



Project
MUSE[®]

Today's Research. Tomorrow's Inspiration.

STRUCTURAL PHYLOGENY IN HISTORICAL LINGUISTICS:
METHODOLOGICAL EXPLORATIONS APPLIED IN ISLAND MELANESIA

MICHAEL DUNN
*Max Planck Institute
for Psycholinguistics and
Radboud University
Nijmegen*

STEPHEN C. LEVINSON
*Max Planck Institute
for Psycholinguistics*

EVA LINDSTRÖM
Stockholm University

GER REESINK

Radboud University Nijmegen

ANGELA TERRILL

Radboud University Nijmegen

Using various methods derived from evolutionary biology, including maximum parsimony and Bayesian phylogenetic analysis, we tackle the question of the relationships among a group of Papuan isolate languages that have hitherto resisted accepted attempts at demonstration of interrelatedness. Instead of using existing vocabulary-based methods, which cannot be applied to these languages due to the paucity of shared lexemes, we created a database of STRUCTURAL FEATURES—abstract phonological and grammatical features apart from their form. The methods are first tested on the closely related Oceanic languages spoken in the same region as the Papuan languages in question. We find that using biological methods on structural features can recapitulate the results of the comparative method tree for the Oceanic languages, thus showing that structural features can be a valid way of extracting linguistic history. Application of the same methods to the otherwise unrelatable Papuan languages is therefore likely to be similarly valid. Because languages that have been in contact for protracted periods may also converge, we outline additional methods for distinguishing convergence from inherited relatedness.*

1. INTRODUCTION. Nonlexical evidence for language relationships is a major blind spot in historical linguistics. Traditional methods—based on the search for cognates in vocabulary constrained by such principles as the regularity of sound change—are powerful, but the lexical signal decays, and in even the largest language families there seems to be a maximum temporal horizon of up to about ten thousand years, beyond which lexical evidence of relatedness is not recoverable (Nichols 1992). But by ten thousand years ago, the peopling of the world was more or less complete. In smaller families the temporal horizon is likely to be much closer to the present. If linguistics is to contribute to the rapidly developing picture of human prehistory emerging from human genetics and archaeology, we need to extract the maximum historical information from the data available, especially in cases where lexical evidence is not informative.

A brief *Science* paper (Dunn et al. 2005) outlined the possibilities of using computational phylogenetic methods applied purely to structural properties of languages, as opposed to lexical items, to extract likely patterns of ancient relatedness. In the current

* This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NL), Vetenskapsrådet (SE), and the EC Sixth Framework Programme under contract no. ERAS-CT-2003-980409. Additional fieldwork data used in this study (i.e. apart from that collected by the authors) were provided by Stuart Robinson (Rotokas), Tonya Stebbins (Mali), William Thurston (Anêm), and Claudia Wegener (Savosavo). Assistance with coding of Oceanic languages from published sources was provided by S. Nordhoff, V. Rodrigues, and K. Ahlén. For permission to use unpublished materials we thank K. Hashimoto (Ata), Stellan Lindrud (Kol), Lloyd Milligan (Mangse), and Dan Rath (Mengen). We thank Michael Cysouw, Nick Evans, Robert Foley, Jonathan Friedlaender, Françoise Friedlaender, Russell Gray, Simon Greenhill, Brian Joseph, Marta Lahr, Gunter Senft, and four anonymous referees for discussion.

article we set out to explain those methods in more detail, show how they can be extended and refined, and push the analysis further to explore how a phylogenetic signal can be distinguished from relatedness through propinquity and possible contact. The whole approach here, though similar to that in McMahon & McMahon 2005, for example, has a number of special advantages: first, in not depending on vocabulary matches, it promises to extend the range of historical linguistics further back in time, and thus suggest deep-time relations between independent well-established language families as well as connections between known families and languages currently considered isolates. Second, it promises to connect linguistic typology and historical linguistics—two fields that have pursued independent paths, even though typological patterns are bound to have at least a partially historical explanation.

The first part of this article concerns methodological preliminaries: we motivate the use of abstract structural features for historical investigations, differentiating our own approach from a number of other recent applications of similar tools, and explaining some of the basic concepts of computational cladistics. We further present a nonmathematical description of the workings of the two main phylogenetic methods to be used, maximum parsimony and Bayesian phylogenetic analysis. Following this, we discuss the languages and linguistic features used in the analysis. The geographic focus of this study, Island Melanesia, contains languages from two groups: the Oceanic languages, which are known to be closely related, and the so-called Papuan languages, which are a residual category of languages whose relationships to one another are far from clear. We go on to describe the database of structural features we employ in the phylogenetic studies.

Next, we present phylogenetic analyses of the two language groups. First, we use the Oceanic languages to test the method, and show how phylogenetic reconstruction based entirely on abstract structural features can recreate the tree independently reconstructed using vocabulary-based methods (largely, the classic comparative method). Then we go on to apply the method to the Papuan languages and show how the results provide a plausible reconstruction of relatedness between languages that cannot be related to one another using vocabulary-based methods. We also turn our attention to what specific grammatical features have made the most contribution to the results, as a way of gauging the relative typological stability of certain linguistic features. Finally, we consider how contact-induced convergence may be distinguishable from inheritance of features from a shared ancestor, again using the Oceanic and Papuan language groups as the basic data points. Instead of taking individual languages as the points of comparison we consider the contributions of linguistic features to the emergent language histories.

2. METHODOLOGICAL PRELIMINARIES. This section describes the well-established comparative method of historical linguistics, with its strengths and its weaknesses. Of particular relevance here is its dependence on shared linguistic form, especially vocabulary, which makes it impossible to apply to languages separated so long ago that any surface traces of cognacy have been eroded. We go on to discuss the use of abstract structural features, that is, the presence or absence of particular categories rather than form, to assess deep-time linguistic relations. To evaluate the information contained in a database of such features, computational methods are required, and we touch upon previous work in this vein. We then describe the two principal methods for finding treelike structures in the data, viz. maximum parsimony and maximum likelihood (MCMC Bayesian phylogenetic analysis; see §2.5), followed by an explication of some basic concepts in the field, and of methods to investigate non-treelike signals in the data.

2.1. THE LIMITATIONS OF SOUND-MEANING CORRESPONDENCES IN THE SEARCH FOR LANGUAGE PHYLOGENY. Nearly all the great storehouse of knowledge about language relatedness accumulated over the last two and a half centuries has been based on sound-meaning correspondences in vocabulary. The method employed is simply known as the COMPARATIVE METHOD (CM), characterized by Harrison (2003:213) as ‘the sine qua non of linguistic prehistory’. The CM can be summarized as a set of instructions (Durie & Ross 1996): (i) Once a preliminary diagnosis of possible language families has been made, a more definitive assessment of genealogical relationships would need to (ii) demonstrate cognate sets (both morphological paradigms and lexical items), (iii) establish regular sound correspondences, and (iv) reconstruct the proto-language of the family, with its proto-phonology and morphemes. On this basis, (v) innovations shared by groups within the family can be tabulated in order to (vi) arrive at an internal classification, a family tree. A final stage (vii) would involve constructing an etymological dictionary, tracing borrowings, semantic changes, and so on, allowing insights into ancestral activities, ecologies, and preoccupations, as reflected in vocabulary fields.

The first stage simply involves a recognition of similarities in lexical and morphological material between two or more languages, but since similarities can arise due to contact or simple chance (Campbell 1998:318–22), it is necessary to proceed with stages (ii) and (iii). Based on the observation that sound changes largely take place regularly throughout the lexicon of a language, true cognate forms are identified and differentiated from other similarities (McMahon & McMahon 2005:8).

Starting from any two languages presumed to be related, a set of reconstructed proto-forms for the immediate ancestor language is built. Once proto-forms for ancestors of two separate CLADES (branches) have been established, the two sets of proto-forms can be compared, and the sound changes that would have been required to separate them considered, after which proto-forms for the immediate ancestor of these two clades can be posited. At this stage cognates can be found that superficially seem quite unrelated, but that nevertheless can be systematically shown to be related by sequences of sound shifts over time (such as Hindi *cakkā* ‘wheel’ and English *wheel*, true cognates derived from Proto-Indo-European **k^wek^wlo-* ‘wheel’ (Hock & Joseph 1996:469)). Working ever backwards, not only a family tree but also a set of proto-forms is thus established. The comparative method produces a phylogenetic hypothesis by explicit methods, at the same time attributing specific linguistic forms to each now-vanished ancestral node in the tree. In language families with long traditions of literacy, such as Indo-European, Dravidian, or Chinese, it is possible to check the inferences directly, at least to some extent.

Nothing so far has the potential to replace the comparative method as the gold standard for historical linguistics. Nevertheless, it has distinct limitations. First, there are limitations concerning the linguistic domain. Not all linguistic material is suitable. It is well known that nonarbitrary forms like onomatopoeic (*blow*, *sneeze*) and nursery forms (*mama*, *papa*) should not count as cognates. Syntax is also notoriously difficult to reconstruct (Antilla 1972:355ff., McMahon & McMahon 2005:15). We return to this matter in §2.2 below.

A second question is how much we can really reconstruct. As no language is free of dialectal variation, the reconstruction of just one proto-form is necessarily an abstraction, albeit an abstraction that is shared with any grammatical description: the full range of variation is rarely recorded. A second part of this problem is that the actual phonetic values are also not reconstructed—although as the CM deals in phonemic contrasts this problem is not serious.

Third, we know that languages can borrow vocabulary; indeed this is by far the most common effect of language contact. Intense contact can even lead to language meltdown as it were, with extensive sharing of vocabulary. Harrison (2003:231–32) details cases in the Oceanic family where ‘[w]e “know” the languages are related but can’t demonstrate that they are by using the logic of the comparative method’. As Campbell puts it soberly, ‘[t]he problem of loans, or potential loans, is very serious’ (Campbell 2003: 271). For example, it is controversial to this day whether Quechua and Aymara are phylogenetically related or merely share large amounts of vocabulary through extensive contact (Adelaar & Muysken 2004, Campbell 1995, McMahon & McMahon 2005).

A fourth limitation of the comparative method is that it has limited time depth. This is not an issue of time itself, nor of the method as such, of course, but a practical and statistical consequence of the erosion of both sounds and meanings over time, compounded by the loss and replacement of vocabulary. Although a theoretical ceiling is impossible to establish, in practice it appears that there is no data to support reconstruction beyond ten thousand years (Nichols 1992), a date very roughly agreeing with some recent results by Gray and Atkinson (2003).¹ For prehistorians wishing to connect living peoples to archaeological traces or migrations, or geneticists wishing to correlate human population biology with linguistic clades, this limit is a serious drawback.

A fifth limitation is that the comparative method deals only with phylogeny—all obvious loans are weeded out and thereby excluded from the domain of enquiry, leaving only information pertaining to constructing a phylogenetic tree. Information on lateral transfer can be included in a later stage, but it does not form part of the process of comparative reconstruction. In fact, of course, languages, whether related or unrelated, are rarely out of contact with one another, and they always hybridize to some extent (Dixon 1997). Such contact phenomena can make the reconstruction of a family tree highly problematic, even in such well-studied families as Indo-European (see e.g. Ringe and colleagues’ (2002) treatment of Germanic, where conflicting signals make a ‘clean tree’ impossible). As we see below, Oceanic is another well-studied family where such hybridization (there known as ‘linkages’) has occurred within the family. As McMahon and McMahon (2005:27) state, ‘although we might be able to group languages into a family with a certain measure of security, using the comparative method for instance, subgrouping is still a matter of considerable unclarity, since the method as it stands does not allow for the quantification of degrees of relatedness’. Actually, rough quantification is possible, for example, by counting the number of sound changes separating two related languages, counting the number of inherited words shared by daughter languages, and so on. But rigorous quantification remains elusive. In response to this problem, McMahon and McMahon argue for supplementing (not replacing) the CM with computational methods. Recent developments in evolutionary biological methods, allowing the representation of relationships as networks, make the estimation of such hybridization more objective, and we utilize these techniques below.

It is worth drawing attention to the fact that all vocabulary-based methods are ultimately based on underlying statistical reasoning (for the CM this is nicely brought out

¹ Assuming a 20 percent loss of cognates per millennium through lexical replacements, 6,000 years of replacement will leave only 7 percent shared putative cognates, which is perhaps the lowest percentage safely distinguishable from chance matches (Nichols 1998). On the unreliability of any such assumptions of constant loss, see for example Blust 2000, which emphasizes the special rate of loss in the Oceanic subbranch of Austronesian.

by Harrison (2003) and McMahon and McMahon (2005)). The assumption is that form-meaning correspondences are, from a systems point of view, arbitrary (as emphasized by Saussure (1916)), and thus could equally be otherwise. A reasonable number of detailed correspondences make chance correspondences, or the identification of false cognates, highly unlikely. Nevertheless, the statistical assumptions are rarely tested in traditional work (but see Kessler 2001). Hence other methods, which make more explicit the statistical assumptions, should be considered favorably in comparison. Computational approaches are not meant to replace the CM, and the trees or networks they produce are not different in shape, but they are different in their ability to precisely quantify the degree of statistical robustness and the level of confidence in an analysis (McMahon & McMahon 2005:48).

A number of recent computational studies, to be reviewed in §2.3, have been able to deal successfully with some of the limitations, for instance by taking into account multiple meanings for single etyma or multiple etyma for single meanings, different rates of change for subgroups of lexical items, and factoring out borrowings. Since such methods apply statistical methods AFTER steps (i)–(iv) of the comparative method, they supply quantified degrees of relatedness and give plausible subgroupings, including inevitable reticulations. But none of these studies are able to overcome the limited time depth inherent to the mutation rate of lexical items.

2.2. STRUCTURAL FEATURES AS HISTORICAL MARKERS. In this section, we ask two basic questions: (i) Can structural features, like lexical features, carry a detectable historical signal?, and (ii) If so, is it possible that structural features can have equal or greater overall time-stability than lexical features? No one doubts that the CM can be as directly applied to grammatical morphemes as to lexemes:

Much of what is called grammatical reconstruction in the literature is just plain vanilla comparative method applied to morphemes in the usual way. The main difference is that the morphemes have glosses like 'to', 'present' rather than 'sun', 'wind' and 'fire'. (Harrison 2003:228)

In their excellent summary, Harris and Campbell (1995) defend the application of the comparative method to syntactic reconstruction. They point out that syntactic patterns—for example, alignment patterns—can be inferred from sentential tokens and can be placed in exact correspondence (e.g. comparing the case marking of specific types of subjects). Moreover, the much repeated claim that there are no regularities in syntactic change is simply false—for example, just as the change /p/ to /f/ is ubiquitous while the reverse is infrequent, the transition of postpositions to case suffixes is commonplace. Case studies demonstrate what can be done: Campbell (1998:250) gives an impressive list of the reconstructable properties of Proto-Uralic syntax. In short, there is no reason for believing that syntactic reconstruction is not a viable goal. In fact, in many cases historical linguists make such reconstructions *en passant*, and we refer below to work on the Oceanic languages in particular (and see Lynch et al. 2002).

While these cases of acceptable reconstruction rely on formal, or substantive (Croft 2004), features as well, in general the use of just morphotactic or syntactic (i.e. typological) similarities is firmly rejected for the establishment of genealogical relationships (inter alia Campbell 1998:323, Croft 2004), unless the lack of morpheme cognacy can be explained (Rankin 2003:197).

We argue that structural features (abstract grammatical properties) can be used to investigate historical relations between languages. Thomason and Kaufman (1988) have

shown that just about any structural or grammatical feature can in fact be transferred from one language to another (see also Curnow 2001), but it is essential here to distinguish probabilities from possibilities; outside special conditions, there will be no borrowing of grammatical properties without prior lexical borrowing (Moravcsik 1978).

There are, however, those special sociolinguistic conditions where these generalizations do not hold: these are essentially cases where the donating language is adopted wholesale by the speakers of another language, in the classic case in full language shift. In this case, the tendency is for substrate influences to be more apparent in structure (phonology, grammar) than in lexicon (Aikhenvald 1996, Thomason 2001, Thomason & Kaufman 1988), due to imperfect learning or interference. Ross (1996, 1999, 2001b) has argued that METATYPY is another kind of contact-induced change that does not necessarily entail lexical correspondences. This process causes the morphosyntactic organization of different languages to become similar when bilingual speakers model the organization of one language on another, as illustrated by Oceanic Takia and Maisin having undergone restructuring on the models of neighboring Papuan languages.

Again these cases require special sociolinguistic conditions and seem to be relatively rare, but they do form a special hazard for syntactic reconstruction, in that the structural features that are due to interference could be mistakenly thought to be inherited from the ancestor of the adopted language.

We hold that the combination of structural features from different domains of a grammar (phonology, morphology, syntax, semantics) can indeed yield distinguishable profiles that allow us to investigate historical relations between languages, whether such relations arise from descent or contact.

Let us review the reasons Harrison gave for rejecting the comparison of grammatical properties. He denies the possibility of using abstract grammatical properties in order to infer genealogical relationships between languages for two reasons: (i) syntactic patterns lack the arbitrariness of sound-meaning pairings in lexical items and functional morphemes, and (ii) there can be no regularity in syntactic change, such as the CM can establish for sound changes in successive stages of language varieties. He considers it axiomatic that 'individual simplex linguistic signs' reside in the lexicon, but that there is no 'grammaticon' for complex linguistic signs. 'Any system of grammatical contrasts is iconic to the extent that it reflects a distinctly human ontology' (Harrison 2003:224). He illustrates his point by comparing two closely related Micronesian languages. Ponapean has a postverbal affix indexing person and number of the direct object, similar to Hebrew, whereas Mokilese transitive verbs are invariant, as in English. While it is easy to concur with his conclusion that this does not signal different genealogical relationships for Ponapean and Mokilese, we would argue that the (dis)agreement between two languages with regard to a particular configuration for a number of syntagmatic constructions is likely to be significant. As Watkins (2001:62) points out, 'the language areas involving Indo-European languages have all been characterized by interdiffusion of grammatical features, but in none can we really speak of convergence to a common prototype, in the sense of loss of linguistic identity'. Watkins continues: 'I do not deny that this is possible, but it remains for me only a theoretical construct'. He concludes that both genetic and typological comparison are necessary in order to draw historical conclusions.

The issue, then, is how many and what types of structural features are needed to allow inferences regarding linguistic relatedness. The answer is an empirical matter. In order not to predetermine the nature and/or number of structural features that would

yield a level of significance that allows what Nichols (1996:48) calls *INDIVIDUAL-IDENTIFYING* evidence, it is best to follow an inductive method. As many abstract structural features from as many parts of the grammar as possible should be investigated. As we show below, there are computational methods to help determine what kind and size of constellations of grammatical properties yield individual-identifying evidence.

We come now to our second question: Do structural features have equal or greater overall time-stability than lexical features? There is little doubt that abstract grammatical patterns can remain stable for millennia. Let us take the issue of word order, one of the most intensively researched areas of typology and historical linguistics. Languages with a long continuous written history demonstrate that long-term stability of word order is certainly possible. The Dravidian languages are synchronically left-branching, canonical subject-object-verb (SOV) languages with harmonic orders of genitive-noun, postpositions, verb-auxiliary, and so forth (Steever 1998). And they have been so for over two thousand years, as far as written records attest. Similarly, 'the order of elements in a Chinese sentence has remained remarkably stable over the last two millennia' (Norman 1988:130). In Indo-European, of course, the fortunes of original SOV order have been much more diverse in the same time frame: Latin's OV and adjective-noun developed into French VO with noun-adjective, an almost complete reversal of Latin's (slightly inconsistent) dependent-head pattern (Harris & Campbell 1995:230). It is clear that some word orders, specifically SOV and SVO, are more stable than others (Nichols 2003:304–5). The Greenbergian word-order harmonies establish statistical tendencies for harmonic word-order characteristics to bundle together within a language (Hawkins 1983:133ff.), making these covariant characters, or features, from a cladistic viewpoint, a matter that is significant for some forms of phylogenetic analysis (see below). But the point here is that language families with long written histories make these issues of statistical time-stability something that can be empirically investigated.

