusawa (2002:122)	True Negative	True Negative	True Negative	True Negative
GB022	Are there prenominal articles?	Proto-Polynesian	Present	Clark (1976:xiv), Chung (1978:23), Clark (1976:53), Crowley (1985:136)	True Positive	True Positive	True Positive	True Positive
GB073	Are there morphological cases for independent oblique personal pronominal arguments (i.e. not S/A/P)?	Proto-Central Pacific	Absent	Kikusawa (2002:1)	True Negative	True Negative	True Negative	True Negative

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB020	Are there definite or specific articles?	Proto-Polynesian	Present	Chung (1978:23), Clark (1976:53)	True Positive	Half	True Positive	True Positive
GB021	Do indefinite nominals commonly have indefinite articles?	Proto-Polynesian	Present	Chung (1978:23), Clark (1976:53), Lynch et al. (2011:73)	True Positive	False Negative	True Positive	True Positive
GB093	Can the P argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004:498-499), Lynch et al. (2011:83), Evans (2001:112)	True Positive	Half	True Positive	True Positive
GB115	Is there a phonologically bound reciprocal marker on the verb?	Proto-Polynesian	Present	Marck (2000:4)	True Positive	Half	True Positive	True Positive
GB333	Is there a decimal numeral system?	Proto-Oceanic	Present	Lynch et al. (2011:72)	True Positive	Half	True Positive	True Positive
GB430	Can adnominal possession be marked by a prefix on the possessor?	Proto-Oceanic	Absent	Ross (2004:511)	True Negative	Half	True Negative	True Negative

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB433	Can adnominal possession be marked by a suffix on the possessed noun?	Proto-Oceanic	Present	Ross (2004:495), Pawley (1973:155)	True Positive	Half	True Positive	True Positive
GB028	Is there a distinction between inclusive and exclusive?	Proto-Oceanic	Present	Pawley (1973:112), Crowley (1985:184), Ross (2004:500), Lynch et al. (2011:67, 75)	True Positive	Half	True Positive	True Positive
GB074	Are there prepositions?	Proto-Oceanic	Present	Pawley (1973:167), Ross (2004:498)	True Positive	Half	True Positive	True Positive
GB058	Are there possessive classifiers?	Proto-Oceanic	Present	Pawley (1973:154), Ross (2004:498), Lynch et al. (2011:69, 77)	True Positive	Half	True Positive	True Positive
GB318	Is plural number regularly marked in the noun phrase by a dedicated phonologically free element?	Proto-Polynesian	Present	Clark (1976:59)	True Positive	Half	True Positive	True Positive

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	**ML result**	**Parsimony result**	**ML result**
GB131	Is a pragmatically unmarked constituent order verb-initial for transitive clauses?	Proto-Polynesian	Present	Marck (2000:4), Chung (1978:15), Pawley (1973:118)	True Positive	True Positive	Half	True Positive
GB131	Is a pragmatically unmarked constituent order verb-initial for transitive clauses?	Proto-Polynesian	Present	Kikusawa (2002:122)	True Positive	True Positive	Half	True Positive
GB193b	Is the order of the adnominal property and the noun N-ANM?	Proto-Oceanic	Present	Ross (2004:497), Lynch et al. (2011:74)	True Positive	Half	True Positive	True Positive
GB091	Can the A argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004:498-499), Lynch et al. (2011:83)	True Negative	Half	True Negative	True Negative
GB094	Can the P argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004:498-499), Lynch et al. (2011:83)	True Negative	Half	True Negative	True Negative
GB260	Can polar interrogation be indicated by a special word order?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	True Negative

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB130b	Is the order of S and V in intranstive clauses VS?	Proto-Polynesian	Present	Kikusawa (2002:122)	Half	Half	True Positive	True Positive
GB158	Are verbs reduplicated?	Proto-Oceanic	Present	Lynch et al. (2011:84), Evans (2001:112)	Half	Half	True Positive	True Positive
GB115	Is there a phonologically bound reciprocal marker on the verb?	Proto-Central Pacific	Present	Pawley (1970:352), Lynch et al. (2011:83)	True Positive	Half	True Positive	Half
GB092	Can the A argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004:498-499), Lynch et al. (2011:83)	Half	Half	True Positive	True Positive
GB130a	Is the order of S and V in intranstive clauses SV?	Proto-Polynesian	Absent	Kikusawa (2002:122)	Half	Half	True Negative	True Negative
GB521	Can tense be marked by a non-inflecting word ("auxiliary particle")?	Proto-Central Pacific	Present	Pawley (1970:347)	True Positive	Half	True Positive	Half
GB090	Can the S argument be indexed by a prefix/proclitic on the verb in the simple main clause?	Proto-Oceanic	Present	Ross (2004:498-499), Lynch et al. (2011:83)	Half	Half	True Positive	True Positive

