The Western Malayo-Polynesian Problem

Alexander D. Smith

UNIVERSITY OF HAWAIʻI

Since the 1970s, Austronesian higher-order subgrouping has often included a Western Malayo-Polynesian (WMP) node within the larger Malayo-Polynesian (MP) group. WMP includes all Malayo-Polynesian languages except for those in Central and Eastern Malayo-Polynesian, although today it is generally not accepted as a valid subgroup. This paper critically assesses WMP, using both linguistic and archaeological evidence. In conclusion, it is claimed that WMP is indeed invalid, and that in its place several primary branches of Malayo-Polynesian are recognized (Western Indonesian, Sumatran, Celebic, South Sulawesi, Chamorro, Palauan, and Moken), plus Central-Eastern Malayo-Polynesian and Philippines, which may represent either a collection of several yet unnamed primary branches or an innovation-defined linkage. Additionally, it is argued that the 6,750 Proto-WMP reconstructions in the *Austronesian comparative dictionary* must be reevaluated, where appropriate, as Proto-MP reconstructions.

1. INTRODUCTION.[1] The current view of Austronesian (AN) higher-order subgroups grew out of earlier proposals (Blust 1977, 1983–84, 1993, 1999) and often includes a Western Malayo-Polynesian (WMP) node within the larger Malayo-Polynesian group, which itself includes the languages of the Philippines, western Indonesia, Malaysia, Sulawesi, Chamic, Malagasy, Chamorro, and Palauan. Blust (1977:2) was the first to make this proposal, which is often repeated in more general literature (Tryon 1995; Bellwood 2007; Blust 2013). It is widely understood, however, that despite the implication that WMP languages may have descended from a discrete Proto-Western Malayo-Polynesian (PWMP) language, WMP does not form a legitimate subgroup and that WMP itself may consist of several, yet unnamed, primary branches of Malayo-Polynesian (see, for example, Ross 1994, 2005; Blust 1984–85:56, 2013:31; Adelaar 2005:14–16; Edwards 2015:93; Pawley 2006; Donohue and Grimes 2008). The aim of the current paper is to critically analyze the implications of a WMP model, and to propose a new subgrouping

1. In the fall of 2014, I sat down in Robert Blust's office and said, "I want to talk to you about Western Malayo-Polynesian," and a smile spread across his face. I want to thank him for encouraging me to pursue this topic, even where some of my conclusions differ from his own. I want to thank Hsiu-chuan Liao, who offered insights that dramatically improved an earlier draft of this paper. I also want to thank two anonymous reviewers whose criticisms made this paper much more complete. All conclusions are my own and any mistakes that remain are my responsibility.

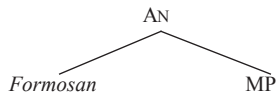
Oceanic Linguistics, Volume 56, no. 2 (December 2017)

that removes WMP and replaces it with several primary branches of MP in an effort to reconcile the linguistic and archaeological realities of the expansion and diversification of Proto-Malayo-Polynesian. Moreover, while many studies not directly concerning higher-order Malayo-Polynesian subgrouping have made similar statements about the invalidity of WMP, this paper differs in that it proposes a model that may replace the WMP model. Specifically, the following observations and proposals are made:

a. The archaeological evidence suggests that speakers of PMP expanded too rapidly over too great a distance for WMP to have developed in a manner that supports it.
b. The linguistic evidence for WMP does not withstand close scrutiny.
c. PMP diversified into at least eight primary branches, giving rise to Proto-Western Indonesian (PWIN), Proto-Sumatran (PSUM), Proto-Celebic (PCEL), Proto-South Sulawesi (PSS), Proto-Central-Eastern Malayo-Polynesian (PCEMP), Moken, Chamorro, and Palauan. Also, the paper acknowledges that the languages of the Philippines may represent more than one primary branch of MP, but the exact internal subgrouping of these languages demands further research. This new internal subgrouping has wide-ranging implications; most importantly, the over 6,750 PWMP reconstructions in Blust and Trussel (ongoing) must be reevaluated, where appropriate, as PMP reconstructions.

2. THE WESTERN MALAYO-POLYNESIAN HYPOTHESIS. The current and most widely accepted AN family tree model is the product of decades of research. The "Out-of-Taiwan" hypothesis (Blust 1984–85, Bellwood 1984–85) was the most ground-breaking and was based on earlier research (Blust 1977) that placed a number of AN primary subgroups in Taiwan, with a single primary group, Malayo-Polynesian, outside of Taiwan. This suggests that Austronesian speakers first settled Taiwan, where the language developed and evolved, and a single group of people moved south, into the northern Philippines and then later into the remainder of Island Southeast Asia, Madagascar, and the Pacific. In the Austronesian family tree in figure 1, Formosan stands for nine separate primary branches of AN found in Taiwan, after Blust (1999). Note that it is common practice to use italics on a family tree to collapse multiple primary branches into a single branch for representational convenience.

FIGURE 1. HIGHER-ORDER AUSTRONESIAN SUBGROUPING

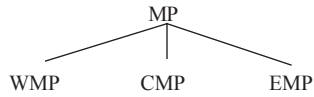


The ultimate correctness of this hypothesis is not being questioned here. Although some have proposed significantly different accounts of AN higher-order subgrouping (Dyen 1962,1965; Starosta 1995; Reid 1982; Wolff 1995; Ross 2009; Sagart 2004, 2013), the general consensus is that the area of highest diversity is Taiwan, and that there is a MP branch which includes all AN languages outside Taiwan, plus Yami.

To return to the current discussion, Blust (1977) first proposed a separation of what he called "Western Malayo-Polynesian" from other Malayo-Polynesian languages,

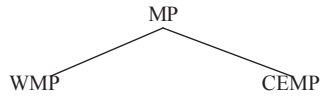
although it appears only on a tree representation, and no justification for the separation is given, as there are no phonological or lexical innovations that offer supporting evidence for WMP. In this early publication, Blust proposed a MP tree with three primary divisions, as shown in figure 2 (EMP = Eastern Malayo-Polynesian, CMP = Central Malayo-Polynesian).

FIGURE 2. HIGHER-ORDER MP SUBGROUPING AFTER BLUST (1977)



Further publications (Blust 1982, 1983–84, 1993, 2009a) have argued for a Central-Eastern Malayo-Polynesian subgroup. Thus, the MP family tree model now suggests a binary split between the languages that stayed behind in the west, and those that moved further east, as shown in figure 3.

FIGURE 3. CURRENT HIGHER-ORDER MP SUBGROUPING AFTER BLUST (1983–84, 1993, 1999)



2.1 THE VALIDITY OF WMP. Although Blust's model might be considered the standard, there is considerable doubt as to the validity of WMP as a subgroup. Ross (2005:4–7) provides a representative example of this doubt, and states that "the literature often refers to a discrete 'Proto-Western Malayo-Polynesian', but there is no evidence ... that such a language ever existed." This statement was made, however, in a paper concerning the subgrouping of Bashiic languages, and, as such, Ross does not further specify what might take the place of WMP, instead stating that PMP probably developed into 20–25 unspecified primary branches. Ross (1994) had earlier attempted to list 24 specific groups into which WMP may be split, but overall, the issue is still in need of specific attention.

Various statements of similar type are found throughout the literature, where authors express doubt in the validity of WMP in papers not directly concerned with MP higher-order subgrouping and without offering an explicit alternative. For example, Blust (1984–85:56, 2013:741, 2014:313) states multiple times that WMP may ultimately consist of more than one primary branch of MP. Other authors have made similar claims, but enumerating each example here is not necessary for the point being made.

There are also numerous competing subgrouping hypotheses that dismantle WMP within larger arguments for a more radically altered family tree. Examples are Reid (1982), Dyen (1962, 1965, 1995), Sagart (2004, 2013), and Wolff (1995). These studies, which are primarily concerned with proposing completely different subgroupings of AN languages, are not considered crucial to the current argument, as this work seeks to alter a specific node within Blust's 1977 model, not to replace the whole thing.

Further, there is no consensus regarding the internal classification of Western Malayo-Polynesian languages, as most scholars do not consider it a valid subgroup. A great number of these languages are morphologically complex, retaining features of PMP phonology, verbal morphology, case, and voice alternations. Others, however, have changed dramatically, from Palauan in Micronesia (Blust 2009b; McManus 1977), to Chamic on the Southeast Asian mainland (Thurgood 1999), to Merap and the Segai-Modang languages in central Borneo (Smith 2017a,b). Several WMP languages or language groups have been claimed to represent primary branches of MP on their own, including Enggano (Edwards 2015), Nasal (Anderbeck and Aprilani 2013), Chamorro (Blust 2000; Reid 2002), and Batanic (Ross 2005). Geographically, the WMP languages are separated into a Philippine group, a Bornean group, a western Indonesian group (essentially the Greater Sunda islands other than Borneo, also including peninsular Malaysia), a Sulawesi group, Chamic, Malagasy, Chamorro, and Palauan. This is, of course, not a linguistically defined hypothesis, but there is some evidence that major geographic boundaries have played a role in the migration patterns of early Austronesian settlers and, as a result, in the genetic affiliation of WMP languages (Blust 2010:47–48).

2.2 HOMORGANIC NASAL SUBSTITUTION AS WMP EVIDENCE. Homorganic nasal substitution (HNS) is a morphophonological phenomenon whereby the final nasal of a prefix, often reflexes of *maŋ-, assimilates to and replaces an initial obstruent in the base to which it attaches. Thus, in Malay, *tulis* 'writing' when affixed with *məŋ-* becomes *mənulis* 'to write', where the cluster *...ŋ-t...* becomes *...n...* and the morphological boundary becomes ambiguous. Blust (1995:633) notes that homorganic nasal substitution is found in "most languages of the Philippines and western Indonesia … and in Palauan and Chamorro." Blust (2000:104) notes that Western Malayo-Polynesian is "a group based largely on the prevalence of homorganic nasal substitution as a functioning process in the formation of active verbs" (although he goes on to state that this claim may in fact be false), and Blust (2013:741) further states that "WMP languages in general … uniquely share the process of homorganic nasal substitution in active verb forms." These statements, along with personal communication with Blust, imply that homorganic nasal substitution may serve as evidence for WMP, although no firm position has ever been taken. Blust (1995, 2000, 2013, pers. comm.) only makes note of its conspicuous distribution, without systematically defending it as subgrouping evidence.

However, even if one were to try to use HNS as evidence for WMP, its presence in CEMP languages in a fossilized form causes issues. Consider the two forms below. PAN *Cakaw 'to steal' and *kuCu 'louse' are both found in CEMP languages with what appear to be fossilized cases of HNS (POC = Proto-Oceanic; data are from Blust and Trussel ongoing).

(1) PAN *Cakaw 'to steal'
PMP *takaw
*paŋ-takaw (*panakaw) 'thief, steal'
*maŋ-takaw (*manakaw)

POc *tako
*nako
*panako 'thief'

(2) PAn *kuCu 'head louse'

PMP *kutu 'head louse'
*maŋ-kutu (*maŋutu) 'to delouse'

POc *kutu
*ŋutu 'louse; to delouse'

Evidence for these reconstructed forms can be found in several CEMP languages, including Ngadha and Li'o *naka* 'steal', Hawu *naʔo* 'steal', Rotinese *nako* 'steal', and Tetun *naʔo-k* 'steal'. Forms with fossilized reflexes of *paŋ-takaw are found in Manam *anako* 'thief', Bauro *hanago* 'steal', Nakanamanga *panako* 'steal', and Rotuman *hanao* 'steal'. Evidence for fossilized HNS in forms reflecting *kuCu can also be found in several CEMP languages, including Tigak *ŋut* 'louse', Mendak *ŋut* 'kill lice by biting', and Ghari and Talise *ŋutu* 'louse'. I am not the first to point out these words, however. In both cases, forms reflecting *Cakaw and *kuCu with what appear to be HNS are listed in Blust and Trussel (ongoing) as doublets. Thus, according to the *Austronesian comparative dictionary*, both *tako and *nako must be reconstructed to POc with near identical meanings. Similarly, *kutu and *ŋutu are reconstructed to POc, again, with near identical meanings. It may be argued, however, that these examples are not doublets, but rather reflect HNS with further deletion of initial *ma-, as numerous languages in western Indonesia and Malaysia have independently reduced the HNS prefix*maŋ- to only the nasal component *ŋ-*. Examples include Kadorih (Kalimantan) *nahkow* and Kayan (Data Dian dialect, Kalimantan) *nakaw*. It thus seems within reason to assume that *nakaw and *ŋutu show signs of the same change and that PCEMP retained at least a small number of fossilized words with HNS.