In language families without such documents, it is perfectly feasible to look at the extant languages, and project back from current word orders to the likely ancestral values (Watkins 1976). Although word-order change can be due to internal factors, it often seems connected to intense language contact between unrelated languages (Harris & Campbell 1995:137–41), as has been suggested specifically for the region of interest here and the interactions between Papuan and Austronesian languages (Foley 1986:281–82).

We have paid more attention to word order because it is sometimes held to be highly labile (Matthews 1982), but clearly other features such as gender, case, and specific grammatical categories (such as an inclusive/exclusive distinction in first-person pronouns) are less controversially time-stable.

One scholar who has championed the use of structural features in language prehistory is Nichols (1992, 1998, 2003). We concentrate here on two points she makes: first, their relative time-stability, and second, their use as diagnostics for prehistoric relations between languages. Nichols identifies 'historical markers', namely specific structural features that can be shown to be persistent (time-stable) inside language families, have low world-wide frequency, a low tendency to be borrowed, and are not given to spontaneous emergence (Nichols 1998:143–45). 'Markers' used by Nichols include ergativity, headmarking, numeral classifiers, identity of stems across singular and plural pronouns of the same person, inclusive/exclusive first-person pronouns, verb-initial word order, nominal classes evident only under possession, and gender and concord classes. The geographic distribution of these features is indeed suggestive of migration

routes, allowing the development of detailed historical hypotheses of language spread and diversification across the globe (Nichols 1992, 1998). Since some of these hypothesized migration routes are vast, if she is right, the time scales involved are also huge, taking us well back in time into the Pleistocene, where it is presumed the comparative method cannot reach.

Nichols makes clear that she does not assume that possession of the same ‘markers’ can necessarily be taken as evidence for inheritance from a common ancestor, for individual markers may have been acquired through language contact (1998:148)—rather, sharing of markers is taken to be indicative of shared geographic origin (e.g. contact along a migration route). Indeed, even abstract patterns of alignment expressed by noncognate verbal marking of arguments have been shown to be susceptible to diffusion in intense contact situations, as Mithun (2007, 2009) argues for North American languages. This finding once again illustrates that ‘anything can be borrowed’: according to Mithun, perhaps not just the grammatical patterns themselves, but also the rhetorical precursors to them, by the process dubbed *metatypy* by Ross (1996). Nevertheless, we think it follows from the studies of language contact cited above that in the majority of cases a sufficiently large cluster of markers can carry a significant historical signal (allowing that what constitutes a ‘sufficiently large cluster’ must be established; an empirical test is reported in §4.1). This signal can contain some phylogenetic information since genuinely related languages, uncontroversially established by the CM, do share typological features to a high degree as well.

Nichols uses these ‘markers’ to arrive at broad conclusions about linguistic prehistory, which have not, however, been universally well received. But the only features of her approach we need to defend are (i) the potential time-stability of grammatical features, and (ii) their possible use as diagnostics for prehistoric relations between languages. This is because our approach differs in a crucial respect. Rather than use a small number of typological features, preselected on the grounds that they make good ‘markers’ on the grounds of time-stability and rarity of independent invention, we instead use over a hundred structural properties of languages that together yield overall typological profiles of the languages under comparison. We have included features that certainly have ‘marker’ characteristics (including some of those used by Nichols) and others that may not. By using many features we greatly decrease the probabilities of chance cooccurrence, so that the shared clusters of features suggest a shared historical association. Like Nichols, we can be agnostic about inheritance vs. diffusion in any particular case, but the signal will at least distinguish order from randomness. The order we obtain may be due to either phylogeny or contact or both. We show (§5) that recent contact can be distinguished from ancient relationships that may be due to either phylogeny or contact or both. By investigating a subset of the features used we can investigate their stability over time at least in an established family.

2.3. PRIOR USE IN LINGUISTICS OF METHODS DERIVED FROM EVOLUTIONARY BIOLOGY. In recent years, a number of studies have been published in which computational cladistic methods have been brought to bear on linguistic data (Gray & Atkinson 2003, Gray & Jordan 2000, Holden 2001, McMahon & McMahon 2003, Minett & Wang 2003, Nakhleh et al. 2005, Rexová et al. 2003, Ringe et al. 2002, Warnow 1997, Warnow et al. 1995). This section briefly summarizes these earlier ventures and clarifies where the present work differs from them despite employing some of the same tools. In this tradition, a property with two or more alternative values is called a *CHARACTER*, and the values are called *STATES*, a terminology we here adopt.

These studies share a number of properties. First, they have all been applied to well-known and well-studied linguistic families, mostly Indo-European (Gray & Atkinson 2003, McMahon & McMahon 2003, Rexová et al. 2003, Ringe et al. 2002), but also Austronesian (Gray & Jordan 2000), Bantu (Holden 2001), and Chinese (Minett & Wang 2003). Consequently, the structure of the family trees in question has already been extensively explored using the comparative method.

Second, all use word lists as their principal, or in most cases only, type of data. Usually, a Swadesh list of 100 or 200 core vocabulary items is compiled for a set of languages already known to be related. Each meaning on the list, for example 'hand', is treated as a character, and the forms in each language are assigned to states according to the cognate sets they belong to, such that French *main* and Spanish *mano* would receive the same value (state) and English *hand* and German *Hand* a different one. Cladistic algorithms are then applied to produce tree structures expressing the relationships between the languages.

Ringe and colleagues (2002) have included more lexical characters (giving a total of 333) and have also added twenty-two phonological and fifteen morphological characters. Given the greater chance of lexical borrowing noted in the previous section, the inclusion of phonology and morphology is certainly an advantage over just word lists in trying to capture the relations among the languages, but it should be noted that the formulation of the characters still relies heavily on the fact that the researchers already have detailed knowledge of the development of the language family through extensive prior work within the comparative method. They thus use very specific phonological changes (e.g. 'medial **k*^w > **g*^w unless **s* follows immediately'), and many of the morphological characters are also parts of the lexicon (e.g. 'abstract noun suffix **-ti*-') rather than abstract categories or combinations of categories. In other words, the added characters of Ringe et al. 2002 incorporate a great deal of specialist comparative linguistic knowledge into the cladistic method; this naturally requires even more prior knowledge than word lists, and effectively restricts this methodology to very well-understood language families.

Modeling evolution as the gain and loss of reflexes of cognate sets, as for example the analysis of Austronesian by Gray and Atkinson (2003) and Indo-European by Ringe and colleagues (2002), does not just yield phylogenetic trees similar to those we already know from the comparative method, but has the additional advantage of providing a measure of statistical robustness (McMahon & McMahon 2005:48). The comparative method, which models evolution as the ordered accumulation of linguistic changes, tolerates little ambiguity. Every identification of an innovation in the sound system of the language is treated as having a probability of 1.0, and any inconsistency must be dealt with by excluding the inconsistent data (e.g. by identifying the source of the conflict as contact-induced change). Perhaps surprisingly, it turns out that a 'perfect-phylogeny' tree can be quite difficult to obtain using computational methods even in the case of well-known families. Working with twenty-four Indo-European languages, Ringe and colleagues (2002) found it impossible to produce a perfect phylogeny tree (see §5 for more discussion) even from a set of characters chosen to exclude the possibility of 'back-mutation' and had to develop a principled method for excluding incompatible characters (which presumably resulted from undetected borrowing or other nonphylogenetic processes).

Several papers address method specifically. For example, McMahon and McMahon (2003) discuss the choice of the most retentive characters (still words in lists), showing

quite different results depending on the character set used. Minett and Wang (2003) test two methods of identifying areas of borrowing by mathematical means. Warnow and colleagues (2005) investigate properties of character evolution and parallel development, and both Warnow and colleagues (2005) and Nakhleh and colleagues (2005) work to develop methods that can handle tree structures and network structures in a single model, building on the work and data of Ringe et al. 2002; note, however, that they both rely on the linguist to identify the borrowing at the data-coding stage. Only one paper, Warnow 1997, discusses the analysis of languages that are not known to be related, with reference to Johanna Nichols's work (1990, 1992) but no formalization of a method.

The present study starts from very different premises, and therefore the application of cladistic methods follows a different path. Here, the object of study is a group of languages that are not known to be related—in fact, the genetic relationships among them are very much at issue.

2.4. MAXIMUM PARSIMONY. In evolutionary biological methods, and similarly in historical linguistics, the essential problem in phylogenetics is to choose the tree that best fits the data. In traditional studies in linguistics, this has been done by hand, art, experience, and intuition. But as the number of data points and the number of taxa increase, such traditional methods do not suffice. For twelve taxa (in our case, languages), for example, there are over 13,000 million possible trees—different possible branching arrangements (Felsenstein 2004:23), and these not only cannot be inspected by hand, they cannot be practically enumerated by machine either! Hence the need for powerful computational algorithms that will find the best or most likely tree by various heuristics. In this section and the following, we sketch the two algorithms—maximum parsimony and a (Bayesian) maximum likelihood method—that we employ below (§4).

We first describe maximum parsimony, as a well-established method with thoroughly explored strengths and weaknesses (its origins can be found in Edwards & Cavalli-Sforza 1963). Maximum parsimony is a measure that seeks to minimize the amount of evolutionary change in a tree: the basic rule is 'minimize independent evolutions of the same feature or character state'. This is achieved by locating each change in character state at the highest possible node in the tree, so that the least number of changes account for all the attested states. Some versions of parsimony are constrained to allow only unidirectional changes. Unidirectional parsimony may be used where linguistic characters represent irreversible changes, such as the mergers used by Ringe and colleagues (2002). Other forms of parsimony allow bidirectional changes and are appropriate for different kinds of data (e.g. abstract structural features, discussed below). A number of different algorithms (both exhaustive and heuristic) exist for determining which tree is the most parsimonious for the observed character states of a set of taxa. As the number of possible trees increases rapidly with an increasing number of taxa, some kind of heuristic algorithm is likely to be necessary for a reasonably large set of taxa. With very small data sets (and not more than four or five taxa, with fifteen and 105 possible trees respectively), a maximum-parsimony analysis can be done by hand.

Parsimony is simple to apply in its heuristic forms (see Appendix D for details of the software used in this study), and ad hoc adjustments to the analysis are easier to implement. For example, if it is independently known that some characters of the taxa under consideration have more phylogenetic significance than others, it is relatively easy to add this information in the form of character weighting to a parsimony analysis (character weighting is actually presumed in the Bayesian phylogenetic analyses discussed below, §2.6).

The major weakness of parsimony is the phenomenon of ‘long branch attraction’ (Felsenstein 2004). If two distantly related taxa are both highly divergent, then the most parsimonious account of their history is to infer that they are both derived from a highly divergent common ancestor. The two taxa are reconciled by having them converge with each other first, before converging with the rest of the tree. This is a particular problem in cases such as those we face in our own analysis, where the characters under analysis have limited state possibilities and/or unequal rates of change. Parsimony also makes assumptions about rate of change that may be inappropriate for linguistic data. It should be noted that Tuffley and Steel (1997) have shown that maximum parsimony is equivalent to a special case of maximum likelihood, the ‘no common mechanisms’ model, which presumes that rates of change of characters cannot be classified into rate classes. This model is probably not appropriate for typological data, since it leads to the counterintuitive prediction that it should be impossible to talk about innovative and conservative linguistic features.

Interpreting a parsimony analysis is not always straightforward. A useful statistical test of the strength of the phylogenetic signal in the data is called BOOTSTRAP RESAMPLING. It is always possible that a small number of ‘badly behaved’ characters (for example, characters distributed according to some regular, but not phylogenetically motivated, principle) are biasing the maximum-parsimony analysis. Bootstrap resampling replaces a single maximum-parsimony analysis with a great number of analyses, each done on a randomly selected subset of the data. If, for example, a single character was responsible for a particular idiosyncratic bifurcation in the analysis computed from the full data set, then this character will be absent in many of the trees in the bootstrap analysis, and thus the idiosyncratic split would also be absent in those trees. If all the trees in the results of the bootstrap analyses contain a particular bifurcation, it can be said that this bifurcation has complete (100 percent) support. If only 90 percent of the trees have a bifurcation, then its support is 90 percent, and so forth. A CONSENSUS TREE is a single tree representation of the main message inferable from the complete set of bootstrap trees. It is built up by cumulatively adopting the bifurcations present in the bootstrap in descending order from highest frequency, discarding lower-frequency bifurcations that conflict with higher-frequency ones, until a complete tree is drawn. The bootstrap support percentage of each split in the tree is conventionally written on the branch, giving a statistical estimate of our confidence in that branch.² High bootstrap values, however, are no guarantee of the accuracy; if the analytic model used is inappropriate an incorrect answer can easily be consistently found. For example, by the ‘long branch attraction’ phenomenon discussed above, if two independent taxa are highly divergent compared to the rest of the taxa, a parsimony analysis will tend to infer for them a shared ancestral node in the tree. This analytic artifact is quite stable, and chances are that the falsely inferred parent node for these two taxa will occur in all trees of the bootstrap sample.

2.5. MAXIMUM LIKELIHOOD. Maximum-likelihood methods assume an explicit model, and seek the model parameters (tree topology and character-state transition probabili-

² Consensus networks have not traditionally been used with bootstrap trees. The reasons are partially historical: consensus networks have become commonplace in phylogenetic analysis only after the heyday of parsimony analysis. Bootstrap trees are NOT equi-probable phylogenetic hypotheses, and we are not aware of any exploration of the appropriateness or proper interpretation of a consensus network of bootstrap trees.

ties) that are most likely to produce the observed data. This is computationally an extremely complex task, currently unfeasible for large numbers of taxa, and here we describe a heuristic technique for maximizing the likelihood function called (METROPOLIS-COUPLED) MONTE CARLO MARKOV CHAIN BAYESIAN PHYLOGENETIC ANALYSIS (henceforth MCMC Bayesian phylogenetic analysis).

MCMC Bayesian phylogenetic analysis is much more complex and unintuitive than parsimony, but has a number of advantages. It incorporates more realistic models of evolution, which can build in independently known facts about the evolutionary behavior of particular characters (for example, different likelihoods of gain vs. loss of a character state). Empirically, Bayesian phylogenetic inference has been shown in simulations to be more likely than methods such as parsimony to retrieve a phylogenetic signal present in the data. It also allows a greater degree of confidence in the results obtained, and is less likely to produce false positives, that is, detecting a signal of relatedness where none exists (Ronquist 2004).

Rate of change of each character is part of the model, and the inferred tree includes information about the rate of change of each character, as well as the overall amount of change on each branch. Inferred rates can be used to make statements about the stability of a character (see for example Pagel et al. 2007 on lexical stability), and branch lengths can be converted to relative chronology (see e.g. Gray & Atkinson 2003, who used a rate-smoothing algorithm under different models to evaluate the dating of Indo-European).

The ‘model’ in a Bayesian phylogenetic analysis is a crucial analytic decision: the method itself does not presume any particular model of evolution. It is the responsibility of the analyst to use a model that is sufficiently rich to represent the historical relationships within the data. This model is in essence a probabilistic estimation of how the observed data came to be produced. A minimal model would include a description of tree topology, branch lengths, and a set of individual transition probabilities for each character. Models can further build in anything that can be formally expressed, for example, varying rates of change, different assumptions about the number of families within the set of taxa, and so forth. The main limitation is computational power, and indeed many Bayesian phylogenetic analyses stretch current computing power to the limits of practicability, requiring weeks or months of processing on a supercomputer.

The Bayesian phylogenetic analysis has the following steps:

- (1) An initial hypothesis specifies the priors: an initial set of transition probabilities and a tree topology. If there is a strong tree signal in the data, the precise values chosen do not matter too much; if the signal is weak then the results will make sense only if the prior probability values are already close to reality. There are mathematical tests that allow one to diagnose whether the assumption of a flat prior probability distribution is valid.
- (2) The parameter values applied to the model allow the likelihood to be calculated that this hypothetical probability distribution produced the observed data (‘likelihood’ is the same as the ‘probability of the observed data given the model’, that is, $L = P(\text{data}|\text{model})$).
- (3) A slight random perturbation is applied to the parameter values to produce a new model near to the old model in the ‘parameter space’ (the randomness in this step is what makes it a ‘Monte Carlo’ process).
- (4) The likelihood calculation is repeated for the new model, that is, the likelihood of this new set of parameter values is calculated (i.e. step 2 above, the

probability that the data could have been produced by this new set of parameter values). The likelihood of the new model is compared to the likelihood of the old (preperturbation) model.

- a. If the likelihood is lower, discard it with a probability proportional to the difference in likelihoods (so a high chance of discarding a very much lower likelihood, and a lower chance of discarding a small difference—this is what is called ‘metropolis coupling’). If the new values are not discarded, they are adopted as a new set of priors.
- b. If the likelihood is higher, these parameters become the priors of the new round.

This step is the Bayesian inference: we update our beliefs about the model based on the information gained from the new observation.

- (5) Take the current priors and return to step 3.

Steps 3 and 4 are repeated many—usually millions—of times (this is the Markov chain), and a sample—perhaps every 20,000th—of the results (tree, model parameter settings, likelihood values) is saved. At the outset of the process the likelihoods fluctuate wildly, and the overall likelihood that the current parameter settings could have produced the observed data is low. In later iterations of the Markov chain the search space is closer to the optimum values, and so acts as an attractor basin, and the likelihood fluctuations are small. These later iterations of the Markov chain move around within the optimum zone freely, but are unlikely to leave it (and if they do manage to leave, they will return quickly to the optimal values again). The trees produced after this equilibrium has been reached are a random sample of the equilibrium zone, and thus can be considered to be all equally likely hypotheses.

The search process can be looked at as a hill-climbing algorithm, where elevation is analogous to the goodness of fit of the tree. Standing at a point, you measure the elevation of another point some way off. If the point is higher, you move to it. If it is lower, you move to it only sometimes—with a high probability if the target is not much lower, and a low probability if it is very much lower (this allows you to escape small local peaks). If there is one major peak, the search procedure will eventually take you there, and further searching will only mean wandering around this peak zone. An obvious problem arises where there are two widely spaced optima. This can be tested empirically, by running multiple analyses of the same data to check that the result is stable. This problem can also be addressed using ‘simulated annealing’, a method that improves the performance of Markov chain optimization processes by adding slowly decreasing amounts of randomness to the search parameters (Felsenstein 2004:52–53).

The set of trees generated by the Bayesian phylogenetic analysis is made up of equally probable phylogenetic hypotheses. Since there may be conflicting phylogenetic signals present in the data, a consensus network (discussed below) is a good tool for summarizing the phylogenetic information contained in the Bayesian tree sample.

2.6. SOME IMPORTANT CONCEPTS: ROOTING, MULTIPLE TREES, AND NETWORKS. The representation of phylogenetic data in a tree is familiar in linguistics.³ The usual phylogenetic tree has a ROOT, giving it temporal directionality reflecting a hypothesis

³ Linguistic trees, however, are generally drawn with their roots at the top, rather than at the bottom as is usual in biology.

about the path of historical development of the elements from a common ancestor. Under biological applications of tree building, a root has to be explicitly chosen. An unrooted phylogenetic tree (i.e. a tree without a root) represents a system of developmental pathways without any hypothesis about direction of change. Depending on the selection of the root, a single unrooted tree can have a number of rooted trees associated with it.