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB025b	Is the order of adnominal demonstratives and nouns N-Dem?	Proto-Oceanic	Present	Lynch et al. (2011:72)	True Positive	Half	True Positive	Half
GB030	Is there a gender distinction in independent 3rd person pronouns?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half
GB031	Is there a dual or unit augmented form (in addition to plural or augmented) for all person categories in the pronoun system?	Proto-Oceanic	Present	Ross (2004:498), Kikusawa (2001:109), Lynch et al. (2011:69), Pawley (1973:173)	True Positive	Half	True Positive	Half
GB035	Are there three or more distance contrasts in demonstratives?	Proto-Oceanic	Present	Lynch et al. (2011:72)	True Positive	Half	True Positive	False Negative
GB051	Is there a gender/noun class system where sex is a factor in class assignment?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	False Positive
GB054	Is there a gender/noun class system where plant status is a factor in class assignment?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB065b	Is the order of possessor noun and possessed noun possessed-possessor?	Proto-Oceanic	Present	Pawley (1973:155-156), Ross (2004:512), Lynch et al. (2011:77)	True Positive	Half	True Positive	Half
GB075	Are there postpositions?	Proto-Oceanic	Absent	Ross (2004:500)	True Negative	Half	True Negative	Half
GB113	Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?	Proto-Oceanic	Present	Pawley (1973:171), Wilson (1982:123), Lynch et al. (2011:80, 92), Evans (2001:112)	True Positive	Half	True Positive	Half
GB117	Is there a copula for predicate nominals?	Proto-Oceanic	Absent	Lynch et al. (2011:86)	True Negative	Half	True Negative	Half
GB170	Can an adnominal property word agree with the noun in gender/noun class?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half
GB172	Can an article agree with the noun in gender/noun class?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB192	Is there a gender system where a noun's phonological properties are a factor in class assignment?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	False Positive
GB198	Can an adnominal numeral agree with the noun in gender/noun class?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half
GB431	Can adnominal possession be marked by a prefix on the possessed noun?	Proto-Oceanic	Absent	Pawley (1973:117), Ross (2004:494, 497)	True Negative	Half	True Negative	Half
GB432	Can adnominal possession be marked by a suffix on the possessor?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half
GB082	Is there overt morphological marking of present tense on verbs?	Proto-Oceanic	Absent	Lynch et al. (2011:84)	True Negative	Half	True Negative	Half
GB521	Can tense be marked by a non-inflecting word ("auxiliary particle")?	Proto-Oceanic	Absent	Lynch et al. (2011:84)	True Negative	Half	True Negative	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB264	Is there a polar interrogative particle that most commonly occurs neither clause-initially nor clause-finally?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	Half
GB023	Are there postnominal articles?	Proto-Oceanic	Absent	Pawley (1973:112), Ross (2004:498)	True Negative	Half	True Negative	Half
GB047	Is there a productive morphological pattern for deriving an action/state noun from a verb?	Proto-Oceanic	Present	Lynch et al. (2011:70)	True Positive	Half	True Positive	Half
GB052	Is there a gender/noun class system where shape is a factor in class assignment?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half