Although the above proposal removes HNS from consideration as evidence for a WMP subgroup, HNS itself remains a uniquely Malayo-Polynesian phenomenon. This adds another piece of supporting evidence for the validity of Malayo-Polynesian, because HNS has not been found on the island of Taiwan. The simplest explanation is that HNS was innovated once in PMP.

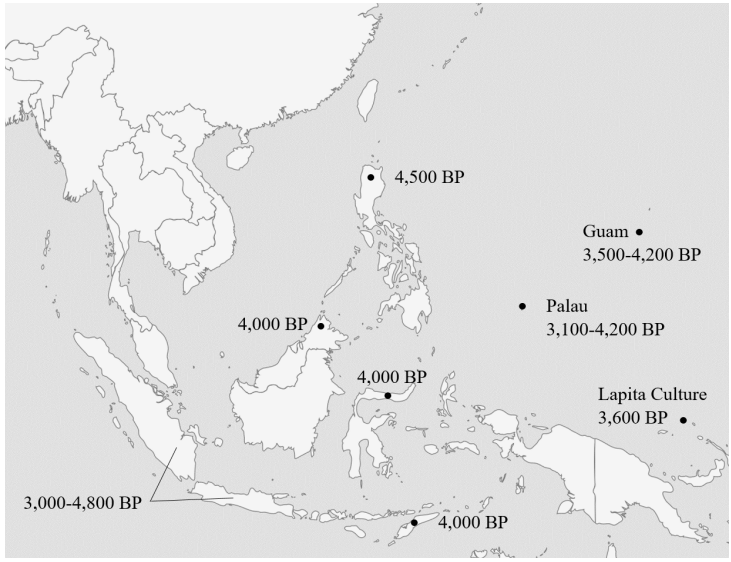
3. THE ARCHAEOLOGICAL HISTORY OF ISLAND SOUTHEAST ASIA AND IMPLICATIONS FOR WMP. Proto-Austronesian was spoken in Taiwan. This can be inferred from the family tree, which places the area of greatest linguistic diversity in Taiwan (Blust 1984–85, 1999). PMP was likely spoken in the northern Philippines. Archaeological evidence indicates that the Philippines were first settled by agriculturalists around 4,500 BP (Bellwood 2007). Borneo, Java, Sumatra, Sulawesi, Timor, Halmahera, the Marianas (Guam), and Palau were settled between 4,000 and 3,500 BP (Bellwood 2007; Kirch 2002; Ward, Athens, and Hotton 1998). This suggests a short period of in-situ development somewhere in the northern Philippines before PMP speakers expanded southward. After the initial southward expansion out of the northern

Philippines, it seems that much of Island Southeast Asia was settled nearly instantaneously (from a historical point of view).

In figure 4, specific dates of the earliest evidence for Austronesian Neolithic culture are plotted on a map of Island Southeast Asia. Archaeological sites in Sabah, northern Borneo, show evidence of Neolithic culture as early as 4,000 BP (Bellwood 1988, 1989). Archaeological material evidence for the settlement of Guam and Palau place Austronesian arrival at later dates, but more recent analysis of soil samples suggests that the forests of Guam and Palau were being cleared through large scale burning as early as 4,300 BP in Guam and between 3,100 and 4,200 BP in Palau (Kirch 2002, Ward, Athens, and Hotton 1998). Radiocarbon dates of pottery shards from Sulawesi yield a wider range of dates. At Ulu Leang in Southern Sulawesi, Glover (1976) reports dates as early as 4,500 BP, while Bulbeck (1992) dates pottery from this site at 3,500 BP (chapter 7 of Bellwood 2007 contains a thorough overview). Also, in locations as far west as Sumatra and Java, pollen core samples yield evidence of forest clearing as early as 4,500–4,800 BP, although permanent, large scale clearing does not appear until closer to 3,000 BP (Bellwood 2007; Flenley 1988; Stuijts 1993). Note, however, that these samples were taken from interior locations at high altitudes, which leaves open the possibility that permanent large scale forest clearings were indeed occurring along the coast closer to 4,500 BP than in the interior. Finally, evidence for the arrival of Neolithic culture in modern Timor has been dated to as early as 4,000–4,500 BP (Glover 1977a,b).

The presence of Neolithic culture in Borneo, the Philippines, Sulawesi, Guam, Palau, and the Lesser Sunda Islands is straightforwardlylinked with the arrival of Austronesian-

FIGURE 4. EARLIEST EVIDENCE FOR NEOLITHIC AUSTRONESIAN ARRIVAL THROUGHOUT SOUTHEAST ASIA AND THE PACIFIC



speaking peoples, as these areas were either uninhabited (Guam, Palau) or were inhabited by people without agriculture (Borneo, Philippines, Sulawesi). Sumatra, however, is close enough to Mainland Southeast Asia that, although it remains unlikely, one may conceive of a scenario where Austroasiatic agriculturalists preceded Austronesians, although there is no direct evidence for this and it remains speculative. One must then be careful when associating dates in Sumatra with Austronesian arrival.

3.1 HOW HISTORICAL POPULATION MOVEMENT INTERACTS WITH THE FAMILY TREE MODEL. Linguistically, when a group of languages are sisters on a family tree, the implication is that speakers of a single language split into two (or more) groups simultaneously and then developed separately. When a group of languages are cousins on a family tree, the implication is that speakers of a single language developed and diversified in situ, then a single group from within the community split off, became isolated, and diversified further. Figures 5 through 8 offer visualizations of how population movement implies certain genetic relationships among languages.

In figure 5, the horizontal line represents the movement of a single group of people who speak the same language, and the small off-shoots represent settlers who stay in each location. The resulting family tree is a rake model, as shown in figure 6. The reason for this logical relationship is that, according to the scenario in figure 5 speakers at location A, B, C, and D speak the same language at the time of settlement. Thus, the languages evolve from the same protolanguage, which is formalized in the rake model.

FIGURE 5. SETTLEMENT PATTERNS ASSOCIATED WITH RAPID POPULATION EXPANSION

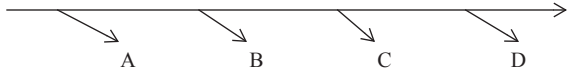


FIGURE 6. RAKE FAMILY TREE ASSOCIATED WITH RAPID POPULATION EXPANSION

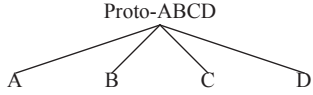
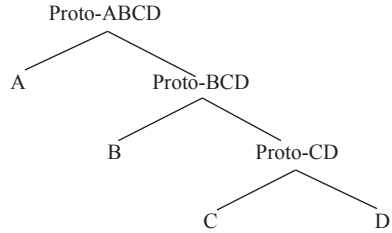


Figure 7 shows a different scenario, with a slower moving population and a series of locations. At each location, the entire population settles, and sometime later, a particular group from within the settled population moves somewhere else. The result of such a population movement would be a tree with nested subgroups, with clearly defined internal relationships like that shown in figure 8.

Of course, this is all fairly simplified, and does not include alternative models for the diversification of languages (like those discussed in François 2014), but the principles are generally supported by archaeological and linguistic evidence. Thus, given an archaeological record that suggests rapid population movement over a large area, historical linguistic evidence should support a rake model over a nested model. Similarly, where linguistic evidence supports a rake model, archaeological evidence should show signs of

FIGURE 7. SETTLEMENT PATTERNS ASSOCIATED WITH SLOW POPULATION EXPANSION AND INTERMITTENT PERIODS OF IN SITU DEVELOPMENT

FIGURE 8. NESTED FAMILY TREE MODEL ASSOCIATED WITH SLOWER POPULATION EXPANSION



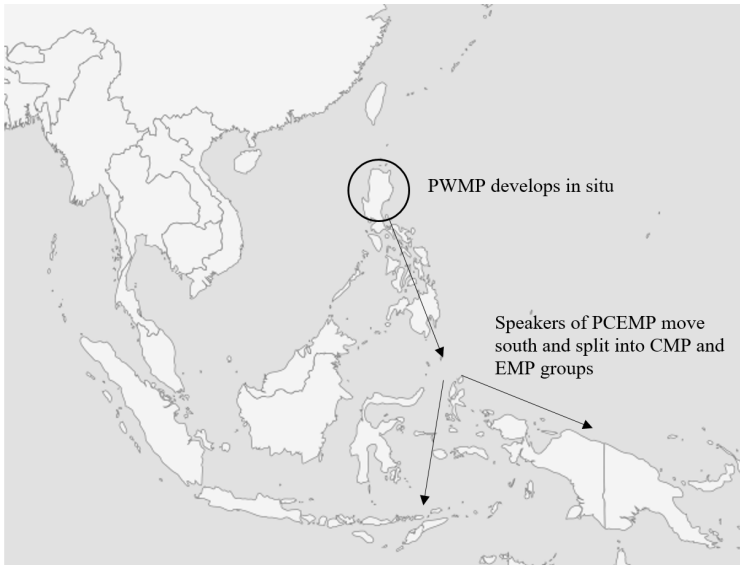
rapid population movement. Where these expectations fail, we must then search for an explanation as to why. In Island Southeast Asia, we find a mismatch between the AN family tree and archaeological evidence. Figure 4 suggests that much of Island Southeast Asia was settled by AN speaking bearers of Neolithic culture at the same time. This, in turn, suggests rapid population expansion of the type in figure 5. The AN family tree with a WMP node, however, is of the type associated with slow expansion and intermittent periods of in situ development.

3.2 WESTERN AND CENTRAL-EASTERN MALAYO-POLYNESIAN AND MIGRATION THEORY. An Austronesian family tree that includes a binary division between WMP and CEMP makes an implicit prediction about the migration history of Austronesian speaking peoples in Island Southeast Asia. Ignoring the archaeological evidence, it suggests that a group of speakers from an ancestral PMP-speaking community left their homeland and moved into the Lesser Sunda islands and further east into the Pacific. Those who stayed behind must have experienced a period of in-situ development, long enough for PWMP to differentiate itself from PMP. Only later could this group of now WMP speaking peoples expand into the areas that they currently occupy. This is the only plausible scenario that I am aware of that explains how the binary branching model of MP might have come to be, as WMP is currently spread over too vast an area for it to have developed as a unit over the entirety of its distribution. Figure 9 gives a visualization of how this scenario might have played out.

If figure 9 were true, then it forces one to further postulate a second expansion out of the northern Philippines that is responsible for the settling of Borneo, Sulawesi, Palau, and the Marianas. This is a highly unlikely scenario, as it implies that when CEMP speakers broke away from PMP and moved into eastern Indonesia, they did not settle the large landmasses of Borneo or Sulawesi, but rather went straight to eastern Indonesia.

Only at a later date could a group of WMP speakers go through essentially the same scenario, expanding into the remainder of Island Southeast Asia. Such a scenario ignores the principle of simplest moves. Also, the archaeological dates for Island Southeast Asia (figure 4) do not agree with figure 9. The linguistic model and the archaeological facts are, thus, at an impasse.

FIGURE 9. MIGRATION HISTORY PREDICTED BY THE WMP FAMILY TREE MODEL



4. THE INTERNAL SUBGROUPING OF MALAYO-POLYNESIAN. I have attempted to show that the WMP hypothesis is flawed in two major respects. First, the binary tree model of MP implies a specific migration history that is irreconcilable with the archaeological facts. While this is not necessarily damning evidence (archaeology and linguistics often disagree, for various reasons), it does cast doubts on the current linguistic model. Second, HNS, a phenomenon that might be viewed as the only innovation linking WMP languages together, is better interpreted as a retention from PMP, as evidenced by fossilized reflexes of HNS in CEMP. In the following sections, attempts are made to organize former WMP languages into linguistically defined primary branches of MP. Eight branches are defended, plus languages of the Philippines that possibly include more than one primary branch of MP, with much of the evidence and arguments drawn from established literature. These groups are:

1. Western Indonesian. A group that consists of all indigenous languages of Borneo, plus the Austronesian languages of Sumatra (excluding Batak, Barrier Islands languages, and Nasal), Javanese, Madurese, Balinese, Sasak, and Sumbawa. The hypothesis of a WIN subgroup was first proposed in Blust (2010) and expanded

upon in Smith (2017a), but WIN as defined here is revised from these earlier proposals that included Batak, Barrier Islands languages, Nasal, and Moken.