If the direction of change is not known, there are a number of ways that the root of a tree may be determined. First, the root of the tree may be determined by defining an OUTGROUP, a taxon or clade that is presumed to belong to a branch outside the branches that the rest of the taxa belong to; it stands in for an ancestor by defining which of the nodes in the tree is the ancestral, root node. For an outgroup to be applicable, it must be possible to assume that all branchings of the tree occurred after the outgroup split off. It is not valid to pick an arbitrary unrelated taxon as an outgroup: while most phylogenetic methods force all the taxa included in the analysis to appear in a single phylogenetic tree, one cannot be confident that a taxon that is not truly related to the other taxa would join to an unrooted tree at the root. An unrelated taxon intended as an outgroup may join the tree within an otherwise genetic subgroup, motivated by some surface similarity of form.

Another principled way of rooting a tree is midpoint rooting. In midpoint rooting, it is assumed that the two most distantly separated branches of the tree are equidistant from the ancestor, and thus that the root is equidistant from them. The basic assumption of this rooting method is that there has been a more-or-less constant rate of change. This assumption is probably not valid in cases of real linguistic change, but makes for a good first hypothesis.

Since both these ways of rooting trees are problematic for the kind of data in our analysis, we mostly use unrooted trees, so it is important in what follows for the reader to be able to ‘read’ them. The following analogy may help: think of an unrooted tree as a collapsed mobile on the floor. We can pick it up and suspend it at different points (e.g. half-way between (A, B) and (C, D) on the left in Figure 1a, or between A and (B, C, D) on the right), where these points are different possible roots.

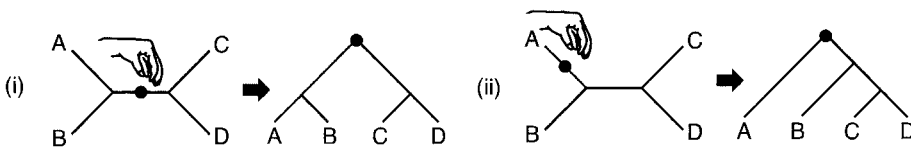


FIGURE 1a. Two different ways to root an unrooted tree.

The structural relations between clades remain constant, but the two different rooting choices imply different subgroupings.

A further important point for interpreting what follows is that, since computational phylogenetic methods are statistical, the outcome of an analysis may not be a single tree. Methods may generate a distribution of trees representing, for example, the degree of certainty for aspects of the hypothesis. When the result of an analysis is a tree set rather than a single tree, it is useful to have tools to summarize the data within the tree set. Figure 1b shows a result tree set of only three trees (real analyses may generate thousands). We summarize the relationships between the trees using majority rules consensus trees, and consensus networks.

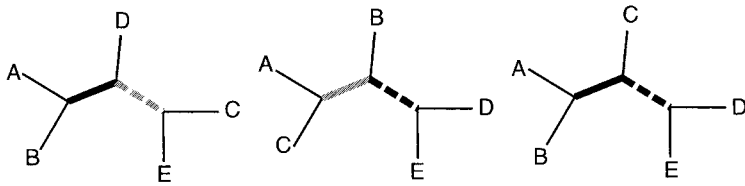


FIGURE 1b. Tree sample showing identical branches between trees.

The MAJORITY RULES CONSENSUS TREE is a tree built by tabulating all the bifurcations present in the tree set ordered by frequency. The three trees in the simple tree set are coded in Figure 1c to illustrate the bifurcations shared by more than one tree. The figure below shows the bifurcations present in the tree set—there are two trees in which (A, B) form a branch distinct from C, D, and E; two trees with (D, E) forming a branch; one with (A, C); and one with (C, E). This figure then shows the majority rules consensus tree generated from this data: the (A, B) branch and a (D, E) branch are most frequent in the tabulated splits (present in two of the three trees in the tree set), and so are added to the consensus tree. The number written on the branch gives the percentage of trees in which the branch occurs, thus providing an indicator of the relative confidence one can have in each branch.

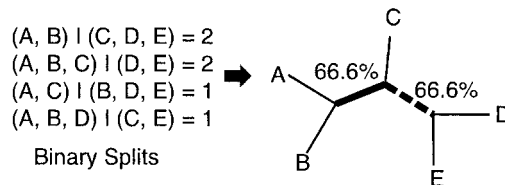


FIGURE 1c. Majority rules consensus tree.

Once these two splits are added to the consensus tree the tree is fully resolved, and the lower frequency splits (A, C) and (C, E)—which each occur in only 33.3 percent of the tree set—do not appear.

While it is useful to have numeric scores making explicit the relative confidence in each of the nodes of a tree, it can be a pity that the lower-score splits in the data set are thrown away. Conflict in the tree set may be indicative of real processes, such as concurrent (simultaneous) tendencies in linguistic change motivated separately by inheritance and by contact. A consensus network is a device for summarizing this conflicting information.

The CONSENSUS NETWORK starts out, like the majority rules consensus tree, from a set of trees to be summarized. The binary splits in the tree set are likewise tabulated (repeated in Figure 1d). The consensus network is drawn by showing conflicting splits as parallelograms. A split is shown as a set of parallel lines with length proportional to the support for that split in the tree set. In the figure below, the long black parallel lines (plain and dashed) represent the instances of these splits in the tree set (drawn with the same color/pattern); the short gray lines represent the conflicting splits, for which there is correspondingly less support.

The NEIGHBORNET method (first use in linguistics reported in Bryant et al. 2005) produces a network visually similar to a CONSENSUS NETWORK. This is no coincidence:

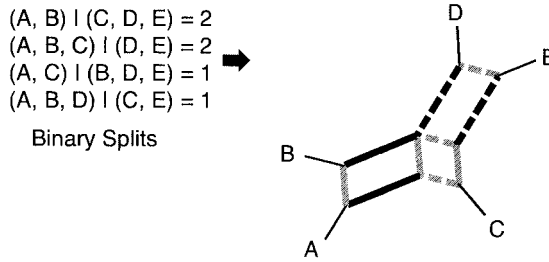


FIGURE 1d. Consensus network.

the NeighborNet graph is produced from a set of binary splits, just like the consensus network. The difference is in the source of the binary splits. The splits used to generate a consensus network are gathered from the binary splits present in a set of phylogenetic trees. In contrast, the NeighborNet method generates the network from distances (measures of overall difference) calculated from tabulated data. A set of tabulated data (such as that in Table 1 in §4.3 below) can be partitioned into the same sets of binary splits as in the consensus network illustration above, and would thus produce an identical network. But the interpretation would be different. While the two types of network look the same, a consensus network provides a summary of the phylogenetic information in a set of phylogenetic trees, while a NeighborNet network provides a ‘phenetic’ summary of the surface similarities between a set of taxa.

This concludes the methodological preliminaries, and we now turn to the languages of Island Melanesia that constitute the data for our study.

3. THE LANGUAGES OF ISLAND MELANESIA: TESTING THE POTENTIAL FOR STRUCTURAL PHYLOGENETICS. This area of multiple islands and island chains is home to languages of very different stocks: more than one hundred languages of the Oceanic branch of the Austronesian family (Lynch et al. 2002:97), and an estimated twenty-five non-Austronesian,⁴ or so-called Papuan, languages whose interrelations are poorly understood, and which are clearly relict languages of pre-Austronesian populations (Dunn et al. 2002).

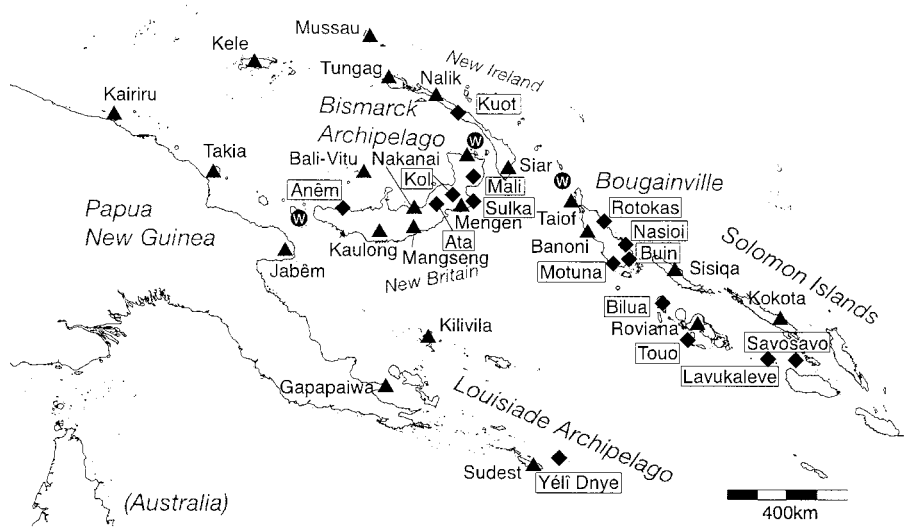
The archaeological record shows presence of modern humans by 40,000 years ago, while the bearers of Austronesian languages arrived only 3,200 years ago.

Relatively speaking, the Oceanic languages have been extensively researched, and their phylogenetic relationships are for the most part reasonably well established (see Lynch et al. 2002). Confusingly to nonspecialists, ‘Papuan’ denotes no established language family—rather, it is a negatively defined areal grouping, denoting all those languages in the region that are not Austronesian. Genealogical relationships among the Papuan languages of Island Melanesia, the region of our research, are most uncertain.⁵

Map 1 shows the region and the languages of our sample.

⁴ To give an exact number of Papuan languages in Island Melanesia is not possible. Some of the languages that have been identified in the literature as non-Austronesian very likely belong to an Oceanic lineage (Dunn & Ross 2007, Ross & Næss 2007). Further, the language named Baining actually consists of a number of separate languages.

⁵ All languages used in the analysis are listed in Appendix A, together with their Ethnologue codes and sources.



MAP 1. Map of Island Melanesia showing the languages under investigation (▲ Oceanic subgroup of Austronesian; ◆ Papuan).

3.1. PRIOR WORK AND THE EXISTING STATE OF THE ART. We refer to the Papuan languages of Island Melanesia as the EAST PAPUAN languages, but note that this is a mere geographic label and does not entail subscribing to the phylum proposed by Wurm (see below) or other previously proposed groupings sometimes so named. These languages have fallen within the scope of a number of controversial claims about distant genetic relatedness, which may have come about as a result of the fact that vocabulary-based methods do not work here—the separation of the languages is at such a time depth as to have eroded any traces of cognacy. For example, Todd (1975) shows that among the four Papuan languages of the Solomon Islands, it is difficult to establish cognate sets beyond shared Austronesian loans. She shows that for a 180-word list, only three words are potential cognates; and many more words have shared similarities through obvious shared Austronesian loans. Ross's (2001a) reconstruction of pronoun paradigms among the Papuan languages of Island Melanesia is equally tentative. Ross notes that 'if there is a genealogical relationship among the island languages, it may be of much greater time depth than that of the [Trans New Guinea] phylum' (2001a:311).

Despite the difficulty of applying the comparative method, however, scholars have hypothesized genealogical groupings based on other types of methods. Most controversial of these was Greenberg's (1971) hypothesis of a giant Indo-Pacific grouping, which included all non-Austronesian languages from the Andaman islands to Tasmania, excluding mainland Australia. Greenberg's method has been severely criticized by historical linguists, for example, Campbell (2003) and Trask (1996), and the Indo-Pacific hypothesis itself is judged invalid by regional experts, for example, Crowley and Dixon (1981) and Pawley (2007). The Indo-Pacific hypothesis in any case presents few specific predictions about the Papuan languages of Island Melanesia, beyond a basic division by island group, New Britain vs. Bougainville vs. Central Solomons. Greenberg also noted that the lexicostatistical classification of Allen and Hurd dividing the Bougainville languages into two groups seemed correct in as far as the data that was then available

(Allen & Hurd 1965). The next major hypothesis is Wurm's 'East Papuan phylum' (1982:231–57). The Allen and Hurd classification was adopted unchanged by Wurm (1982) as his Bougainville Super Stock; see Figure 2. Although Wurm worked with considerably more data than Greenberg, the basis for these suggestions is Greenbergian in style, that is to say, the use of hand-picked features to yield a subjective judgment of relatedness. Wurm here used structural features, claiming that there was too much basic (i.e. core, Swadesh-type) lexical borrowing from Oceanic languages to make lexical data trustworthy. Wurm noted gender, elaborate verb morphology, and pronoun paradigms as evidence for the groupings and subgroupings, but his account is not explicit about the data and the method for arriving at the judgment. This is an important study for us, since it bases its inferences, albeit in an informal way, on a set of structural data (§4.2).

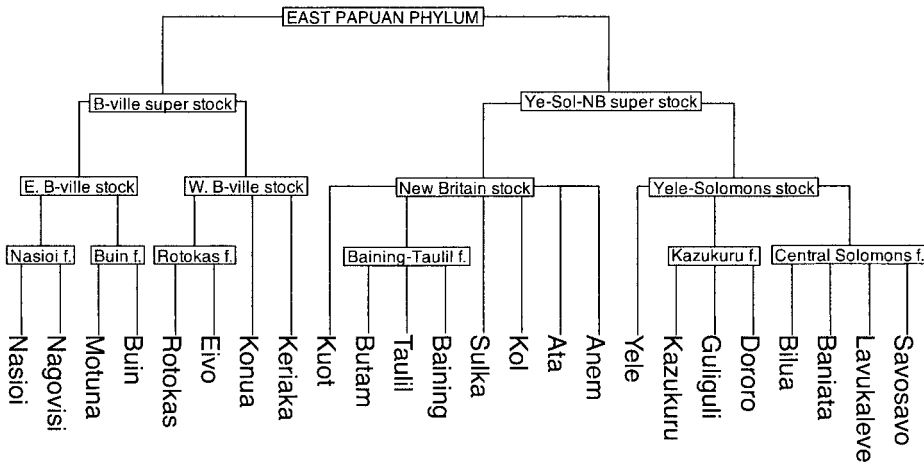


FIGURE 2. Wurm's (1982) East Papuan phylum.

Ross 2001a is the most recent attempt to establish long-range groupings among the East Papuan languages. This was carried out as part of a larger survey of mainland Papuan languages, which followed up on Wurm's suggestions that the pronouns alone may carry the key to establishing relatedness. Ross shows that correspondence between forms in certain pronoun paradigms suggests five families and three isolates in the East Papuan languages (see Figure 3), while offering no clues as to the overall connections between them. He admits that even some of these groupings, especially the connection between Yéî Dnye and his West New Britain family, are questionable.

1. Yéî Dnye (Rossel Island)-West New Britain (Anêm, Ata)
2. East New Britain (Baining, Taulil, Butam)
3. North Bougainville (Konua, Rotokas)
4. South Bougainville (Nagovisi, Nasioi, Motuna, Buin)
5. Central Solomons (Bilua, Touo (= Baniata), Lavukaleve, Savosavo)

Isolates:

- Kol in East New Britain
- Sulka in East New Britain
- Kuot in New Ireland

FIGURE 3. Ross's (2001a) East Papuan groupings.

As Ross (2001a) notes, these groupings are based on a single source of evidence—they are intended as no more than heuristic suggestions for future work, and they cannot themselves be taken to have established any phylogenetic relations. Below we discuss the extent to which Ross's groups are confirmed by the structural phylogeny method (§4.2).

3.2. THE LANGUAGE SAMPLE. For this study we selected twenty-two Oceanic languages (Figure 4) from most of the major divisions for which adequate data are available, covering the area in which the East Papuan languages are found, and sampled at approximately the same density. This sample is slightly different from the one used in Dunn et al. 2005, as in this case we wanted to have representatives of the major subgroups recognized for the Oceanic languages of Melanesia.

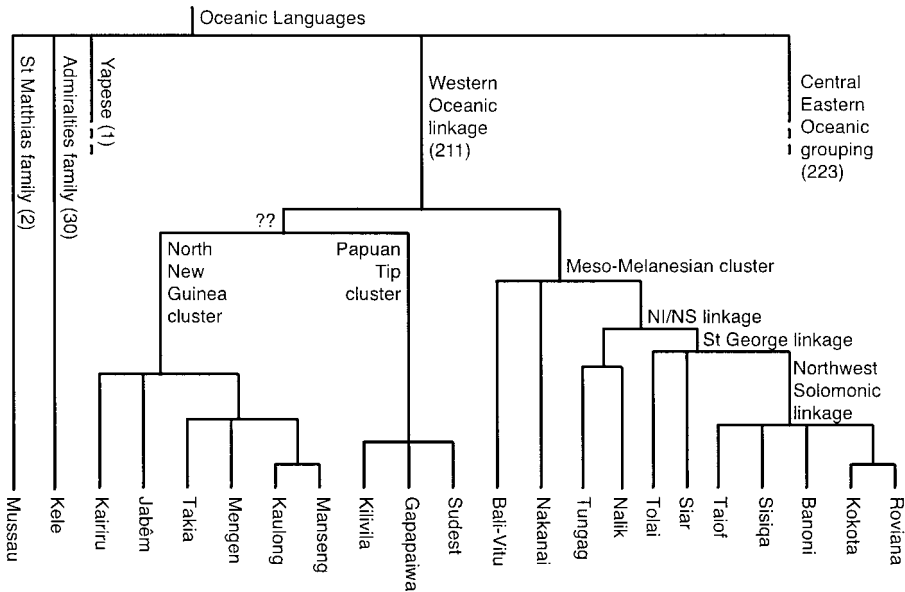


FIGURE 4. Selected Oceanic languages according to CM tree (Lynch et al. 2002).

It is necessary to briefly discuss the status of Oceanic subgroups. Lynch and colleagues (2002:92) assert that Oceanic is a well-defined subgroup of the Austronesian family, because all Oceanic languages reflect a certain set of innovations relative to reconstructed Proto-Malayo-Polynesian. Within Oceanic, however, not all subgroups can be defined strictly by this criterion of shared innovation.

Lynch and colleagues posit three possible primary subgroups of Oceanic: (i) Admiralties family, (ii) Western Oceanic linkage, and (iii) Central/Eastern Oceanic, of which only the Admiralties family is defined by shared innovation (2002:96). They suggest that ‘the Admiralties languages, perhaps together with the St. Matthias languages and Yapese, represent an early Oceanic offshoot’ (2002:98). For this reason we include in our sample Kele and Mussau as representatives of the Admiralties and the St. Matthias groups, respectively.

This linguistic classification does not follow the strict comparative method in that it admits groupings that are not innovation defined, but rather ‘innovation linked’. The term ‘linkage’ as used by Ross and others means that the languages in question share a polythetic set of innovations, and are most likely descended from a dialect network, with the consequence that it is not possible to reconstruct one single proto-language.

This deviation from the strict comparative method is necessary due to a reticulate linguistic prehistory with repeated contact, making it difficult to reconstruct many of the intermediate subgroups between Proto-Oceanic and the lowest-level contemporary clades (i.e. groups of contemporary sister languages). So, for example, the Western Oceanic linkage split over time into three further linkages, (i) Meso-Melanesian, (ii) Papuan Tip, and (iii) North New Guinea (Lynch et al. 2002:99), and each of these into further subgroups, linkages without a clear single parent language or lower-level families for which a common ancestor could be reconstructed. The unity of Western Oceanic is therefore not firmly established. Note too that using the ‘cognate-birth, word-death model’, Greenhill and Gray (2005) do not find support for a Western Oceanic clade within Oceanic.