GB053	Is there a gender/noun class system where animacy is a factor in class assignment?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half
GB059	Is the adnominal possessive construction different for alienable and inalienable nouns?	Proto-Oceanic	Present	Ross (2004:492, 511-512), Lynch et al. (2011:69)	True Positive	Half	True Positive	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB079	Do verbs have prefixes/proclitics, other than those that only mark A, S or P (do include portmanteau: A and S + TAM)?	Proto-Oceanic	Present	Pawley (1973:142), Ross (2007:292), Lynch et al. (2011:83)	True Positive	Half	True Positive	Half
GB089	Can the S argument be indexed by a suffix/enclitic on the verb in the simple main clause?	Proto-Oceanic	Absent	Ross (2004:498-499), Lynch et al. (2011:83)	True Negative	Half	True Negative	Half
GB115	Is there a phonologically bound reciprocal marker on the verb?	Proto-Oceanic	Present	Pawley (1973:172), Ross (2004:495, 513)	True Positive	Half	True Positive	Half
GB155	Are causatives formed by affixes or clitics on verbs?	Proto-Oceanic	Present	Lynch et al. (2011:74, 83), Pawley (1973:130), Evans (2001:344)	True Positive	Half	True Positive	False Negative
GB171	Can an adnominal demonstrative agree with the noun in gender/noun class?	Proto-Oceanic	Absent	Ross (2004:498)	True Negative	Half	True Negative	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB193a	Is the order of the adnominal property and the noun ANM-N?	Proto-Oceanic	Absent	Ross (2004:497), Lynch et al. (2011:74)	True Negative	Half	True Negative	Half
GB285	Can polar interrogation be marked by a question particle and verbal morphology?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	Half
GB286	Can polar interrogation be indicated by overt verbal morphology only?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	Half
GB291	Can polar interrogation be marked by tone?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	Half
GB299	Can standard negation be marked by a non-inflecting word ("auxiliary particle")?	Proto-Oceanic	Present	Lynch et al. (2011:88)	True Positive	Half	True Positive	Half
GB083	Is there overt morphological marking on the verb dedicated to past tense?	Proto-Oceanic	Absent	Lynch et al. (2011:84)	True Negative	Half	True Negative	Half
GB084	Is there overt morphological marking on the verb dedicated to future tense?	Proto-Oceanic	Absent	Lynch et al. (2011:84)	True Negative	Half	True Negative	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB121	Can tense be marked by an inflecting word ("auxiliary verb")?	Proto-Oceanic	Absent	Lynch et al. (2011:84)	True Negative	Half	True Negative	Half
GB312	Is there overt morphological marking on the verb dedicated to mood?	Proto-Oceanic	Present	Lynch et al. (2011:84)	False Negative	Half	True Positive	True Positive
GB131	Is a pragmatically unmarked constituent order verb-initial for transitive clauses?	Proto-Central Pacific	Present	Kikusawa (2002:122)	False Negative	True Positive	False Negative	True Positive
GB522	Can the S or A argument be omitted from a pragmatically unmarked clause when the referent is inferrable from context ("pro-drop" or "null anaphora")?	Proto-Oceanic	Present	Ross (2004:495)	True Positive	Half	True Positive	Half
GB262	Is there a clause-initial polar interrogative particle?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	False Positive
GB132	Is a pragmatically unmarked constituent order verb-medial for transitive clauses?	Proto-Oceanic	Present	Ross (2004:497), Pawley (1973:117)	True Positive	Half	True Positive	False Negative