2. Sumatran. A group consisting of the Sumatran Barrier Islands languages (Enggano, Nias, Mentawai, Sixule, and Simeulue), Batak languages, and Nasal, spoken in southern Sumatra. This proposal is similar to an earlier proposal by Nothofer (1986), with the addition of Nasal.
3. Celebic. The languages of Sulawesi, excluding the Philippine languages in the north and the South Sulawesi group in southwest Sulawesi. This group is phonologically defined in Mead (2003).
4. South Sulawesi. The languages of southwest Sulawesi, including the major regional languages Buginese and Makasarese. South Sulawesi includes the Tamanic languages of Borneo, whose speakers likely migrated from Sulawesi (Hudson 1978; Adelaar 1994, 1995).
5. Central-Eastern Malayo-Polynesian. All AN languages of the CMP and EMP groups, which includes Oceanic languages, the languages of the Bird's Head Peninsula, Halmahera, and the Lesser Sunda Islands as defended in Blust (1983–84, 1993, 2009a, 2012).
6. Chamorro of the Mariana Islands.
7. Palauan of the Republic of Palau.
8. Moken, spoken along the western coast of Thailand and Burma.
9. Philippines. Not a subgroup as much as a loosely related group of languages that may contain multiple primary branches. Traditionally, the Philippine subgroup consisted of all AN languages of the Philippines, excluding Sama-Bajaw but including Yami in Taiwan and the Philippine languages of northern Sulawesi (Blust 1991, 2005; Zorc 1986).

4.1 WESTERN INDONESIAN. In a section of his paper on the Greater North Borneo (GNB) hypothesis titled "Wider horizons," Blust (2010:91–96) first proposed a Western Indonesian subgroup that includes "all languages of western Indonesia outside of Sulawesi." The evidence for such a group is entirely lexical, and is restricted in most part to terminology that was innovated by PWIN speakers to name formerly unknown plants and animals. Also, the original Western Indonesian hypothesis has such wide scope that it includes some languages (Sumatran Barrier Islands languages, Nasal, Moken, and Batak) for which none of the proposed lexical innovations have been located. These will be discussed at greater length below. The original list contained only 12 innovations, and as a result the PWIN hypothesis has not yet gained wide acceptance. Additional research in Smith (2017a) pushed the number of PWIN lexical innovations to 34, including additional cases where previously unknown animals were apparently named by the very first AN settlers in Borneo. The fact that lexical evidence for PWIN is mostly confined to such terms is significant, as Blust (1982) demonstrated the value of major floral and faunal boundaries for linguistic subgrouping. The reasoning is that if Borneo was settled by multiple groups, then it should be impossible to reconstruct a single term for previously unknown animals to a single protolanguage. If, however, Borneo was settled by a single group of PWIN speakers, then it logically follows that any new

plants and animals that they encountered would have been named once, and that name would then have been inherited by all daughter languages.

This argument is only valid where a concerted effort has been made to eliminate borrowing as an explanation for such innovative terms. For PWIN, such efforts have been made, and the 34 PWIN lexical innovations (listed below) are well supported overall. Data are from Blust (2010) and Smith (2017a). Evidence that is not present in Blust (2010) is expanded upon below, with every piece of evidence supporting the reconstruction of newly discovered WIN innovations reprinted from Smith (2017a).

(3)

*əluŋ 'river mouth; estuary'	*puRaʔ 'crab'
*jaʔa 'chin; jaw'	*buRis 'silver-leaf monkey'
*suŋay 'small river'	*kəniw 'orangutan'
*tupay 'tree shrew'	*ukəd 'western tarsier'
*pinaŋ 'betel nut'	*kuliR 'clouded leopard'
*kubuŋ 'flying lemur'	*təlaʔus 'barking deer'
PMP*sawa/ PWIN*pəŋanən 'python'	*kəniw 'eagle'
*tuqan/*RimbaʔRimbaʔ 'primary forest'	*ma-tuRun 'binturong'
*biRuaŋ 'sun bear'	*giRam 'river rapids'
*kəlabət 'gibbon'	*iban 'parent-in-law/child-in-law'
*bəduk/*bəRuk 'pig-tailed macaque'	*gətəm 'to harvest'
*dəŋən 'river otter'	*likaw/*liŋkaw 'eyebrow'
*pəlanuk 'mouse deer'	*tiliŋ/*siliŋ 'to fly'
*kəjut 'surprised; startled'	*madam 'rotten'
*duRian 'durian'	*bə-təRiʔ 'pregnant'
*butbut 'coucal'	*qulun 'outsider'
*tiuŋ/*kiuŋ 'myna bird'	*kəlasi 'red-leaf monkey'

(4) ***kəlasi 'red-leaf monkey'**

Malayic	Ketapang *kəlasi*, Keninjal *kəlasi*, Seberuang *kəlasəy*
Land Dayak	Ribun *kəlasi*
Barito	Ngaju *kəlasi*, Maanyan *kəlahi*
Basap	Lebo *kəlasi*
Kayanic	Ngorek *kəlasi*, Busang *haseʔ*, Long Gelat *kənsaę*, Gaai *kalsay*, Kelai *kəlsay*
Punan	Punan Bah *kələlasey*
Kenyah	E Penan *kəlasi*, W Penan *kəlasi*, Lebo' Vo' *kəlasi*, Lepo Gah *kəlasi*, Lepo Laang *kəlasi*, Lepo Sawa *kəlasi*, Badeng *kəlasi*
Dayic	Long Lellang *kəlasih*

The red-leaf monkey is physically similar to the silver-leaf monkey, but has a maroon colored coat. The word *kəlasi is widely distributed in languages where primary data are available. Lobel (2016) does not contain 'red-leaf monkey', but even without Sabahan evidence, it can be reconstructed to Western Indonesian. The only Land Dayak witness, Ribun *kəlasi*, is an apparent Malayic borrowing as *l failed to become *h*.

(5) ***puRaʔ 'crab'**

Basap	Lebo *puraʔ*
Barito	Tunjung *puhaʔ*, Benuaq *puya*, Taboyan *puya*, Maanyan *puya*, Dusun Witu *puya*

Dusunic Kujau *puha?*, Minokok *pua?*, Dusun Talan *pua?*, Dumpas *pua?*
Paitanic Sungai Beluran *pua?*, Lingkabau *pua?*, Lobu *pua?*, Kuamut *pua?*, Serudong *pua?*

(6) ***buRis 'silver-leaf monkey'**
Barito Kadorih *buhih*, Ngaju *buhis*, Maanyan *buhis*, Dusun Witu *buhis*, Bentian *buis*, Taboyan *buis*
Kayanic Balui Liko *bui*, Busang *bui*, Long Gelat *uheh*
Kajang Sekapan *bəgʷi*, Kejaman *bəgʷi*, Lahanan *bəgʷi*
Punan Punan Tubu *oih*, Punan Bah *bui*, Ukit *bui*, Buket *bui*

Two widely recognized species of black-coated long-tailed monkeys are known in Borneo, the long-tailed macaque and the silver-leaf monkey. The major difference between the two is behavior; they do not occupy overlapping territories, as the silver-leaf monkey is found in middle canopies, while the long-tailed macaque is more often found near human habitats. Physically, macaques are gray with hairless faces while silver-leaf monkeys are black or dark in color with dark faces.

(7) ***kəniw 'eagle'**
Barito Paser *koniw*
Melanau Sarikei *kəñiw,* Kanowit *kəñiu*
Punan Ukit *koñu*, Buket *koñu*
Müller-Schwaner Hovongan *koñu*, Kereho *koñu*, Seputan *koñʉ*, Aoheng *koñʉ*
Dayic Long Semadoh *kanʊy*, Long Bawan *kəniw*
B-LB[2] Narum *hañʊy*
SW Sabah Brunei Dusun *kaniw*, Burusu *kanʊy*, Rungus Dusun *koniw*, Kujau *kəniw*, Dumpas *koniw*, Serudong *kaniw*, Tatana *kandiu*, Papar *kandʊy*, Gana *kəniw*, Tidung *kanʊy*
NE Sabah Idaan *kənnʊy*, Seguliud *konnʊy*

Blust and Trussel (ongoing) list Idaan *kənnʊy*, Abai Sembuak *kanʊy*, Bisaya Limbang *kanʊy*, Tagol Murut *kandʊy*, and Lun Dayeh (Long Semadoh) *kənʊy* as near cognate with Tagalog *banoy*, but these forms are only similar because of sporadic changes of *-iw to *-uy, a common sporadic change in MP.

(8) ***ukəd 'western tarsier'**
Land Dayak Hliboi *koa̲d*, Sungkung *əʰkəd*, Golik *ŋkət*, Jangkang *ŋkut*
Barito Taboyan *ukər*, Benuaq *ukar*, Tunjung *ukər*, Bentian *ukər*
Melanau Dalat *ukəd*, Kanowit *ukət*
Kajang Sekapan *kət*, Kejaman *kət*, Lahanan *kət*
Punan Tubu *iket*, Punan Bah *iket*, Beketan *iket*, Lisum *iket*, Ukit *ikət*
Kenyah W Penan *kət*, Lebo' Vo' *kət*, Lepo Gah *kət*, Lepo Tau *kət*

The western tarsier (*Cephalopachus bancanus*) is found only on Borneo, southern Sumatra, and adjacent off-shore islands. It is not found in the Philippines or Taiwan and, therefore, was initially encountered by Austronesian speaking peoples upon first arrival

2. B-LB = Berawan-Lower Baram.

in northern Borneo. There is no Sabahan evidence for *ukəd, but this is due to a lack of overall data.

(9) ***təlaʔus 'barking deer'**

Barito	Kadorih *tarouyh*, Taboyan *təlawus*, Paser *təlaus*, Benuaq *təlaus*, Tunjung *təlauh*
Basap	Tabalar *təlaʔos*, Segai *təlaus*, Batu Putih *təlʔus*
Kayanic	Ngorek *təlaʔoh*, Merap *klaʔawh*, Long Naah *təlaʔuh*, Data Dian *təlaʔuh*, Balui Liko *təlaʔu*, Busang *təluʔu*, Bahau *təlʔoh*, Long Gelat *uh*, Modang *oa̯h*, Kelai *us*
Kajang	Sekapan *təlaʔu*, Kejaman *təlaʔu*, Lahanan *təlaʔu*
Punan	Tubu *təlauʔ*, Punan Bah *təlouʔ*, Beketan *təlahuʔ*, Lisum *təlahuʔ*, Punan Aput *təlauʔ,* Ukit *təlahuʔ*, Buket *təlahuʔ*
Müller-Schwaner	Kereho *taaʔu*, Seputan *toraʔu*, Aoheng *tərəʔu*
Kenyah	Sebop *təlaʔo*, W Penan *təlau*, Lebo' Vo' *təlaʔo*, Lepo Gah *təlaʔo*, Lepo Sawa *təlaʔo*, Lepo Tau *təlaʔo*, Badeng *təlaʔo*
Dayic	Long Semadoh *təlao*, Long Bawan *təlau*, Bario Kelabit *təlaʔo*
B-LB	Miri *təlaʔaw*, Narum *təlaʔaw*, Kiput *təlaaw*, Long Jegan *təlaʔo*, Long Terawan *təlao*

The barking deer (*Muntiacus*) has a wide distribution in Southeast Asia. PAN *sakəC, which is reconstructed using only Formosan evidence, supports the inference that the barking deer was known to PAN speaking peoples. The deer is not, however, found in the Philippines, and speakers of PMP apparently lost the PAN word before reencountering the barking deer in Borneo.

(10) ***kuliR 'clouded leopard'**

Barito	Ngaju *kulih*, Kapuas *kuleh*, Maanyan *kuli*, Dusun Witu *kuli*, Taboyan *kuli*, Paser *kuli*, Benuaq *kuli*, Tunjung *kuleh*
Kayanic	Ngorek *koleh*, Merap *kluyh*, Long Naah *kuleh*, Data Dian *kuleh*, Balui Liko *kuleh*, Busang *kuleh*, Bahau *kuleh*, Long Gelat *kəleh*, Modang *kəlih*, Gaai *kleh*, Kelai *kleh*
Melanau	Kanowit *kuli*
Kajang	Sekapan *kuɮi*, Kejaman *kuɮi*, Lahanan *kuɮi*
Punan	Tubu *kuli*, Punan Bah *kuli*, Beketan *kuɭi*, Lisum *kuɭi*, Punan Aput *kuɭi*, Ukit *kuli*, Buket *kuli*
Müller-Schwaner	Hovongan *kuri*, Kereho *kuri*, Seputan *kuri*, Aoheng *kuri*
Kenyah	W Penan *kuli*, Lebo' Vo' *kule*, Uma Pawe *kule*, Lepo Gah *kule*, Lepo Laang *kule*, Lepo Sawa *kule*, Lepo Tau *kule*, Badeng *kole*
Dayic	Pa' Dalih *kuir*, Long Bawan *kuir*,
B-LB	Kiput *kulay*

The clouded leopard (*Neofelis*), like the barking deer, has a wide distribution in Southeast Asia and was known to PAN speaking peoples (PAN *lukəNaw). It, too, is not found in the Philippines, and the PAN word was lost by PMP speakers before they moved into northern Borneo.