Four of the five defining innovations of Meso-Melanesian are also found in other Western Oceanic languages. And not all subgroups of the linkage exhibit the morpho-syntactic innovations identified by Ross (1988:271). We take this to mean that the CM has not (yet) succeeded in demarcating the exact boundaries of the subgroups of Western Oceanic. This is important to remember when we come to the comparison of Ross’s results and the clades our method reveals (§4.1). The Meso-Melanesian linkage has a primary division into Bali-Vitu, one of the most conservative languages of Oceanic, the Willaumez linkage, and the New Ireland-Northwest Solomonian linkage, which spreads over an extensive geographic region, from New Hanover to Santa Ysabel of the Solomon Islands. It is in the region of the New Ireland-Northwest Solomonian linkage that most East Papuan languages are found. Thus, in addition to Bali-Vitu, we selected Nakanai and Tolai spoken on New Britain; Tungag, Nalik, and Siar spoken on New Ireland; Taiof and Banoni from Bougainville; and Sisiqa, Roviana, and Kokota from the Solomon Islands.

The structure of the Papuan Tip Oceanic language group is much simpler, exhibiting a number of defining innovations so that a single proto-language can be reconstructed, despite the fact that no single innovation is found in all of the daughter branches. Our sample includes three languages: Kilivila, Sudest, and Gapapaiwa.

The North New Guinea linkage has a far greater internal diversity than either the Meso-Melanesian or the Papuan Tip linkages, especially around the Vitiaz Strait, which separates New Britain from the New Guinea mainland, and along the south coast of New Britain. This area is close to the Willaumez peninsula, the center of diversity of the Meso-Melanesian linkage, which suggests that this area is the likely homeland of the Western Oceanic linkage. At the extremes of the North New Guinea linkage are languages that are much more closely related to each other than to other languages of the linkage. This is true for the Schouten linkage, from which we sampled Kairiru, the Huon Gulf family, represented by Jabêm, and two families from the Ngero/Vitiaz linkage, the Bel family, represented by Takia, and the Mengen family, represented by Mengen. From the South New Britain network we sampled Kaulong and Mangseng, geographically close to some of the few remaining Papuan languages. Altogether then our sample contains twenty-two Oceanic languages, shown in Fig. 4.

Of the more than twenty Papuan languages found in Island Melanesia, we included all languages for which enough data were available, either from published or unpublished sources or from fieldwork carried out by the authors or their colleagues.⁶ The total number of languages considered is fifteen, ordered geographically as follows.

⁶ We particularly thank Claudia Wegener (Savosavo) and Stuart Robinson (Rotokas), who collected data during fieldwork toward their dissertations at the Max Planck Institute for Psycholinguistics, Nijmegen.

- New Britain: Anêm, Ata, Kol, Sulka, and Mali (one of the five Baining languages)
- New Ireland: Kuot, the only Papuan language spoken there
- Bougainville: Rotokas, Nasioi, Motuna, and Buin
- Solomon Islands: Bilua, Touo, Lavukaleve, and Savosavo
- Louisiade Archipelago: Yélfî Dnye, the only Papuan language spoken there (Rossel Island)

3.3. CHOICE OF LINGUISTIC FEATURES OR CHARACTERS. For this study a questionnaire-based database was constructed in which linguistic structural features were coded for presence/absence in each of the target Oceanic and Papuan languages. The full list of 115 characters is given in Appendix B.⁷ These structural characters are abstract, that is, coded without respect to their formal expression. They are selected, on the one hand, to provide broad typological coverage—they include the kind of features one expects to be discussed in a sketch grammar. On the other hand, they have been chosen in part because they are expected to distinguish between the languages of Island Melanesia—there would be little point in including features shared by all the languages, or applicable to none. A previous survey had established some of the variation of the Island Melanesia region (Dunn et al. 2002). In addition, we drew on wider generalizations in the literature about the typological variation in Melanesia and New Guinea. Foley (1986, 1998) discusses a number of features in phonology, morphology, and syntax on which Austronesian and Papuan languages generally diverge. Foley (2000) expands this list for Papuan languages, without systematically identifying their presence or absence in the various Papuan lineages. Lynch and colleagues (2002) provide a typological overview of the Oceanic languages for phonology, parts-of-speech, morphology, and syntax.

On the basis of these typological generalizations we included, for example, phonological characters, such as ‘phonemic distinction between /l/ and /r/’, ‘fricative phonemes’, and ‘word-final consonants’, since presence of the first two and absence of the third tend to be far more common in Oceanic languages than in Papuan languages.

In the domain of morphology, Oceanic languages typically have tense-aspect-mood (TAM) indicated by preverbal particles, often as a portmanteau with the subject marker, while object marking, if present, is expressed by a postverbal suffix or enclitic. In Papuan languages, different strategies are found depending on the major grouping, but in general verbs are the morphologically most complex word class, often marking subject and object as well as a rich TAM system. The majority of Papuan languages are head-marking (Nichols 1992). Other traits figuring in Nichols’s typological studies (e.g. Nichols 1992, 1995, 1997, 2003) involve number (plural neutralization) and gender marking, ergativity, numeral classifiers, an inclusive/exclusive distinction, and others. All of these features were coded for the languages in question. Whereas nouns are generally devoid of inflectional morphology indicating number or gender, both in Austronesian languages and the majority of Papuan languages of the mainland (those of the hypothesized Trans New Guinea (TNG) family), this feature is found in languages in the islands and along the north coast, with complicated noun-class systems in a few areas. Again, our questionnaire contains a number of specific questions, splitting out various factors of these domains. Similarly, alignment of arguments in the clause is coded not in terms of the high-level

⁷ The *Science* paper (Dunn et al. 2005) used 125 features—the 115 features used here reflect a refinement based on subsequent discussion (Donohue & Musgrave 2007, Dunn et al. 2007).

concepts ‘ergative’ or ‘active-stative’, but in terms of specific questions regarding treatment of S, A, and O in basic and complex constructions.

Traits that we know to be uninformative in both language groups are not coded. For example, although tone is a phonological feature in various Papuan groups, and is one of the characteristics of Nichols’s Pacific Rim (1998) ‘province’, it is not coded, since it occurs only once in our sample (in Jabêm). Other possible features such as polysynthesis, proximate/obviative case distinctions, or the existence of a word class of verbs are also not coded, since they are either entirely absent in our sample (in the case of obviative) or overwhelmingly present (in the case of a word class of verbs) in the languages of the world.

Characters that show strong dependencies of covariance (so-called ‘linguistic implicational universals’) were excluded, although characters with weaker tendencies to covariance were not excluded where the current state of linguistic typological knowledge does not allow us to systematically distinguish functionally motivated covariance from phylogenetic or areal patterns. For example, although postpositions and prepositions are crosslinguistically strongly associated with OV and VO word order, our questionnaire treats them as separate characters, because we have evidence that the correlation is not perfect in our region nor in languages of wider New Guinea.

We should emphasize that our list of 115 characters is not intended to be a definitive instrument to be used off-the-shelf for languages of other regions. It is not even claimed that it contains all the possible relevant features for our region, for although we have conducted primary fieldwork to establish the values of characters especially for the Papuan languages, in other cases we have had to rely on published sources. The techniques to be employed require substantially complete data matrices, that is, few or no cells left unfilled, and this establishes a practical limit to the character list. The main point, however, is that, in contrast to the approaches of Greenberg, Wurm, and Nichols, we have made an effort to provide a large body of basic features for each language, which together give a broad typological profile, regardless of whether any given feature seems typologically significant. The resultant phylogenies are thus not likely to reflect a sampling bias.

4. POSSIBLE PHYLOGENIES. In this section, we apply some of the biological methods described above (§2) to the matrix of characters described in the previous section (§3.3) and given in Appendices B and C.

4.1. STRUCTURAL PHYLOGENETICS RECAPITULATES CM ON OCEANIC LANGUAGES. Our aim here is to test whether the structural data—consisting of phonological and morpho-syntactic properties only—can carry a phylogenetic signal. To do so, we compare a computational analysis of the structural data on Oceanic languages to the trees obtained by application of traditional vocabulary-based methods, and in particular to the recent definitive treatment of the Oceanic languages in Lynch et al. 2002.

For a preliminary inspection of the information present in the Oceanic structural data, a NeighborNet network was generated from the matrix of typological characters (Figure 5). This network shows an encouraging degree of congruence with the comparative method tree derived by vocabulary-based methods (Fig. 4 above). The Meso-Melanesian cluster is well defined, with some of the internal structure recovered, such as Tungag and Nalik of New Ireland and Kokota and Roviana of the Solomons. The North New Guinea cluster and the Papuan Tip linkage are interdigitated, which may reflect shared history, as suggested by Lynch and colleagues (2002:108). Our later analyses, Figs. 6 and 7 below, consistently fail to distinguish the North New Guinea

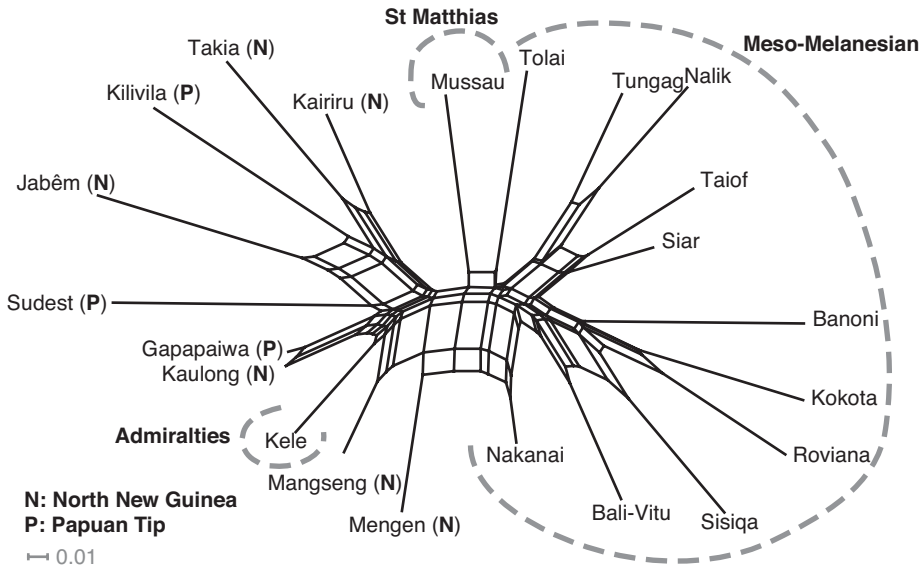


FIGURE 5. NeighborNet of Oceanic languages.

cluster and the Papuan Tip cluster as different clades. The Admiralties language Kele is inserted within the North New Guinea cluster.

We turn now to phylogenetic analysis proper, using both types of algorithm described above, maximum parsimony and Bayesian phylogenetic inference. The parsimony analysis reconstructed considerably less phylogenetic detail than the Bayesian phylogenetic analysis, but for the sake of methodological comparison both results are shown.

A consensus tree of the full bootstrap sample for a maximum-parsimony analysis is shown in Figure 6 (see Dunn et al. 2005 for an earlier application of this method to a similar dataset).

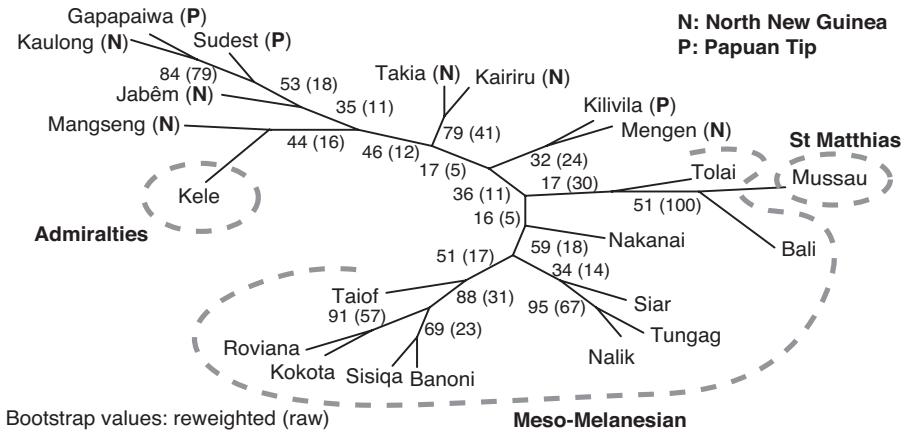


FIGURE 6. Consensus tree for bootstrapped parsimony analysis of Oceanic languages.

The consensus tree of the parsimony analysis (Fig. 6) is not appreciably better than the NeighborNet analysis (Fig. 5). Most of the tree is not resolved to the 50 percent

level, even after reweighting by consistency (Farris 1989, Dunn et al. 2005). The scores above 80 percent are compatible with existing clades in the comparative method tree (with the exception of the Gapapaiwa—Kaulong node), as are some of the lower scores. Meso-Melanesian is weakly distinguished, except for the insertion of Mussau. The bootstrap support for the incorrectly identified Mussau/Bali-Vitu clade decreases with reweighting, showing that it is an artifact of lower-stability features.

The questionnaire that produced the data matrix contains a large number of questions and is fairly robust. When using more consistent features (as in biology, where parallel evolution is comparatively rare), a small number of features is sufficient to produce a high quality tree. With linguistic typological data, the use of a large number of features drastically reduces the probabilities of chance parallel developments, potentially allowing very faint traces of ancient relatedness to be detected. We have no valid grounds for selecting a particular small subset of features as being more likely to preserve a phylogenetic signal, and so we avoid the charge of ‘hand-picking’ features by including in our sample the widest feasible range of noninterdependent typological phenomena (see §3.3). We use a data-driven reweighting method to amplify the signal of the features that are more consistent in the tree and reduce the ‘noise’ produced by highly variable features. While reweighting by consistency produces a better result, it is not a widespread practice in phylogenetics, and should be viewed with caution. Where different characters behave differently, model-based methods such as Bayesian phylogenetic inference can be expected to perform better.

In the Bayesian phylogenetic analysis of this same data, we tested a range of different model assumptions. The results of the highest likelihood model (the model that best explains the data) are shown in Figure 7. This model included distinct parameters for gain and loss of each feature, and assumed features could be classified into four different rate classes (i.e. the analysis inferred values for $2 \times 4 = 8$ parameters).

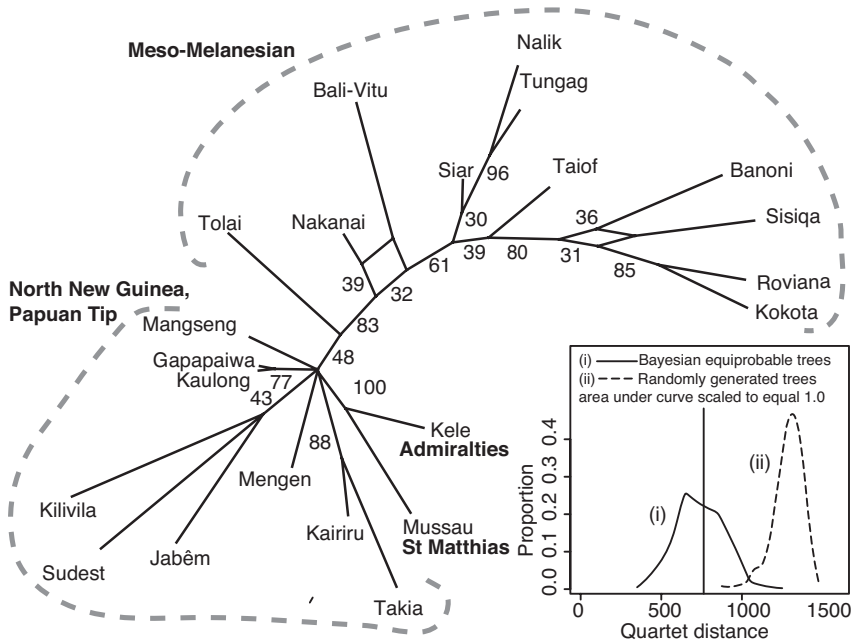


FIGURE 7. Consensus network of the Bayesian posterior tree sample for the Oceanic languages. Inset: comparison of distances from CM tree to (i) Bayesian tree sample, and (ii) randomly generated trees.

The Bayesian network provides a clearer phylogenetic hypothesis with a better match to the CM tree; it is more treelike in form than the NeighborNet network in Fig. 5, and more consistent. The Meso-Melanesian languages are clearly separate from the other languages, and some of the internal structure of this clade is also present (the Northwest Solomonic linkage).

The two histograms inset into Fig. 7 show the tree distance (measured using the quartets distance metric, Felsenstein 2004:530) from the CM tree to (i) each tree in the Bayesian equi-probable tree sample, and (ii) an equal number of randomly generated trees with the same number of taxa. Very few of the randomly generated trees are as near to the CM tree as even the most distant of the Bayesian sample.⁸

We set out to test whether a phylogenetic analysis based on abstract structural features can recapitulate the results obtained by the comparative method applied to sound-meaning correspondences. The outcome is encouraging. It is clear that Bayesian tree inference is superior to a parsimony analysis in detecting a historical signal, and that network analyses can reflect the location and degree of reticulation in the data, as independently noted by scholars using traditional techniques on sound-meaning correspondences.

These results give reason for optimism that a whole new branch of historical linguistics may be within our reach, a branch that would cross-fertilize linguistic typology and historical linguistics, and that may allow us to reach further back in time than the traditional methods based on sound-meaning correspondences allow. We hope that this exercise will encourage those, like the Indo-Europeanists, with much larger bodies of data and ancient literary sources, to undertake a larger-scale test of the hypothesis advanced here, namely that structural phylogeny is an important new tool for exploring historical relationships between languages.

4.2. STRUCTURAL PHYLOGENETICS ON THE PAPUAN LANGUAGES. Given the success of the method as applied to the Oceanic subgroup of Austronesian, whose genealogical relations are independently known, we proceed with a cladistic analysis of the Papuan languages of Island Melanesia, for which only a few low-level families so far have been convincingly identified (Dunn et al. 2002, Ross 2005).

We briefly summarize the interpretation of data similar to those reported earlier (Dunn et al. 2005) as a NeighborNet network in Figure 8. This network is consistent with the geographic distribution of these languages (the island group that each language belongs to is indicated on the graph).

Most computational cladistic methods proceed from the assumption that all taxa are to be included on a single tree. In biology this is not a problem; it is safe, for example, to assume that any pair of mammalian taxa have a common ancestor at some point. But for languages this is not necessarily so. Language histories are short. If our species has had language for over 100,000 years, then the handful of families that can be related at six to ten thousand years antiquity represent less than 10 percent of their history. The vast majority of language families have known histories that are much shorter again. Any or all of the languages making up the Papuan language group might be related or unrelated at levels of antiquity that we can tell nothing about. The null

⁸ As a referee points out, greater-than-chance degree of congruence between a tree sample and a reference tree will be found even when only portions of the trees in the sample are reliably congruent (e.g. a few reliably recovered clades). A better test would take into account the uncertainty in the topology of the reference tree (cf. Greenhill & Gray 2005).