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB297	Can polar interrogation be indicated by a V-not-V construction?	Proto-Oceanic	Absent	Lynch et al. (2011:89)	True Negative	Half	True Negative	Half
GB065a	Is the order of possessor noun and possessed noun possessor-possessed?	Proto-Oceanic	Present	Lynch et al. (2011:76), Pawley (1973:155-156)	True Positive	Half	True Positive	Half
GB022	Are there prenominal articles?	Proto-Oceanic	Present	Crowley (1985), Lynch et al. (2011:70), Pawley (1973:112)	True Positive	Half	Half	False Negative
GB105	Can the recipient in a ditransitive construction be marked like the monotransitive patient?	Proto-Oceanic	Absent	Pawley (1973:118)	True Negative	Half	Half	Half
GB131	Is a pragmatically unmarked constituent order verb-initial for transitive clauses?	Proto-Oceanic	Present	Pawley (1973:118), Lynch et al. (2011:86)	False Negative	Half	False Negative	True Positive

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB140	Is verbal predication marked by the same negator as all of the following types of predication: locational, existential and nominal?	Proto-Eastern Polynesian	Absent	Clark (1976:130)	True Negative	Half	False Positive	Half
GB159	Are nouns reduplicated?	Proto-Oceanic	Present	Lynch et al. (2011:70)	Half	Half	True Positive	Half
GB130a	Is the order of S and V in intranstive clauses SV?	Proto-Central Pacific	Absent	Kikusawa (2002:122)	False Positive	Half	False Positive	True Negative
GB132	Is a pragmatically unmarked constituent order verb-medial for transitive clauses?	Proto-Polynesian	Absent	(Chung 1978:15), Pawley (1973:118)	False Positive	Half	False Positive	True Negative
GB132	Is a pragmatically unmarked constituent order verb-medial for transitive clauses?	Proto-Polynesian	Absent	Kikusawa (2002:122)	False Positive	Half	False Positive	True Negative
GB132	Is a pragmatically unmarked constituent order verb-medial for transitive clauses?	Proto-Central Pacific	Absent	Kikusawa (2002:122)	False Positive	False Positive	False Positive	True Negative

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB133	Is a pragmatically unmarked constituent order verb-final for transitive clauses?	Proto-Oceanic	Present	Pawley (1973:118)	False Negative	Half	False Negative	True Positive
GB068	Do core adjectives (defined semantically as property concepts such as value, shape, age, dimension) act like verbs in predicative position?	Proto-Oceanic	Present	Lynch et al. (2011:63, 74), Pawley (1973:113)	True Positive	Half	Half	Half
GB049	Is there a productive morphological pattern for deriving an object noun from a verb?	Proto-Oceanic	Present	Lynch et al. (2011:70)	True Positive	Half	Half	Half
GB130b	Is the order of S and V in intranstive clauses VS?	Proto-Central Pacific	Present	Kikusawa (2002:122)	Half	Half	False Negative	Half
GB086	Is a morphological distinction between perfective and imperfective aspect available on verbs?	Proto-Oceanic	Present	Lynch et al. (2011:84)	False Negative	Half	False Negative	Half

Grambank ID	Question	Proto-language	Historical linguistics finding	Reference	Glottolog-tree		Gray et al 2009 tree	
					Parsimony result	ML result	Parsimony result	ML result
GB140	Is verbal predication marked by the same negator as all of the following types of predication: locational, existential and nominal?	Proto-Oceanic	Absent	Pawley (1973:143-6), Lynch et al. (2011:88)	Half	Half	False Positive	Half

H Tables of feature stability

Grambank ID	Parsimony cost	0	1	Languages
GB110 VSupplTA	0	169	0	169
GB149 InverseBoundV	0	191	0	191
GB315 DIMGender	0	184	0	184
GB336 NUMBodyTally	0	188	0	188
GB167 PROLogophore	1	156	1	157
GB192 GenderPhono	1	199	1	200
GB196 PRO2PMascFem	1	210	1	211
GB197 PRO1PMascFem	1	210	1	211
GB260 QWordOrder	1	161	1	162
GB285 QPartVMorph	1	165	1	166
GB303 AntipassiveFree	1	131	1	132
GB314 AUGgender	1	185	1	186
GB402 VSupplSee	1	134	1	135

Table 18: Most stable features, Parsimony Glottolog-tree (including all with a cost of 1 or less).

Grambank ID	Parsimony cost	0	1	Languages
GB065a OrderPosrPosd	51	100	119	219
GB069 AttrAdjLikeV	51	114	64	178
GB318 PluralFree	52	91	108	199
GB324 QV	52	114	68	182
GB408 Accusative	52	113	84	197

Table 19: Most unstable features, Parsimony Glottolog-tree.

Grambank ID	Parsimony cost	0	1	Languages
GB133 TransVFinalOrder	7	205	13	218
GB422 ComplThinkKnowPost	8	151	9	160
GB188 AUGBound	9	105	10	115
GB264 QPartMedial	9	156	10	166
GB330 RELCorr	9	153	9	162

Table 20: Most stable features, Parsimony Glottolog-tree, only features with at least a distribution of 95%/ 5%.