(11) ***kəRiw 'orangutan'**

Dusunic	Rungus Dusun *kogiw*, Papar *kogiw*, Kujau *kəgiw*, Minokok *kəgiw*, Dusun Talan *kogiw*, Dumpas *kogiw*
Paitanic	Sungai Beluran *kəgiw*, Lingkabau *kəgiw*, Kuamut *kagiw*
Murutic	Tatana *kagɯy*, Nabaay *kagɯy/kagiw*, Gana *kəgiw*, Tagol Murut *kahɯy*, Bulusu *agɯy*
Barito	Kereho *kahiuʔ*, Ngaju *kahiuʔ*, Maanyan *keuʔ*, Dusun *keuʔ*
Basap	Lebo *kəriu*
Kayanic	Modang *kahjo*, Gaai *kahjeə̯w*, Kelai *kahjaw*

The orangutan is today found only in the forests of Borneo and Sumatra. Although *kəRiw is nearly absent in central Borneo, its presence in SW Sabah, Kayanic, and Barito provides sufficient genetic diversity to reconstruct it to PWIN.

(12) ***pəŋanən 'python'**

Land Dayak	Hliboi *ŋanun*, Sungkung *ŋanən*, Golik *pəŋanin*
Barito	Kadorih *paŋanon*, Ngaju *paŋanən*, Kapuas *pəŋanən*, Maanyan *paŋanen*, Dusun Witu *pəŋanen*, Taboyan *pəŋanən*, Benuaq *pəŋanən*, Tunjung *pəŋanan*
Basap	Lebo *pəŋaʔan*
Kayanic	Ngorek *pəŋanən*, Merap *pəŋanan*, Long Naah *pəŋanən*, Data Dian *pəŋanən*, Balui Liko *pəŋanən*, Bahau *pəŋanan*, Long Gelat *pənnan*, Modang *pənnan*, Kelai *pɛʔ ŋan*
Punan	Punan Tubu *pəŋanen*, Punan Bah *pəŋanən*, Punan Lisum *pəŋanən*, Punan Aput *pəŋanen*, Ukit *pəŋanən*, Buket *pəŋanən*
Müller-Schwaner	Hovongan *poŋanon*, Kereho *poŋanon*, Seputan *poŋanon*, Aoheng *pəŋanon*
Kenyah	Sebop *pəŋanən*, E Penan *pəŋanən*, W Penan *pəŋanən*, Lebo' Vo' *pəŋanən*, Uma Pawe *pəŋanən*, Lepo Gah *pəŋanən*, Lepo Sawa *pəŋanən*, Lepo Tau *pəŋanən*

Although reflexes of *pəŋanən are absent in Sabah, the word is widely attested throughout the rest of Borneo. It is found in Greater North Borneo as well as Barito, which suggests that the word was innovated in WIN. Malayic languages retain PMP *sawa 'python', which means that two words must be reconstructed for python in PWIN. In many communities where primary research was performed, a semantic distinction between two types of python, a long python and a short and stout python, were recorded. This may have been why two names appeared in WIN, but because it is not strictly a replacement innovation, it is not as strong as other lexical evidence.

(13) ***ma-tuRun 'binturong'**

Malayic	Ketapang *bənturun*, Kapuas Iban *turun*,
Barito	Ngaju *tuhun*, Dusun Witu *munin tuyun*, Bentian *munin tuyun biaŋ*
Land Dayak	Sungkung *tuutn*, Ribun *muntuhun*
Kajang	Sekapan *mətun*, Kejaman *mətun*, Lahanan *mətun*

The binturong, or "bear-cat," is well-known to the people of Borneo, and two words with wide distributions are in competition with one another; *kitan and *matuRun. Although *kitan is much more robustly attested in the languages where it is found, *matuRun is the only word for 'binturong' that crosses important subgrouping boundaries. It is found in three major Greater North Borneo subgroups, plus Barito.

(14) ***giRam 'river rapids'**

Land Dayak	Sungkung *giapm*
Barito	Kadorih *kiham*, Maanyan *kiham*, Dusun Witu *kiham*, Tunjung *kehapm*
Basap	Segai Basap *kiham*
Kayanic	Data Dian *giham*, Balui Liko *giham*, Bahau *giam*, Long Gelat *giim*, Kelai *giəm*
Kajang	Sekapan *geam*, Kejaman *giam*, Lahanan *giam*
Punan	Punan Bah *giam*, Beketan *giham*, Punan Lisum *giham*, Punan Aput *giham*, Ukit *giam*, Buket *giam*
Müller-Schwaner	Seputan *keham*, Aoheng *keham*
Murutic	Burusu *giram*

(15) ***Rimbaʔ 'jungle'**

Malayic	Malay *rimba*, Ketapang *rimbo*, Keninjal *ɣimaʔ*, Seberuang *ɣimbaʔ*, Mualang *ɣimaʔ*
Land Dayak	Jangkang *ɣimaʔ*, Ribun *himo*
Barito	Kadorih *himbaʔ*, Ngaju *himbaʔ*
Basap	Lebo *rimaʔ*
Müller-Schwaner	Kereho *hiwaʔ*
Kenyah	Sebop *vaʔ*, Penan *vaʔ*, Lebo' Vo' *vaʔ*, Lepo Tau *mpaʔ*
B-LB	Long Terawan *maʔ*
Murutic	Burusu *limbaʔ*

This word poses an issue for reconstruction, as PMP *tuqan 'primary forest' is retained with a wide distribution in Borneo, but *Rimbaʔ is well represented in a number of languages that span the entirety of the island. It is, thus, necessary to reconstruct two words for primary forest.

(16) ***qulun 'outsider'**

Malayic 'slave'	Ketapang *hulon*, Kapuas Iban *ulun*
Land Dayak 'slave'	Benyadu *ulutn*
Melanau 'slave'	Sarikei *ulun*, Mukah *ulun*
Barito 'person'	Kadorih *urun*, Ngaju *uluh*, Kapuas *uluh*, Maanyan *ulun*, Dusun Witu *ulun*, Taboyan *ulun*, Paser *ulun*, Benuaq *olutn*, Tunjung *ulutn*
Basap 'person'	Lebo *ulun*
Kayanic 'person'	Ngorek *ulun*, Merap *lʊŋ*, Long Naah *kəlunan*, Data Dian *kəlunan*, Balui Liko *kəlunan*, Busang *kəlunan*, Gaai *luə̯n*, Kelai *loə̯n*
Kenyah 'person'	Sebop *kəlunan*, Long Wat *kəlunan*, E Penan *kəlunan*, W Penan *kəlunan*, Lebo' Vo' *kəlunan*, Uma Pawe *kəlunan*, Lepo Gah *kəlunan*, Lepo Laang *kəlunan*, Lepo Tau *kəlunan*

Bintulu ‘person’	*ulun*
Dayic ‘person’	Long Semadoh *ləmulun*, Long Bawan *lun*
B-LB ‘person’	Miri *jəmulon*, Narum *jəmunawn*, Kiput *bulun*, Long Jegan *lamulawŋ*, Long Terawan *ləmulon*
SW Sabah ‘person’	Southern Bisaya *ulun*, Rungus Dusun *ulun*, Kadazan Papar *uhun*, Dusun Talan *tulun*, Dumpas *ulun*
NE Sabah ‘person’	Idaan *ulun*, Seguliud *ulun*, Begak *ulun*

This is one of the most widely attested innovations in Western Indonesian, but it is not included in Blust’s list of WIN lexical innovation because of a single witness outside of Western Indonesian: Ngadha *ulu* ‘person; human being; counting word for children’. Ngadha is spoken on Flores and is no stranger to Malayic influences. Given how widely attested *qulun is in WIN, it seems that a single outlier in Ngadha is best considered a remnant of contact. If more examples are found outside of the WIN area, then this word would have to be reconsidered, but there is additional evidence that suggests that *qulun was a WIN replacement innovation; the distribution of PMP *qaRta ‘outsider’. Reflexes of *qaRta abound in the Philippines, Palau, Sulawesi, and CEMP but are otherwise unattested in languages that fall within Blust’s Western Indonesian subgroup. The simplest interpretation of this is that PMP *qaRta was replaced by *qulun in PWIN and that the only attestation of this word outside of WIN is a borrowing from a Malayic source. The key piece of supporting evidence for this interpretation is again found in Ngadha. Although *ulu* is found in this language, *ata* (< *qaRta) is also attested with a very similar meaning, ‘person, human image, enemy, someone’. Ngadha’s closest relative, Li’o, also reflects *qaRta as *ata* ‘people’ but does not have a reflex of *ulun. All of the available evidence, then, points to PWIN *qulun.

(17) ***iban ‘reciprocal affine; parent-in-law; child-in-law’**

Land Dayak	Benyadu *ibatn* ‘child-in-law’, Bekati *bat* ‘child-in-law’, Jangkang *ibatn* ‘child-in-law’, Ribun *ibatn* ‘child-in-law’, Golik *ibətn* ‘child-in-law’
Barito	Tunjung *ewan* ‘parent-in-law’
Kayanic	Ngorek *eban*, Merap *bi͡ə̃*, Long Naah *divan*, Data Dian *difan*, Busang *divaŋ*, Bahau *difan*
Melanau	Kanowit *iban*
Kajang	Sekapan *ivan*, Kejaman *ivan*, Lahanan *ivan*
Punan	Punan Bah *ivan*
Kayan	E Penan *kivan*, W Penan *kivan*, Lebo’ Vo’ *iban*, Uma Pawe *divan*, Lepo Gah *iban*, Lepo Sawa *iban*, Lepo Tau *iban*
Dayic	Bario Kelabit *iban*, Long Bawan *eban*
B-LB	Kiput *ibin* ‘parent-in-law’
Dusunic	Southern Bisaya *iwan*, Brunei Dusun *iwan*, Limbang Bisaya *yuan*, Rungus Dusun *ivanan* ‘parent-in-law’, Dumpas *iwan* ‘parent-in-law’
Paitanic	Sungai Beluran *iwan* ‘parent-in-law’, Kuamut *iwanon* ‘parent-in-law’
Murutic	Tatana *iwan*, Papar *iwan*, Timugon *iwan*, Tagol Murut *iwan*, Burusu *iwan*

(18) ***gətəm 'harvest'**

Land Dayak	Benyadu *ŋutupm*, Bekati *ŋutup*, Sungkung *ŋaʰtipm*
Barito	Kapuas *məŋətəm*, Benuaq *ŋotapm*, Tunjung *ŋətəpm*
Basap	Batu Putih *aŋətəm*
Müller-Schwaner	Kereho *ŋotom*, Seputan *ŋotom*
Kayanic	Kelai *tam*
B-LB	Narum *gutəm*

(19) ***likaw/*liŋkaw 'brow'**

Barito	Kadorih *likow*, Ngaju *liŋkaw*, Kapuas *liŋkaw*, Bakumpai *liŋkaw*
Basap	Tabalar *ligaw*
Kayanic	Ngorek *leko*, Long Naah *likaw*, Data Dian *likaw*, Balui Liko *liko*, Busang *liko*, Bahau *likaw*
Melanau	Dalat *likaw*, Kanowit *likaw*
Kajang	Sekapan *likaw*, Kejaman *likaw*, Lahanan *likaw*
Punan	Punan Bah *likuow*, Beketan *likow*, Punan Lisum *likow*, Punan Aput *likow*, Buket *liko*
Müller-Schwaner	Hovongan *diko*, Kereho *diko*, Seputan *liku*, Aoheng *liku*
Kenyah	E Penan *likaw*, W Penan *likaw*, Lebo' Vo' *likaw*, Uma Pawe *likaw*, Lepo Gah *likaw*, Lepo Tau *likaw*

(20) ***siliŋ/*tiliŋ 'to fly'**

Barito	Maanyan *samidiŋ*, Dusun Witu *samidiŋ*, Taboyan *mən-siliŋ*
Melanau	Balingian *siliaŋ*, Matu *tileaŋ*, Sarikei *tiliaŋ*, Mukah *siliaŋ*, Dalat *tili*
Müller-Schwaner	Seputan *ñeleŋ*
Basap	Lebo *təmiliŋ*
Kayanic	Long Gelat *məliaŋ*, Gaai *mleaŋ*

The initial consonant is inconsistent even in very closely related languages (within Melanau, for example). The word is, however, well attested in a variety of languages in southern Borneo, plus Basap. Directionality is difficult to determine, and it must have been retained in the languages in question from a more ancient source.