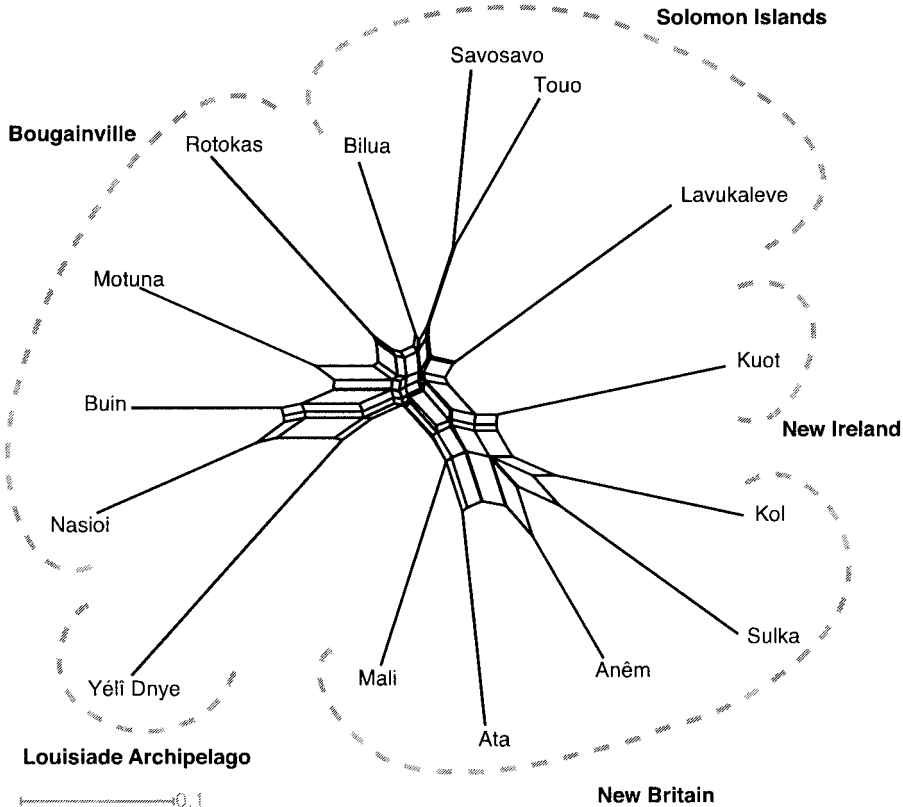


FIGURE 8. NeighborNet of the Papuan languages in Island Melanesia.

hypothesis, that no relationship can be shown between these languages, however, is not supported by the results.

Just as in §4.1 with the Oceanic languages, a Bayesian phylogenetic analysis here on the Papuan languages yields a more perspicuous picture than the NeighborNet graph, as well as the parsimony tree reported earlier in Dunn et al. 2005. The relationships between the East Papuan languages are shown by a Bayesian consensus network in Figure 9. This network isolates the areas where the most conflict is found, clarifying the separation of a number of plausible clades.

Overall, the network exhibits strong splits that coincide with the main archipelagos, namely the Bismarcks, Bougainville, and the Solomon Islands, with some undisputable minor groupings that agree with known genealogical relationships (see §3.1). The position of the Solomons languages, however, is anomalous in relation to the geographic locations of the archipelagos. These languages are positioned in the network between the Bismarcks and Bougainville, while a historical scenario would suggest an increasing distance along the path Bismarcks-Bougainville-Solomons. A plausible solution earlier offered (Dunn et al. 2005:2075) suggests a common ancestor and a separation of Solomons from Bougainville following the rising sea-level at the end of last glacial period (~ 9,000 BP). Another explanation would involve different contact histories, an important question that is investigated in §5.

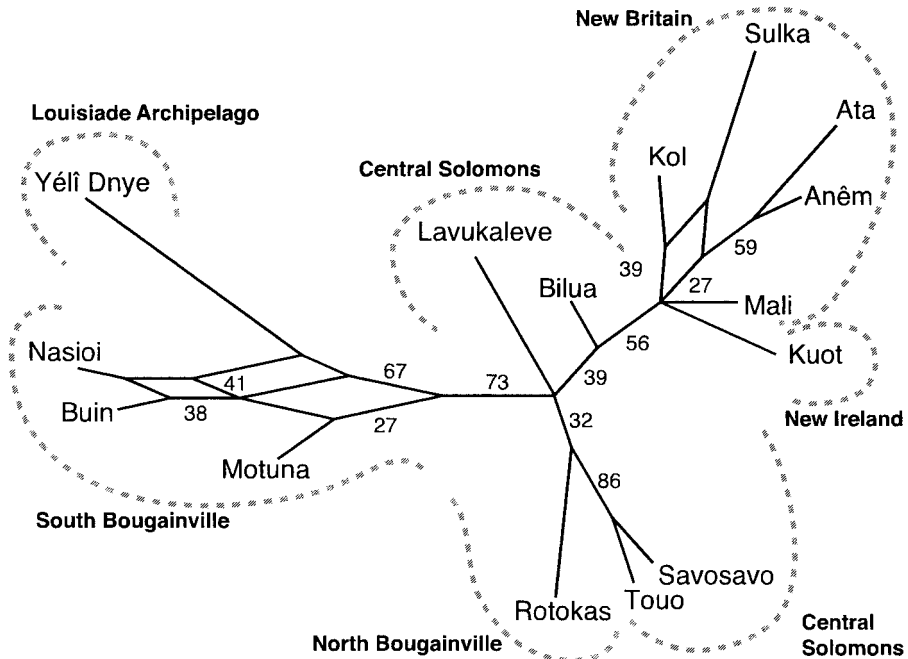


FIGURE 9. Bayesian consensus network of Papuan languages.

The Papuan languages network in Fig. 9 is similar in some details to the previous classifications by Wurm (1982) and Ross (2001a), given in Figs. 2 and 3. Splits in the consensus network reflect that in many trees of the Bayesian posterior sample there are monophyletic clades containing the languages of the Bismarcks Archipelago (New Britain, New Ireland). The Bismarcks clade was proposed by Wurm, but Ross did not find support for it, with the exception of a clade containing Anêm and Ata (also supported by our analysis). The South Bougainville languages (Buin, Nasioi, and Motuna) form a clade both in our analysis and in the proposals by Ross and Wurm.

Yélf Dnye (the overall structural and geographic outlier of the Island Melanesian Papuan languages) is positioned between the South Bougainville languages and the Solomons languages, although there are conflicting trees that instead associate Yélf Dnye with Nasioi and, to a lesser extent, Buin. Rotokas, from North Bougainville, forms a weakly supported clade with two Solomons languages, strongly differentiated from the languages of South Bougainville.

One important caveat should be mentioned with regard to the classification we extract from the trees and networks presented here. The computations that build such trees from a set of features are forced to attach all languages tabulated in the database. Thus, the 'wild cards' in our network, Rotokas of North Bougainville and Yélf Dnye of Rossel Island, may be the victims of this Procrustean treatment. We know that the Bougainville languages belong to two subfamilies, but our sample of necessity contained three languages of the South-Bougainville family and only Rotokas from the Northern family. Hence Rotokas was forced into the tree, even though it lacked its immediate relatives, Konua (a.k.a. Raipoisi) and Eivo. Similarly, Yélf Dnye has no evident relatives, and we have no a priori reason to hypothesize that it should go in any particular place in

the network, yet it had to find some logical place in the network as defined by the sample of languages and the spread of the 115 features. In the various trees and networks we generated, Yélf Dnye invariably attaches close to the Bougainville languages. One reason may be that it is exactly this group of languages that appear to have been least ‘contaminated’ by Oceanic influences (see §5). It does not immediately imply that we now have found a true genealogical niche for Yélf Dnye. At the same time, we could speculate a bit on the basis of some of Wurm’s ideas. Wurm and colleagues (1975: 942) said of the first expansion of the Trans New Guinea phylum in New Guinea Island that it moved ‘into the south-eastern tail-end of the New Guinea mainland [and] drove out another, earlier language group there which moved on to Rossel Island in the Louisiade Archipelago [i.e. present-day Yélf Dnye], and to the New Britain-New Ireland area, perhaps superimposing itself upon even earlier languages there’. In other words, the attachment of Yélf Dnye to our tree may be more than just an artifact of the method.⁹ This possibility could be explored by applying the same instruments that we use here to a much larger sample of Papuan languages from different stocks and wider geographic regions.¹⁰

Note also that the Papuan data and the Oceanic data are the result of very different historical processes, and the phylogenies inferred from these data thus cannot easily be compared within a phylogenetic framework. The differing patterns of diversity between the Papuan and the Oceanic languages are compared below using distance and clustering methods (§§5.1–5.2); importantly, typological distance measures between homogeneous pairs of Oceanic languages are generally smaller than between pairs of Papuan languages (compare also, for example, the lengths of the terminal branches in Figs. 5 and 8; the Papuan languages have had more individual divergence from the center of the graph than the Oceanic languages).

4.3. TYPOLOGICAL STABILITY. The Bayesian phylogenetic inference of the interrelationships between the Oceanic languages is sufficiently close to the CM tree that it is interesting to compare the structural data that went into making it with syntactic aspects of the CM reconstruction of Proto-Oceanic (POc) as presented by Lynch and colleagues (2002:63–89). An important caveat is that the most time-stable items in the database are invariant, and thus phylogenetically uninformative (since they are shared by all languages); the least time-stable items are likely to be phylogenetically uninformative as well. The items that have the greatest influence on the tree topology are those that are a little bit unstable and that changed long enough ago that there are several daughter languages with reflexes of the derived value. The search for the features that are ‘doing the most work’ in the analysis is therefore the search for items of intermediate stability. One can, however, extract useful information from a tabulation of the features that are shared by particular proportions of the data: those that are shared by a large proportion

⁹ Yélf origin myths, for what they are worth, link Yélf Dnye with the Solomon Islands, via a now-submerged island (Pocklington Reef) midway to the Solomons.

¹⁰ Research in progress (Dunn, Reesink, Singer) investigates ‘Structural traces of the Sahul past’ by applying the method we present in this article. The project uses a revised and expanded list of features that are coded for more than 120 languages of a much larger area: one Andaman language, a sample of forty-nine Austronesian languages from the full Austronesian tree from Taiwan into the Pacific, fifty-five Papuan languages, spread from Alor and Pantar and Halmahera in East Indonesia, all of New Guinea to the Solomon Islands, and seventeen Australian languages, representing both the Pama-Nyungan family and non-PN groups. Using the Structure algorithm (see §5.3), a preliminary finding of this work indicates that Yélf Dnye does not cluster with the Papuan languages of Island Melanesia, but with languages of the Trans-Fly area of Papua New Guinea.

of the languages are perhaps the most interesting. Table 1 presents the reconstructed state for a number of features and the proportion of languages in our sample that agrees with this ancestral state, as derived by traditional methods.

NOMINAL DOMAIN			
FEATURE	RECONSTRUCTED STATE	AGREEMENT WITH	
		RECONSTRUCTION	
inclusive/exclusive	Y	22/22	100%
dual-marked noun	N	21/21	100%
plural-marked noun	N	16/21	76%
definite article	Y	8/22	36%
indefinite article	Y	9/20	45%
article-noun order	Y	9/14*	64%
suffix-marked possession	Y	22/22	100%
possessee-marking	Y	22/22	100%
decimal counting	Y	15/22	66%
*8 languages do not have articles			
VERBAL/CLAUSAL DOMAIN			
FEATURE	RECONSTRUCTED STATE	AGREEMENT WITH	
		RECONSTRUCTION	
S-suffix	N	19/22	86%
S-prefix	Y	13/22	59%
A-suffix	N	18/22	82%
A-prefix	Y	13/22	59%
O-suffix	Y	13/22	59%
O-prefix	N	22/22	100%
serial verb construction	Y	18/21	86%
unmarked V-initial	Y?	2/22	9%
unmarked V-medial	Y?	17/22	77%
unmarked V-final	N	17/22	77%
clause-initial negation	Y	13/22	59%
causative affix	Y	18/20	90%

TABLE 1. Frequency of reflexes of reconstructed POc states in Oceanic languages; due to lack of data, not all totals equal twenty-two.

It can be seen that a few features are extremely stable, especially in the nominal domain. All Oceanic languages in the sample have (i) some kind of inclusive/exclusive distinction for first-person plural, and (ii) direct possession by means of a suffix on the possessed noun indexing the possessor. There is no evidence for number marking on the noun in POc, and indeed there is no language in the sample that has dual marking on the noun, despite dual categories in pronouns being quite common across the sample, thus, again, showing a 100 percent agreement with the ancestral state. However, 24 percent of the Oceanic languages allow nominal plural marking. This innovation is possibly due to contact with Papuan languages, 64 percent of which have both plural and dual marking on nouns.

The presence of definite or indefinite articles is not quite so stable. It appears that it is mainly the languages of the Meso-Melanesian linkage that have this feature. Oceanic languages of the other two major branches, the North New Guinea and Papuan Tip linkages, have for the most part lost the article. In some languages there are still vestiges left, for example, the proper noun markers *a* for males and *e* for females in Kaulong (Lynch et al. 2002:393).

In the minority of languages where articles are preserved, 64 percent retain the original article-noun order.

As already pointed out by Lynch and colleagues (2002:72), although the numerals 'one' to 'ten' can be reconstructed for POc forming a decimal system, 'quinary' systems (i.e. compound numbers for 'six' through 'nine', preserving a special term for 'ten') are so widespread that the numbers 'six' through 'nine' must have already been dropping out in the earliest stages of Oceanic. In other words, the noncompound decimal system has only moderate stability, with only 68 percent of our sample retaining the ancestral system. In general, numeral systems do not seem to be robust against contact effects, as shown for neighboring Austronesian and Papuan languages in the Morobe province of Papua New Guinea (Smith 1988). 'Quinary' systems are quite widespread across New Guinea, but especially in the heterogeneous Papuan languages along the northern parts, as already noted by Galis (1960) for the western half of the island.

With regard to the verbal domain, the subject position is predominantly preverbal. In a few languages there is suffixal/enclitical reference to this argument, but this concerns additional number information, as in Bali-Vitu, which has a postverbal *-(a)nga* marking plural for a subject of a transitive verb (Lynch et al. 2002:375).

Significantly, there are no languages in the Oceanic sample that have a prefix or proclitic indexing the object, not even those languages (three out of twenty-two) that have adopted a verb-final order through contact with adjacent Papuan languages. In 59 percent of the sample there is suffix and/or enclitic marking of the object; in the rest the object is expressed by either noun (phrase) or free pronoun. The position and marking options for object are consistent with either a verb-initial or a verb-medial order. Lynch and colleagues (2002:86) appear to favor verb-initial as the reconstruction of POc order, on the basis of the verb-initial order still found in Philippine languages and a rather complex process of reanalysis by POc speakers of aspect-mood adverbial elements and subject proclitics and the grammaticalization of topic fronting. Thus, the two languages in our sample that have a pragmatically unmarked verb-initial order are most probably retentions, rather than innovations. In sum, word order in Oceanic seems to be only moderately stable.

If indeed POc had a clause-initial negator developed out of a negative (matrix) verb (Lynch et al. 2002:88), this feature is rather labile, present in just over half of the languages in the sample. As observed by Lynch and colleagues, there are also a large number of languages with a clause-final negator, in our sample 29 percent; the remaining languages have a preverbal negator. This feature has no strong resilience against areal effects, as all of these orders are also found in Papuan languages.

Finally, the morphological causative affix is rather stable; in 90 percent of the sample there is a reflex of the reconstructed POc **pa[ka]*, allowing a new actor-subject as 'causer' of a basic state or process (Lynch et al. 2002:83).

In conclusion, we can say that the position and function of bound morphemes, possessor in the nominal domain and causative in the verbal domain, are the most time-stable in Oceanic languages, and not restricted to this subgroup of Austronesian. Our sample also suggests that loss of the inclusive/exclusive distinction does not happen easily. Whereas word order of the clause is not impermeable to contact pressure, various indicators of word order, such as subject and object marking in the verbal domain and article-noun in the nominal domain, appear to have a certain robustness. Other features, such as the presence/absence of articles, the position of the negator, and numeral systems, appear to be easily affected by contact with languages with other configurations.

5. DIFFUSION. The diffusion of linguistic features between different languages, or the independent loss of inherited material, can obscure genetic relations, or can render

the CM difficult to apply through lack of regular cognate sets. Small amounts of borrowing can usually be identified and removed from consideration completely, but when there has been too much contact-induced language change, the inheritance profile is blurred; this is one aspect of what Harrison (2003) has called the 'sociohistorical limitations' of the CM.

The biological analog of diffusion—hybridization—occurs more readily between closely related taxa, and many linguists believe that the same applies to structural change in language (though see Thomason & Kaufmann 1988:14ff. for discussion). For the case studies discussed in this article, the consequences of this would be that we predict more diffusion between more recently related languages, and less diffusion across typological boundaries.

Some recent papers have addressed the problem of diffusion using rather different quantitative methods. Ringe and colleagues are concerned with Indo European (IE) historical linguistics, and use computational cladistics as an additional tool for clearing up certain obscurities remaining after application of the CM. Ringe and colleagues (2002) perform a cladistic analysis of the Germanic branch of IE and identify loans that would otherwise obscure the genetic tree. Nakhleh and colleagues (2005) are similarly concerned with identifying loans, including cases of long-term borrowing in which languages remain in contact for a considerable time. Their study is concerned with obtaining what they call a 'perfect phylogeny'—a tree modeling language diversification in which there are no instances of back-mutation (i.e. state changes are in principle irreversible) or parallel evolution (i.e. the initial state cannot occur independently more than once): a perfect phylogeny is 'a phylogenetic tree that is fully compatible with all of the data' (Nakhleh et al. 2005:383). The point of aiming for a perfect tree is that instances where the tree falls short of perfection can generally be identified as instances of language contact. Nakhleh and colleagues use this technique to investigate clades within the Indo-European family that remain controversial, and, using this method, are able to show that 'the IE family, though it did not evolve by means of clean speciation, exhibits a pattern of initial diversification that is close to treelike: the vast majority of characters evolve down the "genetic" tree, and the evolution of the rest can be accounted for by positing limited borrowing between languages' (Nakhleh et al. 2005: 384); in this study borrowings identified in the perfect phylogenetic network amount to 'AT MOST three historically real episodes of contact between the relevant language groups' (p. 407).

Minett and Wang (2003) are also concerned with using computational methods to detect borrowing. They too use the concept of parallel evolution, that is, the notion that a single character should arise in a family tree only once; and where this is not the case, that is, a trait appears more than once, they infer borrowing. They do acknowledge that some linguistic characters (e.g. phonological traits) are more likely than others to have occurred independently in a single tree multiple times.

It appears possible to arrive at a 'perfect phylogenetic' tree or network because it is highly unlikely that cognate sets occur by chance, so that identified cognates can be presumed to have a single source, be it as true cognates with a common ancestral form or due to borrowing. For example, the widespread reflexes of Proto-Austronesian/Oceanic **lima* 'five/hand' are unlikely to be invented independently in any two languages. By contrast, the choice between possessor-possessed or possessed-possessor is 50 percent, and thus two languages can easily agree on this feature by chance. Because of the limited design space inherent to morphosyntactic configurations, it is impossible

to conclude that a pattern occurs unequivocally by descent. Hence, we do not attempt to construct a ‘perfect phylogenetic network’.

It is possible, however, to investigate the relative frequencies of abstract structural features and thus their resilience within a genealogical unit, as we have shown in §4.3. In order to investigate signals of borrowing it is instructive to know which features contribute the most to the diversity found in the sample of languages, as we discuss in the following section. From there we directly tackle the question of the role of contact, by showing how structural distance correlates with geographic distance. Last, we demonstrate an admixture model of the structural features of languages, which treats a language as consisting of recombinations of features inherited from different ancestor languages. This allows us to measure the extent to which it is likely that structural features in any given language may be received from multiple ancestors.

5.1. CONTRIBUTION OF CHARACTERS TO DIVERSITY. In the literature dealing with the heterogeneous Papuan languages and possible contact situations between them and Oceanic languages, one often finds references to ‘typical’ Papuan or Austronesian features (Foley 1986:268, 1998, 2000, Reesink 2005b, Ross 2001b). This, of course, does not mean that such features are ‘Platonic essences’ of different linguistic lineages, as Foley (1998:515) warns. There are no features in our database that occur exclusively in one or the other of the major linguistic groupings, and the import of the phylogenetic approach is that relationships are inferred as structured changes within a feature space, not as a set of individually identifying features. This is a major difference between our approach and the approach described in Nichols 1992.