Grambank ID	Parsimony cost	0	1	Languages
GB110 VSupplTA	0	84	0	84
GB149 InverseBoundV	0	100	0	100
GB167 PROLogophore	0	85	0	85
GB196 PRO2PMascFem	0	108	0	108
GB197 PRO1PMascFem	0	107	0	107
GB260 QWordOrder	0	85	0	85
GB315 DIMGender	0	96	0	96
GB336 NUMBodyTally	0	95	0	95
GB402 VSupplSee	0	70	0	70
GB026 ADJDiscont	1	100	1	101
GB028 Clusivity	1	1	111	112
GB051 GenderSex	1	104	1	105
GB170 ADJGender	1	102	1	103
GB171 DEMGender	1	101	1	102
GB192 GenderPhono	1	104	1	105
GB285 QPartVMorph	1	85	1	86
GB303 AntipassiveFree	1	66	1	67
GB314 AUGgender	1	97	1	98
GB321 GenderUnpredict	1	98	1	99
GB327 RELPost	1	1	96	97

Table 21: Most stable features, Parsimony Gray et al 2009-tree (including all with a cost of 1 or less).

Grambank ID	Parsimony cost	0	1	Languages
GB065a OrderPosrPosd	27	53	59	112
GB069 AttrAdjLikeV	27	57	36	93
GB324 QV	27	58	36	94
GB313 PROPoss	31	53	46	99
GB138 NEGInitial	32	67	41	108

Table 22: Most unstable features, Parsimony Gray et al 2009-tree.

Grambank ID	Parsimony cost	0	1	Languages
GB300 VSupplGive	3	67	4	71
GB188 AUGBound	4	60	4	64
GB422 ComplThinkKnowPost	4	77	5	82
GB081 VInfix	5	95	9	104
GB095 CaseSplitTAM	5	96	6	102

Table 23: Most stable features, Parsimony Gray et al 2009-tree, only features with at least a distribution of 95% / 5%.

Grambank ID	Average rate	gain	loss	0	1	Languages
GB110 VSupplTA	0	0	0	169	0	169
GB149 InverseBoundV	0	0	0	191	0	191
GB315 DIMGender	0	0	0	184	0	184
GB336 NUMBodyTally	0	0	0	188	0	188
GB327 RELPost	0	0	0	2	180	182

Table 24: Most stable features, Maximum Likelihood Glottolog-tree.

Grambank ID	Average rate	gain	loss	0	1	Languages
GB433 POSSSfxPosd	6	3	9	46	162	208
GB131 TransVInitOrder	7	12	3	164	53	217
GB081 VInfix	8	14	2	188	17	205
GB132 TransVMedOrder	9	4	14	58	158	216
GB024b OrderNNUM	9	6	12	72	138	210

Table 25: Most stable features, Maximum Likelihood Glottolog-tree. Only features with at least a distribution of 95% 5%.

Grambank ID	Average rate	gain	loss	0	1	Languages
GB049 NMZObject	82	64	100	61	101	162
GB160 RedupOther	83	76	89	67	80	147
GB105 CaseRecipientObj	86	100	72	98	71	169
GB122 VCompounding	93	96	91	71	70	141
GB140 SameNEGLocExistNom	97	100	94	81	77	158

Table 26: Most unstable features, Maximum Likelihood Glottolog-tree.

Grambank ID	Average rate	gain	loss	0	1	Languages
GB026 ADJDiscont	0	0	0	103	1	104
GB028 Clusivity	0	0	0	1	115	116
GB110 VSupplTA	0	0	0	87	0	87
GB149 InverseBoundV	0	0	0	103	0	103
GB167 PROLogophore	0	0	0	87	0	87

Table 27: Most stable features, Maximum Likelihood Gray et al 2009-tree.

Grambank ID	Average rate	gain	loss	0	1	Languages
GB081 VInfix	0	0	0	98	9	107
GB091 A-ArgSfxV	0	0	0	103	8	111
GB333 NUMDecimal	0	0	0	16	91	107
GB148 AntipassiveBoundV	0	0	0	68	13	81
GB335 NUMVigesimal	0	0	0	90	13	103

Table 28: Most stable features, Maximum Likelihood Gray et al 2009-tree. Only features with at least a distribution of 95% 5%.

Grambank ID	Average rate	gain	loss	0	1	Languages
GB313 PROPoss	94	100	88	55	47	102
GB065a OrderPosrPosd	97	94	100	56	59	115
GB160 RedupOther	99	97	100	37	38	75
GB140 SameNEGLocExistNom	99	100	97	42	38	80
GB105 CaseRecipientObj	99	100	98	45	41	86

Table 29: Most unstable features, Maximum Likelihood Gray et al 2009-tree.