(21) ***madam 'rotten'**

Land Dayak	Benyadu *madap*, Bekati *madam*, Hliboi *medepm*, Sungkung *madepm*, Ribun *modam*
Barito	Kadorih *maram*, Ngaju *maram*,
Melanau	Matu *madam*, Mukah *madam*, Dalat *madam*
Kajang	Lahanan *maram*
Punan	Tubu *məram*, Punan Bah *meram*, Punan Aput *maram*
Müller-Schwaner	Hovongan *maram*, Seputan *maram*
Kayanic	Ngorek *maram*, Merap *marãa*, Modang *məleaŋ*, Gaai *mleaŋ*
Kenyah	Uma Pawe *maram*, Lepo Tau *madam*

(22) ***bə-təRiʔ 'pregnant'**

Land Dayak	Benyadu *batahi*, Bekati, *batahiʔ*, Hliboi *ntiaʔ*, Sungkung *baʔtiaʔ*, Ribun *biti*, Golik *bətiaʔ*, Sanggau *bitti*
Barito	Kadorih *batohiʔ*, Ngaju *batahiʔ*, Kapuas *bətihiʔ*, Tunjung *teheʔ*
Melanau	Dalat *pətaiʔ*, Kanowit *bətehe*ʔ
Kajang	Sekapan *tiiʔ*, Kejaman *taiʔ*, Lahanan *pətaiʔ*
Punan	Punan Bah *bateiʔ*, Beketan *buteʔ*, Punan Aput *bətaiʔ*
Müller-Schwaner	Kereho *botohiʔ*, Aoheng *bətohi*
Lower Baram	Long Terawan *təreʔ*
SW Sabah	Bulungan*bətəri*
NE Sabah	Idaan *təgkiʔ*, Bonggi *togiʔ*

The Long Terawan word is most likely a borrowing, as *R is usually reflected as *k* in intervocalic position. This word is geographically restricted to southern Borneo plus Northeast Sabah, but can be assignable to an ancestor language that covers all of Borneo. Blust (2010:63) lists *təgiʔ as a NE Sabah innovation, with evidence from Bonggi *togi* and Idaan *təgkiʔ*. The larger picture, however, reveals that this word has a more ancient origin, and in this case *g in NE Sabah comes from *R.

The above evidence certainly supports the hypothesis that all languages of Borneo are descended from a common ancestor. Borrowings, which present an issue of special importance in lexically defined subgroups, have been removed from the list, and what remains are innovative words with regular sound correspondences and very robust distribution. The question that remains, however, is how does this impact languages to the west of Borneo, which are sometimes quite different but which Blust (2010) nevertheless includes in WIN? First, the many languages of Sumatra, Mainland Southeast Asia, and western Java that subgroup with Malayic are by default included in WIN, as Malayic is firmly placed in Greater North Borneo, along with Chamic and Sundanese. Rejang, which has an apparently native reflex of *tujuq 'seven' (Rejang *tujuak*, *tojoak*) is also included in Greater North Borneo.[3] In Sumatra, this leaves Batak and Barrier Islands languages (Mentawai, Enggano, Nias, Sixule, Simeulue) unaccounted for, as well as the southern Sumatran languages Nasal and Lampung. To the southeast, all languages that do not subgroup immediately with Malayic (Javanese, Sasak, Madurese, Sumbawa, Balinese), are also "up in the air" as to their ultimate linguistic position. If any of these languages are to be included in Western Indonesian, they will have to exhibit some of the lexical innovations that define it. Otherwise, while Western Indonesian remains legitimate, the number of languages it includes will need to be revised. The remainder of this section is dedicated to determining the linguistic position of these languages. It is argued that the Barrier Islands languages, Batak and Nasal form a subgroup separate from Western Indonesian, and that Moken represents a primary branch of MP on its own. Lan-

3. It is difficult to see how this could be a borrowing, since the only available source, Malay, reflects *q with *h*, while this word reflects *q as *k*. While it remains true that Rejang is quite different from other Greater North Bornean languages, this reflex of *tujuq cannot be brushed aside without assuming multiple parallel innovations.

guages of Java and islands to its east, however, are kept in Blust's Western Indonesian group, pending further investigation.

4.1.1 Java, Bali, and Lombok. Javanese, Madurese, Balinese, Sasak, and Sumbawa all merged *j with *d, so they cannot be immediately removed from Western Indonesian (the reconstructibility of the change *j > *d in WIN is discussed further in 4.1.4). Lexical evidence does suggest that these languages should be included in Western Indonesian, but the evidence itself (see table 1) is not as robust as one might hope. Possible loan-words are marked with "L" in table 1.

TABLE 1. WIN INNOVATIONS PRESENT IN JAVANESE, BALINESE, AND SASAK

PWIN	Javanese	Balinese	Sasak
*suŋay 'small river'	suŋe 'river (literary)'	—	—
*qulun 'outsider'	(h)ulun 'I; servant'	—	—
*butbut 'coucal'	bubut	—	bubut
*kəzut 'startled'	kədut 'muscular twitch'	təŋkəjut 'startled'	—
*biRuaŋ 'Malayan sun bear'	barwaŋ (Old Javanese)	bruaŋ (L?)	bəruaŋ (L?)
*duRian 'durian'	duren (L?)	duren (L?)	duren (L?)
*tikus 'mouse; shrew'	tikus (L?)	tikus (L?)	—

In Javanese, *hulun/ulun*, *suŋe*, and *bubut* are the most important lexical evidence. It is noted that *suŋe* is confined to literary usage, and may ultimately be a Malay loanword, but if native, it can form an argument along with *hulun* and *bubut* that Javanese is indeed part of Western Indonesian. The definition of *bubut* 'a type of owl; coucal' is unambiguous, and it clearly reflects *butbut 'coucal'. The form *hulun*, which means both '1SG' and 'servant', reflects *qulun, one of the most widely attested WIN innovations. Balinese, Sasak, and Sumbawa have significantly less evidence for inclusion in Western Indonesian. Balinese *bruaŋ* 'bear' and *duren* 'durian' could be borrowings, as might Sasak *duren.* Balinese *təŋkəjut* 'startled; shocked' does appear to reflect PWIN *kəjut, which argues for its inclusion. Also Sasak *bubut* 'type of bird' might be a reflex of *butbut 'coucal', but even so, the evidence is scant.

A partial explanation for why there will be less lexical evidence for including these languages in Western Indonesian than one finds on the island of Borneo is the different ecologies of the islands. Much of the Western Indonesian evidence is faunal, based on the fact that many of the animals of Borneo are either not found in Taiwan and the Philippines, or inherited PAN words for large mammals were lost as people moved through the Philippines, and had to be renamed upon reentering the faunal zone of Mainland Southeast Asia. Many of those animals are absent in Java, Bali, Lombok, and Sumbawa, particularly the latter two, which are located across the Wallace Line, implying that large mammals never made it to those islands. In that case, we expect that any inherited words for these animals would be either lost or repurposed.

Blust (2010) included Balinese, Sasak, Sumbawa, Madurese, and Javanese in Western Indonesian partially because of their geographic position, under the premise that the first Austronesian speakers who arrived in this part of Island Southeast Asia came in two separate groups. One followed the northwest coast of Borneo, and the other, responsible

for Barito plus all of the languages mentioned in this section, traveled along the southeast coast. Javanese, and to a lesser extent Balinese and Sasak, do share some apparently native lexical innovations with Western Indonesian, but one must wonder if this limited evidence is enough to justify their inclusion. I take the preliminary position that these languages ultimately belong in Western Indonesian, with an indeterminate relationship to each other and to other WIN languages, but more work is still needed.

4.1.2 Nasal. Anderbeck and Aprilani (2013) provide the most recent work on Nasal, a language spoken by a few thousand individuals in the southern Sumatran province of Bengkulu, surrounded by speakers of Lampung and Malay. No consensus has been reached on Nasal, although Anderbeck and Aprilani posit that Nasal is an isolate within Malayo-Polynesian, with no special linguistic relationship to any other Malayo-Polynesian language.

There are two difficulties one faces when working with Nasal: lack of widely available data, and very high levels of borrowing from both Malay and Lampung. When working with lexical data, which defines Western Indonesian, these problems are only made worse. Although Anderbeck and Aprilani only provide a modified Swadesh list, the *Holle list* (Stokhof 1987) provides some 1,500 forms, although with less precise phonetics. I was able to locate several apparent WIN innovations in these two sources, which are reprinted below:

(23) NASAL

hulon ‘person’	karoh ‘long-tailed macaque’
tupay ‘squirrel’	rimbo ‘primary jungle’
tikus ‘mouse’	durian ‘durian’
bəruk ‘pig-tailed macaque’	lamin ‘attic’

With our limited understanding of Nasal historical phonology, several of these apparent WIN words can be thrown out as borrowings. First, because *R became Nasal *l*, *karoh*, *rimbo*, *durian* can be removed (from *kəRaʔ, *Rimbaʔ, *duRian); while *lamin* ‘attic’, is almost certainly from Malay, because *lamin* is a Greater North Borneo innovation. Greater North Borneo, among other things, innovated *tuzuq ‘seven’, which replaced PAN *pitu. Nasal reflects *pitu as *pitu*, and, thus, cannot be a GNB language. The other words are less obvious. Nasal *tupay* and *tikus* may ultimately be Malay borrowings, but they lack diagnostic phonemes that may serve to identify them as such. There are two possible sources for *beruk*: in PWIN, two words for ‘pig-tailed macaque’ can be reconstructed, *bəRuk (Malay *bəruk*, Tidung *gabok* with metathesis) and *bəduk (Bintulu *bəduk*, Kelabit *bədtuk*, Kayan *bəruk*). Nasal *bəruk* [bəxuk] may reflect either of these, *bəRuk via Malay, or *bəduk through inheritance, but given the sheer number of borrowings, it is best considered a loan. The final word, *hulon* ‘person; human’ typically refers to ‘slave; servant’ in Malay, where it means ‘person’ in Nasal.[4] If *hulon* is considered native, then it argues for including Nasal in Western Indonesian. However, a single word hardly forms a strong argument, especially because it contains no diagnostic pho-

4. This semantic discrepancy may not be so important, as *qulun ‘outsider’ is reflected as ‘person’ throughout Western Indonesian languages. If Malay is the source for *hulun,* it may have independently shifted to ‘person’, or it may have been borrowed from a source (Malayic or not) that had independently shifted *qulun to ‘person’.

nemes that may aid in determining its status as either an inherited word or a borrowing. It does not appear, then, that the evidence supports including Nasal in Western Indonesian.

4.1.3 Enggano. Apart from Nasal, the other Sumatran languages, Barrier Islands languages and Batak, have even less lexical evidence supporting their inclusion in Western Indonesian. In Blust (2010), every piece of possible Batak lexical evidence is a straightforward Malay borrowing.[5] The Barrier Islands languages have none of the lexical innovations that define Western Indonesian. There is, thus, no lexical evidence that Sumatran languages belong to Western Indonesian, a lexically defined subgroup. Among these languages, Enggano is particularly interesting. Edwards (2015) recently proposed that Enggano, the southernmost Barrier Islands language, forms a primary branch by itself, as an isolate within Malayo-Polynesian. Enggano has undergone several sound changes that set it apart from all other languages in western Indonesia, including *k > *ʔ*, *t > *k*, *s > *k*, *ŋ > *h*, *j > *h*, *m > *b*, and *n > *d*. These changes are quite unique, particularly the changes *t/*s > *k*, *ŋ/*j > *h*, and the loss of a nasal distinction in the consonants (although a word-level nasal distinction was innovated, another unique feature of Enggano phonology).

Edwards (2015) discusses in great detail the historical phonology of Enggano, and it is worthwhile to discuss some of this here. First, it is probable that *t and *s merged as *t before *t became *k*. So while the change *s > *k* may seem astonishing, *s > *t is more common, but still not widespread. The fact that *t became *k* is itself interesting, as this change is found nowhere else in western Indonesia. The motivations for a *t > *k* change are not obvious, but like Hawaiian, which also changed *t to *k*, Enggano appears to have undergone this change only after inherited *k shifted to *ʔ*. There is a general implication in languages where *t shifts to *k* that *k must first shift to something else (usually glottal stop or zero). As Blust (2004) points out, this is not a universal implication, but there is a clear tendency for *t > *k* to be preceded by *k > *ʔ*. This is true for Enggano, and helps to put this otherwise isolated sound change in perspective. Finally, the nasal series of stops in Enggano merged with the voiced obstruents, *m > *b*, and *n > *d*. It is probably the case that *ŋ also merged with its nonnasal counterpart, *g, before being further reduced to *h*. This has an impact on the subgrouping of Enggano with relation to other Sumatran languages and is discussed further below.