However, some clusters of features do of course occur more in one lineage than another, and these can be identified statistically, particularly when the classification into lineages is itself a part of the analysis. Rather than just assigning features as more or less characteristic for a certain lineage on the basis of informal frequency judgments, there are mathematical ways to detect feature alignments. One way of investigating which linguistic features contribute most to the language networks presented in §4 is using the standard statistical technique PRINCIPAL COMPONENT ANALYSIS (Baayen 2008). Principal component (PC) analysis is a mathematical procedure that transforms a number of (possibly) correlated variables into a smaller number of uncorrelated variables called principal components. Each of the variables (= the 115 structural features, as defined by our questionnaire) contributes to a limited number of dimensions that show no mutual correlation, which are responsible for the variation found among all the languages in our sample, regardless of their lineage. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

In our case, only five PCs ‘explain’ the variation above a significance level of 0.05, with PC1 accounting for 12 percent of the variation and PC2 for 9 percent, with three other PCs accounting for 6 percent or less. These percentages are rather low for a PC analysis, due to the limited variation within the design space of morphosyntactic configurations. Since the first two PCs, even though they account for a rather small amount of variation, are the most significant, we concentrate on a further analysis of these two components. Figure 10 shows the distribution of languages on the two dimensions given by PC1 and PC2.

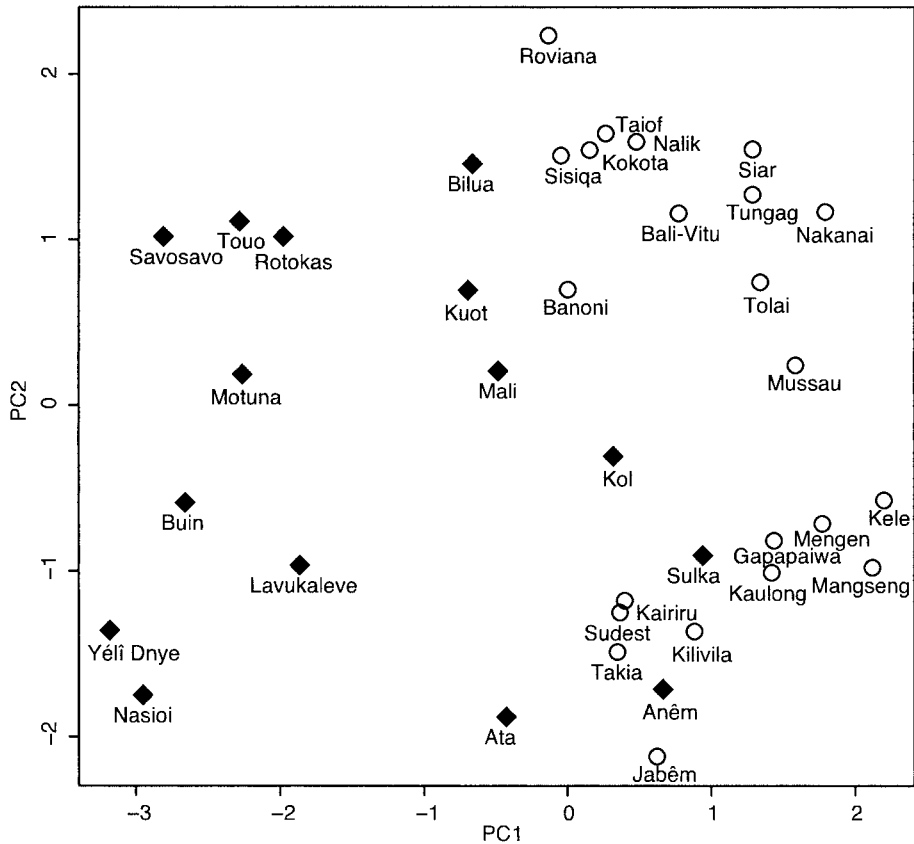


FIGURE 10. PC analysis of combined Oceanic and Papuan language data (components 1 and 2)
(○ Oceanic; ◆ Papuan).

Within the first dimension, a small number of features have a positive or negative loading higher than 0.1, with the bulk of features found around the 0.0 line, meaning that their contribution is negligible. The polarity signs are arbitrary and simply indicate the extremes, maximally separating languages along a dimension. It is worthwhile, therefore, to investigate what the high-loading features are on the dimension that accounts for most of the variation.

Table 2 gives the precise features with highest positive and negative loading on PC1; as can be seen, they include both nominal and verbal morphology and a small number of word order and phonological characteristics. Thus, only twenty-one of the total of 115 features that were coded for each language have a strong contribution, defining PC1 both positively and negatively.

The PC1 is basically the dimension that separates Papuan languages from Oceanic ones, as can be seen along the X-axis in the scatterplot, Fig. 10. Note that the features in Table 2 were not chosen in order to discriminate Oceanic and Papuan languages; they are mathematically identified by the PC analysis as explaining most of the variance. They are not imposed by the researcher, but rather emerge from the analysis.

The only Papuan languages that are on the right side of the dividing line between Oceanic and Papuan are Kol, Anêm, and Sulka of New Britain, consistent with the descriptions that have suggested extensive Oceanic contact phenomena in two of these

	POSITIVE > 0.1	NEGATIVE > 0.1
NOMINAL	number (DU/PL) marked N noun classes/genders conflation of 1 & 2 person pronoun	possessive classifiers suffix-marked possessive
VERBAL	marking of core participants varies ~ Vclass S/A suffix conjugation classes verb suppletion irregular 'give' simultaneous/sequential marking	verb prefix/proclitic realis/irrealis
WORD ORDER	V-final postpositions possessor-possessed order morphosyntactic conflation S/O	V-medial prepositions
PHONOLOGICAL		voicing contrast between stops phonological L and R

TABLE 2. Principal component 1.

languages (Reesink 2005a, Thurston 1982, 1994). The Oceanic languages Sisiqa, Roviana, and Banoni straddle the dividing line.

PC2, represented on the Y-axis of Fig. 10, is less clear-cut, but it still accounts for 8 percent of the variance. Again, only the variables with the highest positive and negative loading on this dimension are given in Table 3. As can be seen, the two PCs are independent from each other; some features (noun classes/genders, possessive classifiers) have a significant loading on both, but with opposite polarity.

	POSITIVE > 0.1	NEGATIVE > 0.1
NOMINAL	possessive classifiers PL-marked N oblique case marking	indef + definite articles conflation of 1 & 2 person pronoun noun classes/genders decimal counting
VERBAL	S/A prefix separate person and number marking on V verb compounds verb + adjunct construction number stem alternation conjugation classes reflexive morphology controlled/uncontrolled morphology	irregular 'give' (in)transitivizing morphology causative by bound affix imperative vs. declarative NEG
WORD ORDER	NP is N-initial clause-final NEG SV of intransitive clause copula	Art N order VS of intransitive clause clause-initial NEG
PHONOLOGICAL		voicing contrast between stops prenasalized stops

TABLE 3. Principal component 2.

This component clearly cuts across the major division of Oceanic and Papuan languages. PC2 distinguishes the two major groups within Oceanic, Meso-Melanesian on the one hand and Papuan Tip/North New Guinea on the other, with Mussau and Kele in the middle. PC2 does not make clear distinctions within the Papuan group.

This result seems to suggest different patterns of diffusion between Papuan languages, which are the descendants of ancient languages present in Island Melanesia, and Oceanic languages, whose common ancestor Proto-Oceanic arrived in the region about 3,200 years ago. As the composition of the two PCs and the spread of languages on these dimensions shows, the full set of linguistic features used in our analysis involves both characters that easily cross such boundaries, either because the design space of a certain linguistic domain does not allow much variation, or because of borrowing between languages, whether related or unrelated. In other words, we have statistically identified clusters of features that are typically associated with Papuan or Oceanic. Thus, in spite of contamination by other signals, clusters of structural features can indeed preserve a phylogenetic signal, confirmed by the congruence between the Oceanic clades arrived at independently by the (less than orthodox) CM and structural phylogenetic methods.

5.2. STRUCTURAL VS. GEOGRAPHIC DISTANCE. While the results reported in §4 are equivocal on the question of whether the connections between the languages are the result of phylogenetic descent or contact, we can directly test to what extent contact between languages of the different lineages is responsible for the obtained groupings. For contact-induced similarities in structure to arise it is necessary that languages are or have been in intense interaction (§2.2 above) and thus in close geographic proximity. In order to compare structural similarities of whatever origin, we calculated the geographic and structural separation of each pair of languages. Geographic distance between pairs of languages was calculated from the geometric center of the region where the languages are currently spoken via waypoints situated between the major archipelagos (rather than direct distances, which do not reflect human accessibility; waypoints are marked on Map 1). The structural distance was calculated as the percentage of disagreement between pairs of feature values in the database (as given in Appendix B). A scatterplot displaying a correlation between geographic and structural distances between the languages of our sample, Figure 11 (Oceanic, Papuan, and mixed pairs), shows that greater structural distance correlates with greater geographic distance.

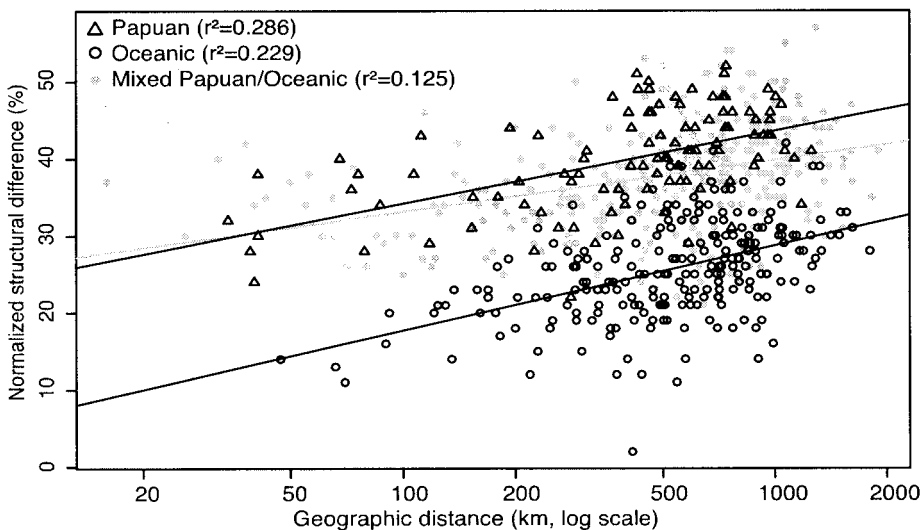


FIGURE 11. Correlations between geographic and structural distance.

This holds for both Oceanic ($r^2 = .229$) and Papuan languages ($r^2 = .286$), but with a greater amplitude in structural distance for the Papuan pairs, reflecting a more ancient dispersal and isolation of Papuan languages than of Oceanic languages.

The correlation between structural and geographic distances for the mixed (Papuan and Oceanic) pairs is much weaker ($r^2 = .125$), indicating that the structural influence of languages from the different stocks on one another is not directly related to geographic proximity, and that the structural-geographic correlation in the Papuan set is largely independent.

While some of the current linguistic patterns can be accounted for by contact history, it is most likely that this occurred against a background of fissioning of populations through isolation by distance, that is, a phylogenetic process. While there is quite some reticulation in the networks we obtained, the clear phenetic dichotomy shown by the PC analysis and the different patterns of the contact signals argue strongly against the proposed explanation of Melanesian linguistic diversity by unrecoverable lateral exchange in situ (Terrell 1986, Terrell et al. 2001).

An alternative diffusional account for this pattern of variation could be motivated by greater ease of diffusion between related taxa. This would assume that Oceanic features have diffused more easily to other Oceanic languages, and likewise Papuan features have diffused more easily to Papuan. This would, however, imply some kind of Papuan unity, that is, that Papuan languages were more similar to each other to begin with.

5.3. 'STRUCTURE' ANALYSIS. In the previous sections we have argued that there are phylogenetic signals in both the Oceanic and Papuan languages of Island Melanesia. There is also evidence of considerable diffusion of abstract structural features between different lineages. In fact, there is hardly any one feature that is found exclusively in one family/linguistic grouping. This is not a problem for the analysis as such, since the structural features as they are defined represent a limited design space within which language variation can occur. Phylogenetic analysis is a statistical reconstruction of the most likely evolutionary pathway to produce the observed variation. But given that there are nonphylogenetic sources of change too, such as lateral transfer (borrowing of features), it also makes sense to look at the data from a feature-centered perspective rather than a language-centered perspective. This recognizes that in their developmental history languages may receive their structural features from different donor languages. This is a model of 'a language' as a collection of structural features that may have their own history different from the aggregate history of the language.

In population genetics, the respective phylogenies of maternal (mtDNA) and paternal (Y-chromosome) mutations can be traced and represented in tree structures. In the case of autosomal markers, which contain recombinant DNA, the establishment of inheritance is more complex, since recombinant DNA is inherited from more than one ancestor for each generation. It is impossible to represent the autosomal DNA relationships between two individuals as a tree, since each recombining element of DNA has its own unique path of descent from the multiple ancestors of each individual. Instead, we can look at the relative contributions of different source populations to each individual. A computational instrument has been developed, called STRUCTURE (Pritchard et al. 2000), that computes the most likely contribution of a given number of ancestral populations to each of the individuals in the sample. The program is told to assume a certain number of ancestral populations and it works out the most likely contribution of each of those ancestral populations to the profile of each individual. The

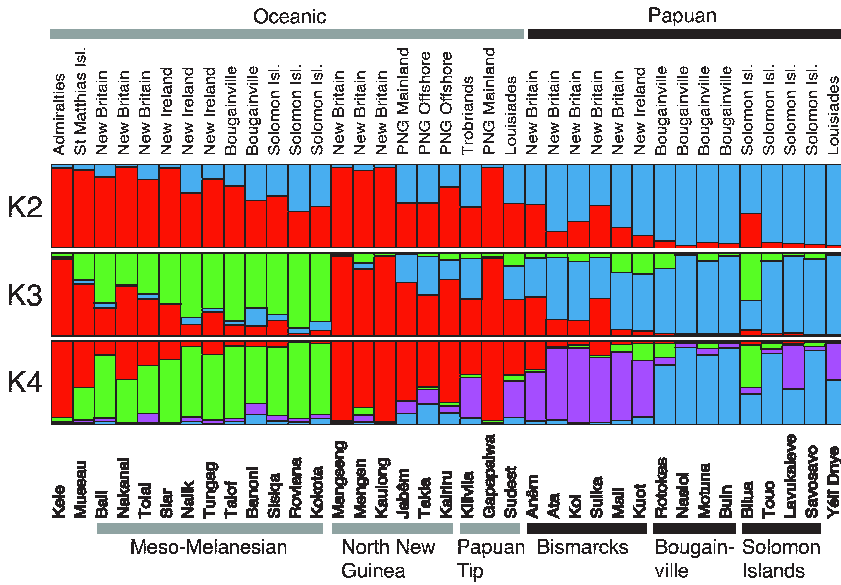


FIGURE 13. Inferred population structure from the highest likelihood run for each of $K = 2$ to $K = 4$.

Four of the New Britain languages (Anêm, Ata, Kol, and Sulka) have elements in common with North New Guinea/Papuan Tip Oceanic languages, which may be a reflection of substrate influence from an ancestral Papuan language.

As the granularity of the analysis is increased by increasing K , the contributing populations look more and more like plausible genealogical units. Importantly, this gives further evidence that the groupings within the Papuan languages are not a product of Oceanic influence. Ongoing research using this method shows more robust results and a higher resolution using a larger number of features (see n. 10).

6. CONCLUSIONS. In this article we have been concerned with finding a method that enables us to detect historical relationships between languages that are not amenable to the comparative method. In particular, the diverse non-Austronesian or Papuan languages of Island Melanesia, surrounded by a far greater number of languages belonging to the Oceanic subgroup of the Austronesian family, have eluded a clear genealogical classification. Earlier attempts by Greenberg (1971) and Wurm (1982), based on loose comparisons, have not been accepted by experts. A stumbling block to any claim of relatedness of these languages is the fact that, apart from some small lower-level families, they lack the lexical and morphological correspondences that would allow an unequivocal identification of cognates, affording the extraction of systematic sound changes and thus eventual reconstruction of ancestral languages.

Yet, the geographic distribution of these languages and the archaeological record that suggests human occupation long before the arrival of the Austronesian-speaking populations requires some account of the relationships between the languages. In order to tackle this question we applied an instrument that attempts to trace language history beyond the threshold of the comparative method. The results of the phylogenetic analysis of the sample of Oceanic languages show a close degree of correspondence to the existing linguistic classification based on sound-meaning correspondences.

We therefore conclude that this instrument can also be applied to a population of languages that are likely to have a much earlier coalescence date. We are not implying

that Island Melanesia has had only one influx of human colonizers since the first migration of c. 40,000 years ago, especially in the light of archaeological findings that the region has known a number of major migrations, both into and out of this area (Friedlaender 1987, Hunley et al. 2007, Lindström et al. 2007, Specht 2005, Spriggs 1997). Data presented here and elsewhere (e.g. Reesink 2005a, Terrill 2003) have shown that lateral exchange of linguistic features between Oceanic and Papuan languages has taken place during the last three millennia. It is not at all implausible to assume that similar contact occurred between the various Papuan groups before that time.

The results of the structural phylogenetic analysis of the Papuan languages, however, suggest a possible historical signal. In particular, the phenetic relationships between the languages (structural distance) clearly correlate with the current geographic position of the languages. A number of factors could motivate this geographic distribution, but we can conclude several things about the structural-geographic correlation, namely (i) it is highly unlikely to be the product of random processes, (ii) it is not the product of recent language contact between Papuan languages (since many of these languages are completely out of contact with each other), and (iii) it is not the product of different degrees of Oceanic contact.

We conclude that there are two possible reasons behind the structure of the Papuan tree, either ancient Papuan-Papuan contact, or descent from a common ancestor. A mixture of both is of course also plausible, although we would tend to favor a phylogenetic origin for the signal of relatedness, given the geographic distribution of the languages. It is not currently possible for us to distinguish common ancestry from ancient contact, but this does not make the linguistic relationships any less interesting, since the Papuan-Papuan contact must have been older than 3,200 years (prior to the Oceanic expansion)—and possibly very much older. This is a period that is otherwise inaccessible to historical linguistics, and our hope is that this article might spur the development of further new methods for exploring linguistic prehistory.

APPENDIX A: LANGUAGES USED IN THE ANALYSIS, ETHNOLOGUE CODES, AND SOURCES

1. AUSTRONESIAN LANGUAGES

Bali-Vitu [bbn] or [wiv]	ROSS, MALCOLM. 2002. Bali-Vitu. In Lynch et al., 362–86.
Banoni [bcm]	LYNCH, JOHN, and MALCOLM ROSS. 2002. Banoni. In Lynch et al., 440–55.
Gapapaiwa [pwg]	MCGUCKIN, CATHERINE. 2002. Gapapaiwa. In Lynch et al., 297–321.
Jabêm [jae]	ROSS, MALCOLM. 2002. Jabêm. In Lynch et al., 270–96.
Kairiru [kxa]	ROSS, MALCOLM. 2002. Kairiru. In Lynch et al., 204–15.
Kaulong [pss]	ROSS, MALCOLM. 2002. Kaulong. In Lynch et al., 387–409.
Kele [los]	ROSS, MALCOLM. 2002. Kele. In Lynch et al., 123–47.
Kilivila [kij]	LAWTON, RALPH. 1993. <i>Topics in the description of Kiriwina</i> . (Pacific linguistics D-84.) Canberra: Australian National University. SENFT, GUNTER. 1986. <i>Kilivila: The language of the Trobriand islanders</i> . Berlin: Mouton de Gruyter.
Kokota [kkk]	PALMER, BILL. 2002. Kokota. In Lynch et al., 498–524. PALMER, BILL. 1999. <i>A grammar of the Kokota language, Santa Isabel, Solomon Islands</i> . Sydney: Department of Linguistics, University of Sydney.
Mangseng [mbh]	MILLIGAN, LLOYD A. 1992. <i>A tentative description of the grammar of the Mangseng language</i> . Ukarumpa: SIL International.
Mengen [mee]	RATH, DANIEL. 1986. Mengen grammar essentials. Ukarumpa: SIL International, MS.
Mussau [emi]	ROSS, MALCOLM. 2002. Mussau. In Lynch et al., 148–66.