4.1.4 Reflexes of *j and the position of the languages of Sumatra. Because there is essentially no evidence supporting the inclusion of Batak, Barrier Islands languages, and Nasal in Western Indonesian, they deserve special attention. The phonological history of these languages is markedly different from WIN, particularly with regard to reflexes of *j. To summarize, every language in Borneo merged PMP *j and *d as *d,[6] as did all GNB languages to the west. Some examples include Iban *si-ida > *sidaʔ* 'they' and *qapəju > *mpədu* 'gall', Kenyah *si-ida > *ida dua* 'those two' and *ŋajan > *ŋadan* 'name', Maanyan *si-ida > *here* 'they' and *qapəju > *aperu* 'gall', Sundanese *qadəp > *harip* 'in front of' and *qapeju > *hampəru* 'gall'. Rejang also reflects the merger of *j and *d, in addition to an apparently native reflex of *tuzuq 'seven'. There are no languages

5. Batak *suŋe* is listed as a native form, but it might also be an early borrowing.
6. Tamanic languages, found in the Upper Kapuas area of West Kalimantan, did not merge *j and *d. These languages, however, subgroup with South Sulawesi (Adelaar 1994) and are, thus, not included in WIN.

that unambiguously belong to Western Indonesian (either Greater North Borneo, or Barito) that do not reflect this merger, and it can likely be reconstructed to the immediate ancestor of these languages. In Taiwan, *j also merged with *d in Paiwan, Puyuma, and Pazeh, but these languages belong to three different primary branches, so merger of *j and *d there does not have much bearing on its importance in WIN. In Western Malayo-Polynesian, *j and *d also merged in several Philippine languages, but elsewhere, in Northern Luzon, *j merged with *g. A more complete list of reflexes of *j in "Western" Malayo-Polynesian languages can be found in appendix 1.

Blust (2010:81–82), in his discussion of Adelaar's proposed Malayo-Sumbawan subgroup, notes that *j > *d is "almost universal in insular Southeast Asia outside of some pockets in northern Luzon, Maloh of Borneo, scattered parts of Sumatra, and some of the languages of Sulawesi." He, thus, rejects *j > *d as evidence for Malayo-Sumbawan, citing its low quality. Blust is certainly correct in rejecting *j > *d as evidence for this lower-level subgroup, especially considering the fact that it is universal in languages of Borneo.

Blust's wording, however, is a bit misleading. First, it is not merely "some of the languages of Sulawesi" that do not show *j > *d, but rather the reconstructed phonologies of both major subgroups in Sulawesi that fail to show this merger: in South Sulawesi *j became *z (Mills 1975) and in Celebic *j became *y (Mead 2003). The only languages on Sulawesi where the change *j > *d is attested are, in fact, later Philippine transplants in northern Sulawesi. On Sulawesi, then, the change *j > *d is the exception, brought about by the expansion of Philippine languages, not the rule.

Second, "scattered parts of Sumatra" suggests that *j > *d occurred in multiple parallel innovations, with only a few languages here and there not showing the change. By far, the languages of Sumatra that do show *j > *d are either Malayic or more closely related to Malayic than the languages of Sumatra, where *j > *d is not attested. The only exception is Lampung, where *j merged with *d, but which does not appear to subgroup with Malayic. Lampung, however, is considered part of the larger Western Indonesian subgroup in this paper, which suggests that *j > *d in that language is also a retention, not an innovation.

Third, because the languages that reflect *j > *d belong to an exclusive subgroup, Western Indonesian, it follows that they inherited *j > *d from a common ancestor. Although it is true that *j > *d is nearly universal in Southeast Asia, this is because of the expansion of a group of languages that inherited it from PWIN, not because of multiple parallel innovations; *j > *d did not happen as many times as Blust suggests. The major exception to this is the Philippines, where *j > *d is attested across multiple "microgroup" boundaries. However, without a clear consensus on the internal subgrouping of Philippine languages (or of the validity of a Philippine group), the question of how many times *j became *d in the Philippines cannot be immediately answered.

Thus, while *j > *d might not be strong evidence, it is also not as weak as Blust makes it seem. The real reason why it is inadmissible as evidence for Malayo-Sumbawan is because it is a retention in these languages, not an innovation. For that reason, *j > *d should not be used to justify any lower-level subgroup in Western Indonesian, but neither is *j > *d alone strong enough to define WIN. Languages where *j did not merge with *d, however, should have their inclusion in WIN immediately questioned on the grounds that *j > *d is reconstructible to PWIN.

Interestingly, all of the languages of Sumatra that have a doubtful relationship to WIN also did not merge *j with *d. As I will argue below, Sumatran Barrier Islands languages, Batak, and Nasal probably merged *j with *g, with diverse changes thereafter. This in turn suggests that these languages do not subgroup with WIN, and should instead be grouped together as the only remaining descendants of a larger Sumatran subgroup that has been greatly affected by the expansion of Malay out of southwest Borneo.

Edwards (2015) argued that Enggano constitutes its own primary branch within Malayo-Polynesian, rejecting an earlier claim by Nothofer (1986) grouping Enggano exclusively with other Sumatran Barrier Islands languages and Batak. In Enggano, *j merged with *ŋ as *h* in medial position and as Ø word-finally, but this odd change did not happen in isolation. Importantly, Enggano merged *m and *n with their nonnasal counterparts, *m > *b, *n > *d. The change of *ŋ > *h* is less odd in this context, as it almost certainly went through an earlier stage where it too had denasalized: *ŋ > *g > *h. The change from nasal to nonnasal, then, rather than specifically targeting *m and *n, was apparently more general, targeting the feature [nasal]. This is a simpler history than *m, *n > *b, *d and *ŋ > *h, as it involves only a single change, but explains multiple reflexes. Edwards (2015:64) seems to support this for reflexes of *ŋ by stating that "based on the observation that *m/*n usually became [b]/[d], it is probable that PMP *ŋ first became [g]." Edwards does not, however, endorse the view that *j has a similar history, and does not consider reflexes of *j significant for Enggano subgrouping, arguing that the history of *j is circumstantial.

Nothofer's Barrier Islands-Batak subgroup is based partially on reflexes of *j. He argues that *j changed to *x in these languages, and is retained as *x* in medial position in at least Nias, Sixule, and Simeulue. Edwards points out that not only does Mentawai not show this change (*j is reflected as *g/ɣ* in Mentawai), the Enggano evidence is contingent on an unobservable intermediate stage, *j > *x > *h.* It is not the change *j > *x that is significant, however, as Nothofer suggested and Edwards rightfully opposed, but *j > *g, a change that includes Nasal and fits the phonological histories of the languages in the area. In Enggano, the merger of *j with *g is only observable via *ŋ, as no reflexes of PMP *g have been located. As noted earlier, *j merged with *ŋ and *ŋ almost certainly merged with *g. While Edwards objects to positing an intermediate stage for *j > *h*, there is a clear preferred chronology regarding the eventual merger of *j and *ŋ based on what we know about Enggano historical phonology and the inherent complexities of a direct change from *j to *h.* There are two logical possibilities, assuming an intermediate stage where *j had become *g: (i) *j merged with *ŋ and *ŋ later merged with *g (24a), or (ii)*j merged with *g, and later, *ŋ merged with *g (24b). A third possibility, which Edwards appears to endorse, is that the changes *ŋ > *h* and *j > *h* are unrelated, and that *j changed immediately to *h* (24 c):

(24) a. *j → *ŋ → *g
*ŋ →

b. *j → *g → *g
*ŋ →

c. *ŋ → *g → *x → h
*j → h

The change *j > *h* is phonetically and phonologically complex and requires multiple feature changes with no clear motivation (*j [gʸ] would have had to had simultaneously lost the palatal, velar, stop, and voice features with no intermediate stage, an unlikely scenario). Thus, it is unlikely that scenario (24c) accurately describes the history of *j in Enggano. Also, *j almost certainly did not merge directly with *ŋ, as shown in scenario (24a), where in all other contexts the nasals merged with nonnasals. Scenario (24b), however, is much more likely, as there is independent evidence that *ŋ merged with *g and, further, that *g became modern Enggano *h.* It also involves only a single feature change for *j (loss of palatalization). If *j > *g rather than *j > *x is recognized as the important sound change for subgrouping in Sumatra, Edwards's other objection, that *j has not become *x* in Mentawai, is less important. Mentawai seems to retain the earlier stage where *j had become *g, while other Barrier Islands languages further reduced *g.

A second advantage of positing a *j > *g sound change is that Nasal, an otherwise isolated language of southwest Sumatra, also displays evidence of this merger. As discussed earlier, Anderbeck and Aprilani (2013) consider Nasal an isolate within Malayo-Polynesian. In Nasal, *j became *k* in final position: *ikəj > *hiok* 'cough', *pusəj > *pusok* 'navel' (data from the Holle List). In medial position, Anderbeck and Aprilani give *pəgus* 'spicy' (< *pəjəs) as evidence for the change *j > *g*, but in the appendix *pahay* 'rice' (< *pajay) contradicts their claim.[7] In either case, however, the change *j > *k* in final position suggests *j became *g, with later terminal devoicing and weakening in medial position that parallels similar developments in Enggano.

Table 2 lists reflexes of *j in medial and final position in Nasal, Batak, and Barrier Islands languages. The multiple changes where *j [gʸ] appears to have become *g*, *x*, *h*, *k*, Ø, all reflect an earlier change, *j > *g, which was then inherited by all languages of Sumatra that do not belong to Western Indonesian.

TABLE 2. REFLEXES OF *j IN SUMATRAN LANGUAGES

PMP	Nasal	Enggano	Mentawai	Nias	Sixule	Simeulue	Toba	Karo
*-j-	g/h	h	g/ɣ	x	x	x	g	g
*-j	k	Ø	Ø	Ø	Ø	Ø	k	k

In conclusion, reflexes of *j in Western Indonesian languages unambiguously point to a merger of *j and *d. This change is universal in languages of Borneo, as well as Malayic and related languages to the west of Borneo. In Sumatra, however, languages that have a questionable relationship to Western Indonesian also reflect a different merger, where *j merged with *g. This change, along with a wholesale lack of lexical evidence that might support grouping these languages with Western Indonesian, supports the conclusion that they are descended from a separate protolanguage, and do not sub-

7. This may not be an inconsistency, as several Malayo-Polynesian languages have fortis reflexes of stops after penultimate schwa, but lenis reflexes elsewhere. Not enough is known about Nasal historical phonology to defend such an explanation, however, and these two reflexes must for now be listed as unexplained irregularities.

group immediately with languages of Borneo. If *j became *g once, it follows that the languages of Sumatra themselves form an exclusive subgroup, one that was altered dramatically with the rise of Malay-speaking empires.[8] This view is more in line with that of Nothofer (1986) and less in line with Edwards (2015) and Anderbeck and Aprilani (2013). It also suggests that Enggano, a language with a rich history of sound change, was at one time much more like the other Barrier Islands languages.

4.1.5 The position of Lampung. Lampung, located in South Sumatra, presents some issues with the above subgroupings. It is a "*pitu*" language, meaning that it does not reflect the change *tuzuq 'to point' > *tuzuq 'seven', which acts as a diagnostic for inclusion in Greater North Borneo (neither does it reflect any of the other lexical replacement innovations of GNB from Blust 2010 and Smith 2017a). It has been under heavy Malay influence as well, which makes identifying borrowings difficult. It has, however, merged *j and *d, which distinguishes it from other Sumatran languages, where *j merged with *g. If Lampung is a Western Indonesian language, it shows little evidence of the fact. Lampung *topay* 'tree shrew' and *tikus* 'mouse' reflect Western Indonesian innovations after Blust (2010), but it is difficult to say if these are borrowings or native. The same is true for Lampung *dəxian* 'durian', although the reduction of prepenultimate *u* to schwa suggests it might not be a borrowing. A single word, however, *kənuy* 'kite; bird of prey', does suggests that Lampung belongs in Western Indonesian. Smith (2017a) identified PWIN *kəniw 'eagle; hawk' as a subgroup-defining innovation, with multiple additional examples of a sporadic *-iw to *-uy change (see the evidence above). What makes this word particularly interesting is that Malay could not have acted as a source, as Malay does not reflect PWIN *kəniw, and reflects *-iw with *-i*. Ultimately, however, the merger of *j and *d, and a reflex of *kəniw, do not provide the kind of hard-hitting evidence needed for a strong subgrouping proposal. While I support the inclusion of Lampung in WIN, the proposal itself will likely be the subject of disagreement.