- Nakanai [nak] JOHNSTON, RAYMOND L. 1980. *Nakanai of New Britain: The grammar of an Oceanic language*. (Pacific linguistics B-70.) Canberra: Australian National University.
- Nalik [nal] VOLKER, CRAIG. 1998. *The Nalik language of New Ireland, Papua New Guinea*. New York: Peter Lang.
- Roviana [rug] CORSTON-OLIVER, SIMON. 2002. Roviana. In Lynch et al., 467–97.
- Siar [sjr] ROSS, MALCOLM. 2002. Siar. In Lynch et al., 410–25.
- Sisiqa [baa] ROSS, MALCOLM. 2002. Sisiqa. In Lynch et al., 456–66.
- Sudest [tgo] ANDERSON, MIKE, and MALCOLM ROSS. 2002. Sudest. In Lynch et al., 322–46.
- Taiof [sps] ROSS, MALCOLM. 2002. Taiof. In Lynch et al., 426–39.
- Takia [tbc] ROSS, MALCOLM. 2002. Takia. In Lynch et al., 216–48.
- Tolai [ksd] MOSEL, ULRIKE. 1984. *Tolai syntax and its historical development*. (Pacific linguistics B-92.) Canberra: Australian National University.
- Tungag [lcm] STAMM, JOSEF. 1988. A grammar of the Lavongai language. *Lavongai materials* (Pacific linguistics D-82), ed. by Clive Beaumont, 1–46. Canberra: Australian National University.
- 2. PAPUAN LANGUAGES**
- Anêm [anz] THURSTON, WILLIAM R. 1982. *A comparative study of Anêm and Lusi*. (Pacific linguistics B-83.) Canberra: Australian National University.
- THURSTON, WILLIAM R. 1994. Renovation and innovation in the languages of northwestern New Britain. *Language contact and change in the Austronesian world*, ed. by Thomas E. Dutton and Darrel Tryon, 573–609. Berlin: Mouton de Gruyter.
- Ata [ata] HASHIMOTO, KAZUO. (n.d.) *Ata grammar essentials*. Ukarumpa: SIL International, ms.
- YANAGIDA, TATSUYA. 2000. Basic clause structure of the Ata language. Paper presented at the Research School of Pacific and Asian Studies (RSPAS), Australian National University, Canberra, April 26, 2000.
- YANAGIDA, TATSUYA. 2002. Ata verb structure: A split in intransitive predicates. Paper presented at the 5th international Conference on Oceanic Languages (COOL5), Australian National University, Canberra, January 15, 2002.
- Baniata—see Touo
- Bilua [blb] OBATA, KAZUKU. 2003. *A grammar of Bilua: A Papuan language of the Solomon Islands*. (Pacific linguistics 540.) Canberra: Australian National University.
- Buin [buo] LAYCOCK, DONALD C. 2003. *A dictionary of Buin, a language of Bougainville*. (Pacific linguistics 537.) Canberra: Australian National University.
- Kol [kol] REESINK, GER. Fieldnotes.
- LINDRUD, STELLAN. 1989. Aspect in Kol. 14-page ms.
- LINDRUD, STELLAN. 2002. Kol noun classes. 26-page ms.
- LINDRUD, STELLAN. 2002. Kol morpho-syntax. 20-page ms.
- LINDRUD, STELLAN. (n.d.) Non-verbal predication in Kol. 36-page ms.
- Kuot [kto] LINDSTRÖM, EVA. 2008. A grammar of Kuot. Stockholm: Stockholm University, ms.
- LINDSTRÖM, EVA. 2002. *Topics in the grammar of Kuot, a non-Austronesian language of New Ireland, Papua New Guinea*. Stockholm: Stockholm University dissertation.
- Lavukaleve [lvk] TERRILL, ANGELA. Fieldnotes.
- TERRILL, ANGELA. 2003. *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
- Mali [gcc] STEBBINS, TONYA. pers. comm.
- Motuna [siw] ONISHI, MASAYUKI. 1994. *A grammar of Motuna (Bougainville, Papua New Guinea)*. Canberra: Australian National University dissertation.
- ONISHI, MASAYUKI. 2002. *An annotated word list of the Motuna language*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

- Nasioi [nas] HURD, CONRAD, and PHYLLIS HURD. 1966. *Nasioi language course*. Port Moresby: SIL International.
 HURD, CONRAD, and PHYLLIS HURD. 1970. Nasioi verbs. *Oceanic Linguistics* 9.1.37–78.
 HURD, CONRAD, and PHYLLIS HURD. 1977. Nasioi projectives. *Oceanic Linguistics* 16.111–78.
- Rotokas [roo] ROBINSON, STUART. 2008. Split intransitivity in the Rotokas language of Bougainville, ms.
 FIRCHOW, IRWIN B. 1987. Form and function of Rotokas words. *Language and linguistics in Melanesia* 15.1-2.5–111.
- Savosavo [svs] WEGENER, CLAUDIA. 2008. *A grammar of Savosavo, a Papuan language of the Solomon Islands*. Nijmegen: Max Planck Institute for Psycholinguistics and Radboud Universiteit Nijmegen dissertation.
- Siwai—see Motuna
- Sulka [sua] REESINK, GER. Fieldnotes
 SCHNEIDER, JOS. 1962 [1942]. *Grammatik der Sulka-Sprache*. Micro-Bibliotheca Anthropos 36.
 THARP, DOUG. 1996. *Sulka grammar essentials*. Ukarumpa: SIL International.
- Touo [tqu] TERRILL, ANGELA, and MICHAEL DUNN. Fieldnotes.
- Yéfi Dnye [yle] LEVINSON, STEPHEN C. 2008. *A grammar of Yéfi Dnye*. Nijmegen: Max Planck Institute for Psycholinguistics, ms.

APPENDIX B: FULL LIST OF CHARACTERS

1. Are there fricative phonemes?
2. Are there phonemic prenasalized stops?
3. Is there a phonemic distinction between /l/r/?
4. Is there a phonemic velar fricative or glide?
5. Is there a voicing contrast between oral (i.e. non-prenasal) stops?
6. Is there phonemic consonant length?
7. Is there phonemic vowel length?
8. Are there contrastive phonation types for vowels? (e.g. nasal, creaky, etc.)
9. Is there lexically determined suprasegmental prominence? (Suprasegmental prominence can be loudness, duration, pitch, that is, stress or tone phenomena (does not include phonemic vowel length).)
10. Are there word-final consonants?
11. Are there consonant clusters?
12. Are there definite or specific articles?
13. Are there indefinite or nonspecific articles?
14. Is the order of NP elements Art N?
15. Are NPs N-initial (except for articles)?
16. Is there an inclusive/exclusive distinction?
17. Are first and second persons conflated in any context?
18. Are second and third persons conflated in nonsingular numbers? (morphologically in any paradigm; disregard pragmatics/politeness)
19. Are more than two degrees of distance morphologically marked in demonstratives?
20. Are any of the spatial demonstratives not speaker-based? (Speaker-based spatial demonstratives are demonstratives that take as their deictic center the speaker. By contrast, some demonstratives take not the speaker but the addressee as the deictic center, for example a demonstrative might mean ‘close to the speaker’; and some take both speaker and addressee as the deictic center, e.g. ‘far from speaker and addressee’.)
21. Is elevation morphologically marked in demonstratives?
22. Are demonstratives classified?
23. Are there declensions (partly) determined by number of the noun? (By noun declensions is meant, for example, nouns divided into groups that have formally different sets of morphological marking. Do not include place names that can act as bare adjuncts.)
24. Are there declensions (partly) determined by gender of the noun? (By noun declensions is meant, for example, nouns divided into groups that have formally different sets of morphological marking. Do not include place names that can act as bare adjuncts.)

25. Are there nouns that are suppletive for number? (Only yes if present for more than two (basic) kin terms.)
26. Can dual number be marked on the noun itself? (Number-marking on N does not count phrase-level clitic or reduplication.)
27. Is number marking prohibited on certain (types of) nouns? (Not including proper nouns, such as place names or personal names.)
28. Are there noun classes/genders? (By noun classes/genders is meant a system of dividing all or almost all of the nouns of a language into morphological classes that determine agreement phenomena beyond the noun itself.)
29. Are there numeral classifiers? (that is, free or bound morphemes that are nonagreeing, noun-categorization devices, the choice of which is determined by lexical selection)
30. Are there possessive classifiers? (that is, free or bound morphemes that are nonagreeing, noun-categorization devices, the choice of which is determined by lexical selection)
31. Are there possessive classes? (that is, different nouns treated differently in possession according to semantically based groupings. Include alienable/inalienable.)
32. Is alienable/inalienable a relevant distinction?
33. Are there different possessive constructions?
34. Can possession be marked on the nominal possessor?
35. Can possession be marked on the nominal possessee?
36. If the order of elements in a possessive construction is fixed, is it possessor-possessed?
37. Is there a decimal counting system? (that is, elements of decimal; even lexical 10, 10 + 5 qualify)
38. Is there evidence for any element of a quinary counting system? (for example, expressions for 5 + 1, 10 + 5 + 1)
39. Are there words for particular amounts of a thing? (e.g. ten possums)
40. Is there lexical overlap between a significant proportion of adjectives and verbs (including zero-derivation)?
41. Does the same lexical set of adjectives function both attributively and predicatively?
42. Is there case marking for core nominal NPs (i.e. S, A, or O function)? (For case marking, includes any affixal marking that appears in the NP and shows the function of the NP in the clause; adpositions are not counted.)
43. Is there case marking for oblique nominal NPs? (e.g. locational, instrumentals, and so on; adpositions are not counted)
44. Are there prepositions?
45. Are there postpositions?
46. Do the same morphemes systematically encode both TAM and person?
47. Do verbs have prefixes/proclitics?
48. Do verbs have suffixes/enclitics?
49. Is a distinction between punctual/continuous aspect available as a morphological choice?
50. Is a distinction between realis/irrealis mood available as a morphological choice?
51. Is the S participant (at least sometimes) marked by a suffix/enclitic? (Pertains to verb morphology.)
52. Is the S participant (at least sometimes) marked by a prefix/proclitic? (Pertains to verb morphology.)
53. Is the A participant (at least sometimes) marked by a suffix/enclitic? (Pertains to verb morphology.)
54. Is the A participant (at least sometimes) marked by a prefix/proclitic? (Pertains to verb morphology.)
55. Is the O participant (at least sometimes) marked by a suffix/enclitic? (Pertains to verb morphology.)
56. Is the O participant (at least sometimes) marked by a prefix/proclitic? (Pertains to verb morphology.)
57. Are variations in marking strategies of core participants based on TAM distinctions?
58. Are variations in marking strategies based on verb classes?
59. Are variations in marking strategies based on clause type, for example, main vs. subordinate?
60. Are variations in marking strategies based on person distinctions?
61. Do verb stems alter according to the number of a core participant?
62. Do verb stems alter according to the person of a core participant?
63. Is number ever marked separately from person on the verb?
64. Are person, number, and any TAM category (i.e. three or more categories in all) marked by portmanteau morphemes on verbs?
65. Are categories such as person, number, and gender related to a single participant discontinuously marked on a verb?
66. Is a noncore participant marked on the verb? (Include affixes, clitics, and satellite particles associated with verbs forming a constituent with the verb on some level, but exclude optional adverbials.)
67. Can recipients be treated as a transitive object, that is, as direct object?

68. Are there syntactically ditransitive verbs?
69. Is negation marked morphologically on the verbs? (i.e. affixation, stem alternation, neutralization of some inflection)
70. Is direction marked on verbs? (Includes affixes, clitics, and satellite particles associated with verbs forming a constituent with the verb on some level, but excludes optional adverbials.)
71. Are there suppletive verbs for number of participants?
72. Are there conjugation classes?
73. Are there (several) verbs that can be used either transitively or intransitively with no morphological marking? (Not counted if only one or two stems; intended here is the 'break' and 'open' type, not 'John eats/eats the bread'.)
74. Is there transitivizing morphology (include clitics)?
75. Is there morphology (include clitics) to mark a reflexive action? (Free word/particle does not count; neither a default P/N coreference.)
76. Is there morphology (include clitics) to mark a reciprocal action? (Free word/particle does not count; neither a default P/N coreference.)
77. Do verbs classify the shape, size, consistency, or position of absolutive arguments by means of incorporated nouns, verbal affixes, or suppletive verb stems? (Not included here are positional verbs that classify a referent in such terms.)
78. Is there a copula for predicate nouns? (e.g. 'John is a teacher'.)
79. Are there serial verb constructions? (i.e. two or more verbs in juxtaposition, functioning as a single predicate, with no morphology to mark their relationship with each other. Each of the verbs is a separate phonological word but the construction as a whole is expressed in one intonational unit.)
80. Is there one or more auxiliary?
81. Is verb compounding a regular process? (i.e. two or more verb stems acting as one phonological and grammatical word)
82. Are there verb-adjunct (a.k.a. light-verb) constructions? (i.e. constructions involving a nonpredicating element expressing the lexical meaning of the construction, in conjunction with a semantically fairly empty verb, which enables the element to function as a predicate by providing the necessary morphology, for example, eye do for 'see', or sneeze hit for 'sneeze')
83. Is there incorporation of any element into verbs?
84. Is there one or more existential verb? (Exclude positional verbs.)
85. Is the verb 'give' morphologically peculiar (different from most other verbs)? (e.g. stem suppletion, different affixation)
86. Is there a notably small number, that is, about one hundred or fewer, of verbs in the language?
87. Is a pragmatically unmarked constituent order SV for intransitive clauses?
88. Is a pragmatically unmarked constituent order VS for intransitive clauses?
89. Is a pragmatically unmarked constituent order verb-initial for transitive clauses?
90. Is a pragmatically unmarked constituent order verb-medial for transitive clauses?
91. Is a pragmatically unmarked constituent order verb-final for transitive clauses?
92. Is constituent order fixed? (Do not consider 'left or right dislocation', accompanied by intonational signals.)
93. Can negation be marked clause-finally? (This includes suffixes on verb-final clauses; prefixes on clause-final verbs do not count; doesn't include elliptical 'Pete didn't'.)
94. Can negation be marked clause-initially? (Don't include elliptical 'Not Mary'.)
95. Is there a difference between imperative and declarative negation?
96. Are verbal and nonverbal predicates marked by the same negator?
97. Are S and O conflated morphologically in at least some basic constructions, that is, simple main clauses?
98. Are S and A conflated morphologically in at least some basic constructions, that is, simple main clauses?
99. Are S and O conflated morphologically across clause boundaries, that is, acting as syntactic pivot?
100. Are S and A conflated morphologically across clause boundaries, that is, acting as syntactic pivot?
101. Do S and O operate in the same way, and differently from A, for the purpose of any syntactic construction?
102. Is there a morphosyntactic distinction between predicates expressing controlled vs. uncontrolled events or states?

103. Is there clause chaining? (i.e. chains of morphologically stripped-down medial clauses that are dependent on a single clause (usually, but not necessarily, final) for their TAM or participant-marking specification)
104. Is there a morphologically marked distinction between simultaneous and sequential clauses?
105. Is the verb 'say' or a quotative construction used in desiderative constructions? (e.g. 'I said for him to go' for 'I wanted him to go')
106. Are there purposive nonfinite subordinate clauses?
107. Are there temporal nonfinite subordinate clauses?
108. Are there complement clauses?
109. Are causatives formed by serial verb constructions?
110. Are causatives formed by bound affixes/clitics?
111. Are causatives formed by constructions involving 'say'?
112. Is topic or focus marked morphologically? (i.e. by affixes or clitics)
113. Is there tail-head linkage? (i.e. a discourse strategy in which the final verb of one sentence is repeated as the first verb of the next sentence)
114. Are verbs reduplicated?
115. Are nouns reduplicated?

APPENDIX C: CHARACTER MATRIX.

	10	20	30	40	50	60
Mangseng	101010000100001100111000?0?00111101111?000010010000101000000					
Rotokas	100010100001?01011?001101?10000101101?110001011111010000000					
Taiof	1000100001111101001?0000?0010011101010??10010011011010100000					
Sudest	11111010011000111001000010101111101110?110111111100101100000					
Anêm	101110000110001100110000101101110010101010010111010101100100					
Bali	1111000000?110100010000?0010111101011??100101110100?0100000					
Mali	1011100001111101100100011100111011010010010111001101100100					
Motuna	100000000111010100000000?0111111011011011100001001010100100					
Kol	1010101001101111000101010011011101011?010010011001111000100					
Sulka	10101110111111000001010100000100011011000010110110101000000					
Nasioi	10000000011000000010110011?010111111011?111?1001101010100100					
Nakanai	101010000001111001100000000011101010?000010011010000100000					
Nalik	101010000100110100000000?0000011101010110010011101010000111					
Tolai	10101001011101000000000000111101011100010011100101100100					
Yéli_Dnye	110100110000100011100001100001110111010101111101111101100					
Jabêm	110?100001000011001100?0?0100100101101???0111011010101000000					
Gapapaiwa	11101000?110000100010000?0?1011110110101?0010011000101000000					
Bilua	1110100000111011000000?001001110111100?0001011100111110100					
Mengen	101011001110101100110000?0000011101001?000010011001011010000					
Kairiru	101000000110000100110000?0000011101001?110011011000101100000					
Roviana	11111000001110100010000?0000011101110??1001001100000100000					
Sisiqa	11110000?01100100010000?000001111110?10001?111011010000000					
Kele	111010001110001100000000000011111010100010010010101000000					
Buin	00001000?11000?0110000001001101110110??111011010010101001?0					
Kokota	101110000011000100100000?00000111011100110010101010000010000					
Kaulong	11101000?110001100000000?0?10111101101??0010011000101000000					
Takia	10101000011000110010000?00001111010010??11?1011110101100000					
Tungag	10101010?110111100000000100000111010110010010110000000001011					
Banoni	10011010011?01011110000?0001011101110??0010111010000100000					
Kilivila	101010001110000100100100001001111011010010110011010101100000					
Siar	10101000011111010010000010010111101010?010010010??0000000000					
Ata	10010000?00000110000000?0010111001001??10110111?01101100100					
Mussau	101101101100?00100100000?0010111101010?010010011000101100000					
Touo	11011001?001110110000000010100001001110010001001110000100100					
Kuot	10101000110000010010101011010011100010001001001110111110101					
Savosavo	11111000?001010111000000110100000101100001001011??0000110100					
Lavukaleve	101110001111001110000011110100000011101110101011001111111010					

	70	80	90	100	110
	•	•	•	•	•
Mangseng	0000000000001?1011?10??0010010110110101000010??11??11				
Rotokas	0000001000010111000?000000110011011101????10011001?0011				
Taiof	0000001?00?0?101001100??0110101001?010100001??0?1?0?1?				
Sudest	??0100??11?001?11?0?000?00100101111101010000??110101?11				
Anêm	1011101000110111001001010010010110110101000011001000110				
Bali	001000000?001?100101?0?00100100011101010??1?001?000?11				
Mali	0000001000010011001000110011010101101100010001110000110				
Motuna	?000001110?01101010100010010001101?110111110?1110101111				
Kol	11100000000101000010?1010?1101010101110001001??11000100				
Sulka	10110000001001010010?0010010010100000101000010001001110				
Nasioi	0010001110110?1101?01?0?10100011100?01??11010??1000011				
Nakanai	000000000000101001?101?00100101001101010000?0010101111				
Nalik	00001010001001010111001100100101011101010000?1110101?1?				
Tolai	0000000001000101001?1?1?00110101010111010100??11??11				
Yéfi_Dnye	111110001111001100000010101000100010111110100100001011				
Jabêm	0000?10000?1????010????0010010?10010101??10111?????01				
Gapapaiwa	0001000?0?0?1?1001?????0100101101101010?????10100?1?				
Bilua	0000001100?00111000000101010010001?101010000?1110100111				
Mengen	00000000100001010100101100100101000?0101000001100101?11				
Kairiru	?00001??0??1?1001?????0100011101101010??0??1000?1?				
Roviana	0000001100?0?1010010000?001100001?11110000??11?1?1?10				
Sisiqa	000100??111001?10010?00??001010101110101?0??011100?1?				
Kele	0000000000001?001011?00100101111101010?001000?10????				
Buin	0011001?0011011100?0?0?00100011000?01??0??????1?0?11				
Kokota	000000?100?0?10000111?11?001100001010101010?0??10001?11				
Kaulong	0001000?00?0?1?1001??0?0100101101101010?????10100?1?				
Takia	110000010110??11001?1?11?010001?00110101001??111?1??00				
Tungag	0000001?00?011?100?1??0010010101110101000001110100?11				
Banoni	0011011101?001?10010?0100?0100001101?10?00?1?11100?1?				
Kilivila	1010101110001000101011001001010101010100000111011111				
Siar	0000001000?0010100101?0001001010111010100?001110100?11				
Ata	??01011100?110??11????00100101100111001100?1?1??1111				
Mussau	0000000001001111001?10?100100100011?010100000001010??1?				
Touo	00000011101011000010??011010001110000101001000001100110				
Kuot	0000001100111011001100000001100101111101000010010000110				
Savosavo	1100101100111100001000011010001101110101?01?0??1001111				
Lavukaleve	0010001010100101001101010010001110110111001011101111110				

APPENDIX D: TECHNICAL NOTE

The Bayesian phylogenetic analyses carried out in this article were done using the BayesPhylogenies program by Pagel and colleagues (2007), NeighborNets were built using SplitsTree4 (Huson 1998, Bryant et al. 2005), and the admixture models of the Structure algorithm were implemented in the package by Pritchard (2000). General statistical tasks (including principal components analysis) and plotting were done using the R statistical computing language (R Core Development Team 2008), including the ‘ape’ library (Paradis et al. 2004), which provides an implementation of the Mantel test, as well as functions for working with phylogenetic trees. The venerable PAUP* program (Swofford 1998, the only nonfree software used in the analysis) was used for the maximum-parsimony analysis.

REFERENCES

- ADELAAR, WILLEM, and PIETER MUYSKEN. 2004. *The languages of the Andes*. Cambridge: Cambridge University Press.
- AIKHENVALD, ALEXANDRA Y. 1996. Areal diffusion in Northwest Amazonia: The case of Tariana. *Anthropological Linguistics* 38.73–116.

- AIKHENVALD, ALEXANDRA Y., and ROBERT M. W. DIXON (eds.) 2001. *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press.
- ALLEN, GERALD, and CONRAD HURD. 1965. *Languages of the Bougainville district*. Port Moresby: Department of Information and Extension Services.
- ANTILLA, RAIMO. 1972. *An introduction to historical and comparative linguistics*. New York: MacMillan.
- BAAYEN, R. HARALD. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- BLUST, ROBERT. 2000. Why lexicostatistics doesn't work. *Time depth in historical linguistics*, vol. 2, ed. by Colin Renfrew, April McMahon, and Larry Trask, 311–31. Cambridge: McDonald Institute for Archaeological Research.
- BRYANT, DAVID; FLAVIA FILIMON; and RUSSELL GRAY. 2005. Untangling our past: Languages, trees, splits and networks. In Mace et al., 67–84.
- CAMPBELL, LYLE. 1995. The Quechumaran hypothesis and lessons for distant genetic comparison. *Diachronica* 12.2.157–200.
- CAMPBELL, LYLE. 1998. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- CAMPBELL, LYLE. 2003. How to show languages are related: Methods for distant genetic relationship. In Joseph & Janda, 262–82.
- CROFT, WILLIAM. 2004. Typological traits and genetic linguistics. Albuquerque: University of New Mexico, ms. Online: <http://www.unm.edu/~wcroft/Papers/Typ-Gen.pdf>.
- CROWLEY, TERRY, and ROBERT M. W. DIXON. 1981. Tasmanian. *Handbook of Australian languages*, vol. 2, ed. by Robert M. W. Dixon and Barry Blake, 395–421. Canberra: Australian National University Press.
- CURNOW, TIMOTHY J. 2001. What language features can be 'borrowed'? In Aikhenvald & Dixon, 412–36.
- DIXON, R. M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- DONOHUE, MARK, and SIMON MUSGRAVE. 2007. Typology and the linguistic macrohistory of Island Melanesia. *Oceanic Linguistics* 46.2.325–64.
- DUNN, MICHAEL; ROBERT A. FOLEY; STEPHEN C. LEVINSON; GER REESINK; and ANGELA TERRILL. 2007. Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Oceanic Linguistics* 46.2.388–403.
- DUNN, MICHAEL; GER REESINK; and ANGELA TERRILL. 2002. The East Papuan languages: A preliminary typological appraisal. *Oceanic Linguistics* 41.28–62.
- DUNN, MICHAEL, and MALCOLM ROSS. 2007. Is Kazukuru really Austronesian? *Oceanic Linguistics* 46.210–31.
- DUNN, MICHAEL; ANGELA TERRILL; GER REESINK; ROBERT A. FOLEY; and STEPHEN C. LEVINSON. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309.2072–75.
- DURIE, MARK, and MALCOLM ROSS (eds.) 1996. *The comparative method reviewed: Regularity and irregularity in language change*. Oxford: Oxford University Press.
- EDWARDS, A. W. F., and L. LUCA CAVALLI-SFORZA. 1963. The reconstruction of evolution. *Annals of Human Genetics* 27.105–6.
- FARRIS, JAMES S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5.417–19.
- FELSENSTEIN, JOSEPH. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer.
- FOLEY, WILLIAM A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- FOLEY, WILLIAM A. 1998. Toward understanding Papuan languages. *Perspectives on the Bird's Head*, ed. by Jelle Miedema, Cecilia Odé, and Rien A. C. Dam, 503–18. Amsterdam: Rodopi.
- FOLEY, WILLIAM A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29.357–404.
- FRIEDLAENDER, JONATHAN S. (ed.) 1987. *The Solomon Islands project: A long-term study of health, human biology, and culture change*. Oxford: Oxford University Press.
- FRIEDLAENDER, JONATHAN S. (ed.) 2007. *Genes, language and culture history in the Southwest Pacific*. New York: Oxford University Press.

- GALIS, K. W. 1960. Telsystemen in Nederlands-Nieuw-Guinea. *Nieuw Guinea Studiën* 4.131–49.
- GRAY, RUSSELL D., and QUENTIN D. ATKINSON. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426.435–39.
- GRAY, RUSSELL D., and FIONA M. JORDAN. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405.1052–55.
- GREENBERG, JOSEPH H. 1971. The Indo-Pacific hypothesis. *Linguistics in Oceania* (Current trends in linguistics 8), ed. by Thomas A. Sebeok, 807–71. The Hague: Mouton.
- GREENHILL, SIMON, and RUSSELL GRAY. 2005. Testing population dispersal hypotheses: Pacific settlement, phylogenetic trees and Austronesian languages. In Mace et al., 31–52.
- HARRIS, ALICE C., and LYLE CAMPBELL. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- HARRISON, S. P. 2003. On the limits of the comparative method. In Joseph & Janda, 213–43.
- HAWKINS, JOHN A. 1983. *Word order universals: Quantitative analyses of linguistic structure*. New York: Academic Press.
- HOCK, HANS HENRICH, and BRIAN D. JOSEPH. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.
- HOLDEN, CLARE J. 2001. Bantu language trees reflect the spread of farming across sub-Saharan Africa: A maximum-parsimony analysis. *Proceedings of The Royal Society of London, Series B* 269.793–99.
- HUNLEY, KEITH; MICHAEL DUNN; EVA LINDSTRÖM; GER REESINK; ANGELA TERRILL; HEATHER NORTON; LAURA SCHEINFELDT; FRANÇOISE R. FRIEDLAENDER; D. ANDREW MERRIWETHER; GEORGE KOKI; and JONATHAN S. FRIEDLAENDER. 2007. Inferring prehistory from genetic, linguistic, and geographic variation. In Friedlaender 2007, 118–40.
- HUSON, DANIEL. 1998. SplitsTree—a program for analyzing and visualizing evolutionary data. *Bioinformatics* 14.68–73.
- JOSEPH, BRIAN D., and RICHARD D. JANDA. 2003. *The handbook of historical linguistics*. Oxford: Blackwell.
- KESSLER, BRENT. 2001. *The significance of word lists*. Stanford, CA: CSLI Publications.
- LINDSTRÖM, EVA; ANGELA TERRILL; GER REESINK; and MICHAEL DUNN. 2007. The languages of Island Melanesia. In Friedlaender 2007, 141–54.
- LYNCH, JOHN; MALCOLM ROSS; and TERRY CROWLEY. 2002. *The Oceanic languages*. London: Curzon Press.
- MACE, RUTH; CLARE J. HOLDEN; and STEPHEN SHENNAN (eds.) 2005. *The evolution of cultural diversity: Phylogenetic approaches*. London: University College London Press.
- MATTHEWS, PETER. 1982. *Do languages obey general laws?* Cambridge: Cambridge University Press.
- MCMAHON, APRIL, and ROBERT MCMAHON. 2003. Finding families: Quantitative methods in language classification. *Transactions of the Philological Society* 101.7–55.
- MCMAHON, APRIL, and ROBERT MCMAHON. 2005. *Language classification by numbers*. Oxford: Oxford University Press.
- MINETT, JAMES W., and WILLIAM S.-Y. WANG. 2003. On detecting borrowing: Distance-based and character-based methods. *Diachronica* 20.289–331.
- MITHUN, MARIANNE. 2007. Integrating approaches to diversity: Argument structure on the NW Coast. *Diversity in language: Perspectives and implications* (Lecture notes 176), ed. by Yoshiko Matsumoto, David Oshima, Orrin Robinson, and Peter Sells, 9–36. Stanford, CA: CSLI Publications.
- MITHUN, MARIANNE. 2009. Core argument patterns and deep genetic relations: Hierarchical systems in Northern California. *Typology of argument structure and grammatical relations*, ed. by Bernard Comrie. Amsterdam: John Benjamins, to appear.
- MORAVCSIK, EDITH A. 1978. Language contact. *Universals of human language, vol. 1: Method and theory*, ed. by Joseph H. Greenberg, 93–122. Stanford, CA: Stanford University Press.
- NAKHLEH, LUAY; DON RINGE; and TANDY WARNOW. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81.382–420.

- NICHOLS, JOHANNA. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66.475–521.
- NICHOLS, JOHANNA. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- NICHOLS, JOHANNA. 1995. Diachronically stable structural features. *Historical linguistics 1993: Selected papers from the 11th International Conference on Historical Linguistics, Los Angeles, 16–20 August 1993* (Current issues in linguistic theory 124), ed. by Henning Andersen, 337–56. Amsterdam: John Benjamins.
- NICHOLS, JOHANNA. 1996. The comparative method as heuristic. In Durie & Ross, 39–71.
- NICHOLS, JOHANNA. 1997. Sprung from two common sources: Sahul as a linguistic area. *Archaeology and linguistics: Global perspectives on ancient Australia*, ed. by Patrick McConvell and Nicholas Evans, 135–68. Melbourne: Oxford University Press.
- NICHOLS, JOHANNA. 1998. The origin and dispersal of languages: Linguistic evidence. *The origin and diversification of language*, ed. by Nina Jablonski and Leslie C. Aiello, 127–70. San Francisco: California Academy of Sciences.
- NICHOLS, JOHANNA. 2003. Diversity and stability in language. In Joseph & Janda, 283–310.
- NORMAN, JERRY. 1988. *Chinese*. Cambridge: Cambridge University Press.
- PAGEL, MARK; QUENTIN ATKINSON; and ANDREW MEADE. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449.717–20.
- PARADIS, EMMANUEL; JULIEN CLAUDE; and KORBINIAN STRIMMER. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20.289–90.
- PAWLEY, ANDREW. 2007. Recent research on the historical relationships of the Papuan languages, or, What can linguistics add to the stories of archaeology and other disciplines about the prehistory of Melanesia? In Friedlaender 2007, 36–58.
- PAWLEY, ANDREW; ROBERT ATTENBOROUGH; JACK GOLSON; and ROBIN HIDE (eds.) 2005. *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*. (Pacific linguistics 572.) Canberra: Australian National University.
- PRITCHARD, JONATHAN K. 2000. Documentation for *structure* software. Oxford: University of Oxford.
- PRITCHARD, JONATHAN K.; MATTHEW STEPHENS; and PETER DONNELLY. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155.945–59.
- R DEVELOPMENT CORE TEAM. 2008. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Online: <http://www.R-project.org>.
- RANKIN, ROBERT L. 2003. The comparative method. In Joseph & Janda, 183–212.
- REESINK, GER. 2005a. Sulka of East New Britain: A mixture of Oceanic and Papuan traits. *Oceanic Linguistics* 44.1.145–93.
- REESINK, GER. 2005b. West Papuan languages: Roots and development. In Pawley et al., 185–218.
- REXOVÁ, KATEŘINA; DANIEL FRYNTA; and JAN ZRZAVÝ. 2003. Cladistic analysis of languages: The Indo-European classification based on lexico-statistical data. *Cladistics* 19.120–27.
- RINGE, DON; TANDY WARNOW; and ANN TAYLOR. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100.59–129.
- RONQUIST, FREDRIK. 2004. Bayesian inference of character evolution. *Trends in ecology and evolution* 19.475–81.
- ROSENBERG, NOAH A.; JONATHAN K. PRITCHARD; JAMES L. WEBER; HOWARD M. CANN; KENNETH K. KIDD; LEV A. ZHIVOTOVSKY; and MARCUS W. FELDMAN. 2002. Genetic structure of human populations. *Science* 298.2381–85.
- ROSS, MALCOLM. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. (Pacific linguistics C-98.) Canberra: Australian National University.
- ROSS, MALCOLM. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In Durie & Ross, 180–217.
- ROSS, MALCOLM. 1999. Exploring metatypy: How does contact-induced typological change come about? Keynote talk given at the Australian Linguistic Society's annual meeting, Perth, 1999.
- ROSS, MALCOLM. 2001a. Is there an East Papuan phylum? Evidence from pronouns. *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton*, ed.

- by Andrew Pawley, Malcolm Ross, and Darrell Tryon, 301–21. Canberra: Pacific Linguistics.
- ROSS, MALCOLM. 2001b. Contact-induced change in Oceanic languages in north-west Melanesia. In Aikhenvald & Dixon, 134–66.
- ROSS, MALCOLM. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Pawley et al., 15–65.
- ROSS, MALCOLM, and ÅSHILD NÆSS. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands. *Oceanic Linguistics* 46.2.456–98.
- SAUSSURE, FERDINAND DE. 1916. *Cours de linguistique générale*, ed. by Charles Bally and Albert Sechehaye, with the collaboration of Albert Riedlinger. Paris: Payot.
- SMITH, GEOFFREY P. 1988. Morobe counting systems. *Papers in New Guinea linguistics* 26.1–132. (Pacific linguistics A-76.) Canberra: Australian National University.
- SPECHT, JIM. 2005. Revisiting the Bismarcks: Some alternative views. In Pawley et al., 235–88.
- SPRIGGS, MATTHEW. 1997. *The Island Melanesians*. London: Blackwell.
- STEEVER, SANFORD B. 1998. *The Dravidian languages*. London: Routledge.
- SWOFFORD, DAVID L. 1998. *PAUP*: Phylogenetic analysis using parsimony (* and other methods)*. Version 4. Sunderland, MA: Sinauer Associates.
- TERRELL, JOHN. 1986. *Prehistory in the Pacific Islands: A study of variation in language, customs, and human biology*. Cambridge: Cambridge University Press.
- TERRELL, JOHN E.; KEVIN M. KELLY; and PAUL RAINBIRD. 2001. Foregone conclusions? In search of ‘Papuan’ and ‘Austronesians’. *Current Anthropology* 42.97–124.
- TERRILL, ANGELA. 2003. Linguistic stratigraphy in the central Solomon Islands: Lexical evidence of early Papuan/Austronesian interaction. *Journal of the Polynesian Society* 112.369–401.
- THOMASON, SARAH G. 2001. *Language contact: An introduction*. Washington, DC: Georgetown University Press.
- THOMASON, SARAH G., and TERRENCE KAUFMAN. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- THURSTON, WILLIAM R. 1982. *A comparative study of Anêm and Lusi*. (Pacific linguistics B-83.) Canberra: Australian National University.
- THURSTON, WILLIAM R. 1994. Renovation and innovation in the languages of north-western New Britain. *Language contact and change in the Austronesian world*, ed. by Thomas E. Dutton and Darrell Tryon, 573–609. Berlin: Mouton de Gruyter.
- TODD, EVELYN. 1975. The Solomon language family. *Papuan languages and the New Guinea linguistic scene* (Pacific linguistics C-38), ed. by Stephen A. Wurm, 805–46. Canberra: Australian National University.
- TRASK, LARRY. 1996. *Historical linguistics*. London: Arnold.
- TUFFLEY, CHRIS, and MIKE STEEL. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bulletin of Mathematical Biology* 59.3.581–607.
- WARNOW, TANDY. 1997. Mathematical approaches to comparative linguistics. *Proceedings of the National Academy of Science* 94.6585–90.
- WARNOW, TANDY; STEVEN N. EVANS; DON RINGE; and LUAY NAKHLEH. 2005. A stochastic model of language evolution that incorporates homoplasy and borrowing. *Phylogenetic methods and the prehistory of languages*, ed. by James Clackson, 75–87. Cambridge: Cambridge University Press.
- WARNOW, TANDY; DON RINGE; and ANN TAYLOR. 1995. *Reconstructing the evolutionary history of natural languages*. (Institute for Research in Cognitive Science report 95-1b.) Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania.
- WATKINS, CALVERT. 1976. Towards proto-Indo-European syntax: Problems and pseudo-problems. *Chicago Linguistic Society* 12.2.305–26.
- WATKINS, CALVERT. 2001. An Indo-European linguistic area and its characteristics: Ancient Anatolia. Areal diffusion as a challenge to the comparative method? In Aikhenvald & Dixon, 44–63.
- WURM, STEPHEN A. 1982. *Papuan languages of Oceania*. (Ars linguistica 7.) Tübingen: Gunter Narr.

WURM, STEPHEN A.; DONALD C. LAYCOCK; CLEMENS L. VOORHOEVE; and THOMAS E. DUTTON. 1975. Papuan linguistic prehistory, and past language migrations in the New Guinea area. *New Guinea area languages and language study* (Pacific linguistics C-38), ed. by Stephen A. Wurm, 935–60. Canberra: Australian National University.

Max Planck Institute for Psycholinguistics
P.O. Box 310
6500 AH Nijmegen
The Netherlands
[michael.dunn@mpi.nl]

[Received 14 June 2007;
accepted 10 March 2008]

[stephen.levinson@mpi.nl]
[evali@ling.su.se]
[ger.reesink@let.ru.nl]
[a.terrell@let.ru.nl]