4.1.6 The position of Moken. Blust (2010:47) places Moken not only in Western Indonesian, but specifically in the Greater North Borneo group, which suggests that its aberrance is not a result of higher-level diversity, but rather recent and rapid sound change. Reflexes of *tuzuq 'seven' are the only evidence available linking Moken to Greater North Borneo (Blust 2010:68–71). However, Moken does not reflect the PWIN merger of *j with *d, as table 3 makes clear, which calls into question its inclusion in WIN.

TABLE 3. REFLEXES OF *j IN MOKEN

	PMP	Moken
*-j-	*qajəŋ 'charcoal'	kayaŋ
	*pajay 'field rice'	pai
	*ijuŋ 'nose'	yoŋ
*-d-	*qudaŋ 'shrimp'	kodaŋ
	*qudip 'life; alive'	kodip

8. Ultimately, more evidence for such a subgroup needs to be located. While it is certainly conspicuous that precisely those languages of Sumatra that do not fit into Western Indonesian reflect the change *j > *g, the entire subgrouping proposal rests on a single sound change.

Although the semantic shift from *tuzuq 'to point' to *tuzuq 'seven' is still considered strong enough to act as a diagnostic to include a language in GNB, there are several cases where either (i) apparent reflexes of *tuzuq 'seven' were borrowed into languages that did not originally reflect this change (Barito languages, for example), or (ii) languages have words that appear similar to *tuzuq 'seven' but suffer from phonological irregularities that make them questionable. In the second case, some lesser-known examples are Lebo' Vo' Kenyah *tuʃək* 'seven', which reflects Proto-Western Lowland Kenyah *tujək, where other Kenyah languages, like Lepo Tau, reflect *tuzuq as *tujoʔ* 'seven', and Land Dayak examples like Benyadu *ijuʔ*, Hliboi *juəy*, and Golik *ijuʔ,* among others that reflect Proto-Land Dayak *ijuʔ, a reconstruction that, because of multiple irregularities, is rejected as a reflex of *tuzuq 'seven' in Smith (2017a). In the case of Moken, Blust (2010:112–13) notes that Moken examples of *tuzuq 'seven' are either borrowings or contain phonological irregularities that "raise questions about their histories." These are Rawai Moken *luju:k* and Klɔŋ Yuan Moken *dɯyu:k* , which both exhibit inexplicable irregularities. Such words, like the Lebo' Vo' and Land Dayak examples above, are more likely chance resemblances, which may appear striking at first, but cannot be used as evidence linking a language to GNB.

No other lexical evidence has been put forward to further support Moken's inclusion in GNB, and considering both the irregularities found in apparent reflexes of *tuzuq 'seven' and Moken's lack of the *j/*d merger, it is necessary to remove Moken from Western Indonesian as well. Additionally, because Moken reflects *j as *y,* it cannot be further grouped with Sumatran languages, where *j became *g. Considering the lack of evidence for including Moken in either WIN or for grouping it with Sumatran languages, and because there are simply no other languages in the area with which Moken might form a subgroup, Moken must constitute yet another primary branch of Malayo-Polynesian in western Island Southeast Asia, bringing the total to three: Western Indonesian, Sumatran, and Moken.

4.1.7 Conclusion. The existence of three primary branches of MP in western Island Southeast Asia raises questions about the history of Austronesian settlement of Island Southeast Asia as proposed by Blust (2010). In that publication, Blust proposes a scenario where the very first Austronesians who entered western Indonesia came through Borneo in a northwest and a southeast group. The southeast group was responsible for the settling of the Barito river area, and all islands to the west where WIN languages are spoken; while the northwest group settled the part of Borneo facing the South China Sea. The expansion of Malayic out of West Kalimantan would have resulted in the leveling of WIN languages in Sumatra. This proposal assumes that Sumatra was settled by a single expansion of PWIN speakers, and was only later overlaid by Malay. Because of the incredibly rapid expansion of PMP speakers into Island Southeast Asia, this means that PWIN would have developed "on the move," with no time to develop as a unit, considering the archaeological facts that show forest clearing in Sumatra as early as anywhere else. What the new proposal does, however, is show that while it is true that speakers of PMP spread rapidly into Sumatra, their languages were not descended from PWIN. Rather, PWIN would have had to have developed for a time elsewhere, while the descendants of the original Austronesian settlers of Sumatra diversified into the multiple, lin-

guistically distinct groups we find today. When Malay speakers expanded into Sumatra, it was not descendants of PWIN that they displaced, but a separate group with no immediate relationship to the languages of Borneo or the southern Greater Sunda islands.

4.2 PHILIPPINES. The languages of the Philippines have been the topic of ongoing debate regarding their position in Austronesian, and the legitimacy of a Philippine subgroup. Earlier writers—Blumentritt (1899), Blake (1902, 1920), Brandstetter (1911), Lopez (1967), and Sheerer (1918)—have more or less assumed a Philippine group. Blust (1991, 2005) and Zorc (1986) have defended such a subgroup with lists of lexical innovations as the main type of evidence; Llamzon and Martin (1976), Llamzon (1975), and Paz (1981) have used the comparative method to reconstruct the phonology, morphology, and lexicon of Proto-Philippines. Many others (Reid 1978, 1982, 2010; Ross 2005; Pawley 1999; Liao 2011) either question or reject outright the Philippine subgroup hypothesis. These views will be discussed later, but first, it will be worthwhile to review subgroups within the Philippines that enjoy wider acceptance than the Philippine subgroup itself. Within the proposed Philippine group, there are several widely recognized "microgroups" for which Blust (1991) gives a thorough review of the literature. Table 4 lists these microgroups.

TABLE 4. PHILIPPINE SUBGROUPS AFTER BLUST (1991)

Bashiic	Bilic	Central Luzon
Central Philippines	Cordilleran	Danaw
Gorontalo-Mongondow	Inati	Kalamian
Manobo	Minahasan	Palawanic
Sangiric	South Mangyan	Subanun

Of these 15 microgroups, seven are grouped together in a Greater Central Philippine subgroup (Blust 1991): Central Philippines, South Mangyan, Palawanic, Danaw, Manobo, Subanun, and Gorontalo-Mongondow. In total, after factoring in Blust's Greater Central Philippine hypothesis, there are 9 recognized Philippine subgroups, as listed in table 5. Bilic is also known as South Mindanao, Cordilleran as Northern Luzon, Bashiic as Batanic, and Kalamian as Calamian.

TABLE 5. PHILIPPINE SUBGROUPS WITH GREATER CENTRAL PHILIPPINES AFTER BLUST(1991)

Bashiic	Bilic	Central Luzon
Cordilleran	Greater Central Philippine	Inati
Kalamian	Minahasan	Sangiric

Through the remainder of this section, Philippine microgroups will often be referred to with the following abbreviations: BAT, Batanic (Bashiic); BIL, Bilic; CLZN, Central Luzon; GCPH, Greater Central Philippines; KAL, Kalamian; MIN, Minahasan; NLZN, Northern Luzon (Cordilleran); SAN, Sangiric.

4.2.1 Disagreements on the validity of a Philippine subgroup. Arguments against the validity of a Philippine subgroup arise from the correct observation that there are no

phonological innovations of high quality that distinguish PPH from PMP. Reid (1978) began questioning the validity of the Philippine subgroup, but did not offer a full rebuke until "The demise of Proto-Philippines" was published (Reid 1982). Although his alternative subgrouping proposal has not gained wide acceptance, the question of the validity of the Philippine subgroup hypothesis persists. Pawley (1999), Ross (2005), and Liao (2011) have made more recent statements doubting the validity of a Philippine subgroup, typically on the grounds that phonological reconstruction of PPH yields a system identical to that of PMP. Lobel (2013:13–14) accepts a Philippine subgroup but only as "a convenient point of reference" and does not take a strong stance on the issue. Aside from widespread doubt on the validity of the Philippine subgroup, the archaeological evidence, as discussed earlier, points to a single rapid population expansion of PMP speakers out of the northern Philippines and into Island Southeast Asia. It is unlikely that this expansion took place over more than 10–20 generations, hardly enough time for substantial sound changes to take root. Also, as already discussed, rapid population expansion usually results in a rake-like tree, with diversity in primary branches. Thus, if the Philippines were settled by Austronesians in the same expansion that settled the rest of Island Southeast Asia, we would expect multiple primary branches of MP in the Philippines. Blust (1991) proposes that a secondary expansion by speakers of Proto-Philippines wiped out the expected past diversity. I have no issues with such a scenario, as population expansion often results in the leveling of linguistic diversity. If it is to be accepted in the Philippines, however, the evidence must unambiguously support the existence of a Proto-Philippine language.

It is unlikely, then, that sound change will be adequate for determining the relationship between Philippine languages. However, if we turn to lexical evidence to try to defend the Philippine subgroup, the evidence must be of high quality, as there are issues inherent in lexical evidence not found in phonological evidence.

4.2.2 Using lexical evidence. Western Indonesian is entirely defined by lexical innovations, with an additional note that *j probably merged with *d at PWIN. The lexical evidence for Western Indonesian is of high quality, and has been extensively vetted for borrowings and chance resemblances. Because Borneo marks the beginning of Mainland Southeast Asian flora and fauna after the Philippines, it also has several innovations that must have been coined by the first settlers of the island, which provide high quality evidence. The innovations themselves are found throughout the island, not in only a few subgroups, and have an additional presence in languages of Java, Bali, Lombok, and Sumbawa. Lexical evidence for a Philippine subgroup should be vetted along the same lines, and the following sections seek to make clear which types of lexical evidence should be considered high quality and with what criteria lexical evidence should be evaluated.

In proposing lexical evidence for a subgrouping hypothesis, it is important to distinguish between two types of lexical innovations: replacement innovations and nonreplacement innovations. Replacement lexical innovations are new lexemes that completely replace a word of the same meaning that has been reconstructed to a higher subgroup. To cite an example from outside the Philippines, Proto-Malayo-Polynesian *isaŋ/*hasaŋ 'gills' is reflected throughout Borneo, in Malay *insaŋ*, Kapuas *hisaŋ*, Lebo Basap *saŋ*, Kanowit *asaŋ*, Long Bawan Lun Dayeh *asaŋ*, and many others. In Kayanic languages, however, there are no reflexes of *isaŋ/*hasaŋ. Instead, one finds the lexical replacement

innovation *ŋad, which is reflected in Data Dian *ŋan*, Balui Liko *ŋar*, Busang *ŋar*, Bahau *ŋa:l*, Long Gelat *ŋin*, Gaai *ŋal*, and Kelai *ŋœl*. In other words, *ŋad is reflected in every Kayanic subgroup and in the vast majority of Kayanic languages (see Smith 2017a). Thus, Proto-Kayanic *ŋad replaced *isaŋ/*hasaŋ, and *ŋad was then inherited by its daughter languages in much the same way that a phonological innovation is inherited. This is a replacement innovation and has more subgrouping potential than any other type of lexical innovation because directionality can be established. That is, because evidence points to PMP *isaŋ/*hasaŋ 'gills', and because reflexes of *isaŋ/*hasaŋ are not found in any Kayanic language, it follows that PMP *isaŋ/*hasaŋ was replaced by Proto-Kayanic *ŋad at some point in the common history of these languages. Thus, the change *isaŋ/*hasaŋ 'gills' > *ŋad 'gills' is the only plausible explanation. The only alternative, that a proposed PMP *ŋad 'gills' was independently replaced by forms resembling *isaŋ/*hasaŋ 'gills' in every other Austronesian subgroup except Kayanic, is quite improbable. Without directionality, lexical innovations show little more than similarity, and similarity alone does not support subgrouping.

Replacement lexical innovations are the strongest form of lexical evidence. However, other types of lexical evidence may be proposed as a supplement. The two main types of nonreplacement innovations are semantic shifts and synonymic innovations. Semantic shifts occur when an existing word changes meaning. In some cases, a word may shift in meaning and come to replace the inherited word. A published case of such a change is found in the proposed Greater North Borneo subgroup (Blust 2010). Here, PAN *tuzuq 'to point' shifted to 'seven'. This shift resulted in the replacement of PAN *pitu 'seven', which gives it added strength as subgrouping evidence. A synonymic innovation is harder to evaluate. Here, an innovated word does not replace the inherited word, and we must reconstruct two synonymous words to the protolanguage. Blust (2010) provides more examples. PMP *(ba)labaw 'rat' is retained in the North Sarawak subgroup, as in Kenyah *bəlabaw*, but an innovated word, Proto-Greater North Borneo (PGNB) *tikus, is found throughout Sabahan and Malayic languages, as in Kadazan Dusun, Iban, and Malay *tikus*. It is possible that *tikus had a slight semantic difference from *(ba)labaw, but that difference cannot be reconstructed based on evidence from modern languages. We have no choice but to reconstruct both *(ba)labaw and *tikus to PGNB with the same meaning. The reason why synonymic innovations are less powerful than replacement innovations is because there is an equal probability that the innovated form diffused throughout the languages in question, and assigning it to a protolanguage requires one to assume an apparently unmotivated innovation of synonymy. It lacks the clearly defined split between the innovated form and the replaced form one finds in replacement innovations, and, like an irregularly attested sporadic sound change[9] that affects some but not all reflexes of a phoneme, it can never provide the kind of hard-hitting evidence necessary for confident subgrouping.

9. There are two types of irregular sound change. One, which provides powerful subgrouping evidence, occurs where a phoneme *x irregularly changes to *y* in a specific word, with that irregular sound change reflected in all daughter languages of the proposed subgroup that it defines. Another type of irregular change, where *x seems to randomly become *y* in a group of languages, but without a match in specific lexemes where it occurs (for example, *b sometimes becomes *w* in a group of languages, but not in specific lexemes), offers no subgrouping evidence.

Finally, lexical innovations must be robustly attested to a greater degree than sound change. Sound change is regular, so even with only a few examples of a change in a larger set of languages, a historical linguist can confidently reconstruct a word that shows a subgroup-defining innovation. Regularity makes this possible, because we can infer that if *x became *y* in one word, that *x also became *y* in a similar environment in other words. With lexical evidence, however, there is no regularity principle that may allow a linguist to reconstruct an innovation to a protolanguage with only two or three examples. Because of this fundamental difference, although a few words may show striking similarities, a robust set of examples is necessary when using lexical data to ensure that what is being presented is not the result of chance, diffusion, or borrowing.

4.2.3 Reevaluating Philippine lexical evidence. Lexical evidence for a Philippine subgroup has generally not been vetted in a manner that distinguishes replacement innovations, semantic shifts, and synonymic innovations, nor is the robustness of each proposed lexical innovation made clear. Because of this, the lists of sometimes hundreds of innovations one finds in the published literature are misleading, as they do not distinguish high-quality from low-quality innovations. In the section below, I scrutinize a fraction of the evidence for a Philippine subgroup from Blust and Trussel (ongoing).[10] Fifty comparisons are considered, and judged on four principles based on the above discussion:

- Principle 1: the innovation should be a replacement.
- Principle 2: the innovation should be robustly attested both in number of individual languages (justifying their reconstruction to a protolanguage within the Philippine group) and in number of microgroups (to justify their reconstruction to Proto-Philippines).
- Principle 3: the sound correspondences between innovations must be regular.
- Principle 4: the innovations should be geographically noncontiguous.

Violation of any of these principles will call into question the validity of the proposed lexical innovation, but note that violation of a principle does not necessarily imply that the evidence should be immediately thrown out, just that it is of low quality. Note that I do not scrutinize this evidence from a belief that a Philippine subgroup is necessarily invalid. Rather, I mean to show that the present lexical evidence does not form a strong argument that the hypothesis of a Philippine subgroup is more supported than a competing hypothesis where Philippine languages represent either more than one primary branch of MP or, perhaps more interestingly, an innovation-defined linkage.[11]

(25) *abag ‘join forces, cooperate in working’
Found in: NLZN: Kankanaey and GCPH: Cebuano
Status: violates principle 2

(26) *abaká ‘Manila hemp: *Musa textilis*’
Found in: BAT: Yami, Itbayaten; NLZN: Ilokano, Ifugaw; CLZN: Kapampangan; GCPH: Tagalog, Bikol, Aklanon, Hanunóo, Maranao; BIL: Tiruray

10. These examples can be found by navigating from the front page of the *Austronesian comparative dictionary* (Blust and Trussel ongoing) to “Proto-form indexes”, then to “PPH,” which will bring the visitor to a complete list of PPH reconstructions in alphabetical order.
11. David Zorc (pers. comm., July 25, 2017) first pointed out this possibility to me, but a substantial amount of research is needed to test such a hypothesis.

Status: OK, but note that Manila hemp is a widely traded item, and its distribution may ultimately be from trade.

(27) *abat 'spirit that causes sickness'
Found in: NLZN: Isnag (sometimes Isneg), Bontok; GCPH: Cebuano
Status: violates principle 2

(28) *anam 'a plant: *Glochidion* spp.'
Found in: BAT: Yami and NLZN: Ilokano
Status: violates principles 2, 4

(29) *aŋas 'face, countenance'
Found in: NLZN: Kankanaey (Northern), Ifugaw; BIL: Tiruray
Status: violates principle 2

(30) *ananay 'exclamation of pain; ouch!'
Found in: BAT: Yami, Itbayaten, Ibatan (may be a loan); NLZN: Ilokano, Ifugaw
Status: violates principles 2, 4

(31) *anayup 'beautyberry: *Callicarpa* spp.'
Found in: BAT: Yami, Itbayaten; NLZN: Kankanaey
Status: violates principles 2, 4

(32) *aŋ(ə)tad 'clearly visible'
Found in: BAT: Ibatan and GCPH: Cebuano
Status: violates principle 2

(33) *atúbaŋ 'to face, confront'
Found in: BAT: Yami, Ibatan; NLZN: Itawis; GCPH: Bikol, Hanunóo, Aklanon, Hiligaynon, Cebuano, Maranao, Subanun, and Manobo
Status: OK

(34) *aum 'a plant: *Melanolepis multiglandulosa* (Reinw.)'
Found in: BAT: Yami; CLZN: Sambal
Status: violates principle 2

(35) *ayaw 'depart, separate from'
Found in: NLZN: Kankanaey; GCPH: Bikol, Aklanon, Mansaka
Status: violates principles 2, 3 due to irregularities in Kankanaey reflexes of *-aw

(36) *bagut 'pull out, as hair'
Found in: BAT: Yami, Itbayaten, Ibatan; NLZN: Ilokano, Casiguran Dumagat, Ibaloy
Status: violates principles 2, 4

(37) *bakaŋ 'divert the attention'
Found in: BAT: Itbayaten; GCPH: Bikol
Status: violates principle 2

(38) *báli 'join, participate in, accompany'
Found in: NLZN: Bontok; GCPH: Bikol; MIN: Proto-Minahasan
Status: violates principle 2

(39) *balinu 'beach morning glory: *Ipomoea pes-caprae'*
Found in: BAT: Yami, Itbayaten, Ivatan; NLZN: Ilokano, Ibanag
Status: violates principles 2, 4

(40) *balítiq ‘banyan’
Found in: BAT: Yami, Itbayaten, Ivatan, Ibatan; NLZN: Ilokano, Isnag; CLZN: Kapampangan; GCPH: Tagalog, Bikol, Aklanon, Maranao, Western Bukidnon
Status: violates principle 1 on the grounds that reflexes of PMP *nunuk ‘banyan’ are found throughout the Philippines.

(41) *balñaw ‘rinse, rinse off’
Found in: NLZN: Casiguran Dumagat; GCPH: Tagalog, Bikol, Aklanon, Cebuano
Status: violates principles 2, 4. Also, the reconstructed form appears to violate Proto-Philippine phonotactics, where heterorganic consonant clusters were only found in inherited reduplicated monosyllables. The word might be reconstructed with a schwa separating the consonants to avoid this, but it is not clear if schwa would have then been regularly deleted in all of the languages where the form appears.

(42) *banisah ‘a tree: *Planchonella obovata*’
Found in: BAT: Yami, Itbayaten, Ivatan; GCPH: Tagalog
Status: violates principle 2

(43) *baŋaq ‘earthenware water jar’
Found in: BAT: Yami, Itbayaten; NLZN: Ilokano, Ibaloy; GCPH: Tagalog
Status: violates principles 2, 4

(44) *baqak ‘old’
Found in: BAT: Itbayaten; NLZN: Ilokano, Isnag, Ibaloy
Status: violates principles 2, 4

(45) *baraŋgay ‘communal boat’
Found in: NLZN: Ilokano; GCPH: Aklanon
Status: violates principle 2. Probably a loan.

(46) *baRat ‘to meet’
Found in: BAT: Itbayaten, Ibatan; GCPH: Masbatenyo, Cebuano
Status: violates principle 2

(47) *batiŋ ‘to catch by the legs’
Found in: BAT: Itbayaten; GCPH: Tagalog; KAL: Agutaynen
Status: violates principle 2

(48) *bantug ‘fame; famous; renowned’
Found in: GCPH: Tagalog, Bikol, Hanunóo, Romblomanon, Masbatenyo, Aklanon, Waray-Waray, Hiligaynon, Cebuano, Maranao, Binukid, Manobo (Western Bukidnon), Mansaka; SAN: Sangir.
Status:violates principle 2. With the exception of Sangir, this word is confined to GCPH, and may be only a borrowing.

(49) *batuR ‘aligned, lined up’
Found in: BAT: Itbayaten; NLZN: Ilokano, Isnag, Bontok
Status: violates principles 2, 3, 4. Reflexes of *R in Ilokano and Bontok are irregular.

(50) *biklaj ‘spread out, unfurl’
Found in: BAT: Yami, Itbayaten, Ibatan; GCPH: Binukid, Tontemboan
See comment under (51)

(51) *buklad ‘unfold, open up, blossom’
Found in: BAT: Itbayaten; GCPH: Bikol, Hanunóo, Aklanon, Cebuano, Maranao, Binukid
Status: both (50) and (51) appear to violate principle 3; *biklaj became Binukid *bilad* but *buklad became Binukid *buklad*: why did Binukid treat *-kl-* differently in these two words?

(52) *bilu ‘blackened’
Found in: NLZN: Kankanaey; GCPH: Mansaka
Status: violates principle 2

(53) *bítu ‘hole, cavern’
Found in: BAT: Itbayaten; NLZN: Ifugaw; GCPH: Cebuano, Maranao, Binukid, Manobo, Mansaka
Status: violates principle 1 on the grounds that reflexes of PMP *lubaŋ ‘hole’ and *liaŋ ‘cave’ are found throughout the Philippines. Also violates principle 2.

(54) *bujas ‘to pluck, as fruit’
Found in: BAT: Itbayaten; NLZN: Ifugaw
Status: violates principle 2, 4

(55) *bulbul ‘body hair, feathers’
Found in: BAT: Yami, Itbayaten; NLZN: Isnag; CLZN: Kapampangan; GCPH: Tagalog, Bikol, Hanunóo, Aklanon, Palawan Batak, Cebuano, Maranao, Binukid, Western Bukidnon Manobo, Mansaka; BIL: Tiruray, Tboli; MIN: Proto-Minahasan.
Status: strong

(56) *bulud ‘borrow, lend’
Found in: BAT: Yami, Itbayaten, Ibatan; NLZN: Ilokano, Bontok, Ifugaw
Status: violates principles 2, 4

(57) *bunal ‘beat up, bruise someone’
Found in: GCPH: Aklanon, Hiligaynon, Cebuano, Binukid, Mansaka; SAN: Sangir
Status: violates principles 2, 4.

(58) *bunuŋ ‘distribute; share’
Found in: BAT: Yami, Itbayaten, Ibatan; NLZN: Ilokano, Casiguran Dumagat
Status: violates principles 2, 4

(59) *buŋtút ‘stench, bad odor’
Found in: BAT: Itbayaten, Ibatan; NLZN: Kankanaey, Ilokano, Ifugaw
Status: violates principles 2,4

(60) *busuáŋ ‘strong free flow of water’
Found in: BAT: Itbayaten; GCPH: Hanunóo, Aklanon, Cebuano, Binukid
Status: violates principle 2

(61) *dakə́l ‘big, large (in size, quantity)’
Found in: BAT: Yami, Itbayaten; NLZN: Dupaningan Agta, Itawis, Bontok, Ifugaw, Pangasinan; CLZN: Kapampangan; GCPH: Bikol, Palawan Batak, Maranao, Binukid, Western Bukidnon; SAN: Proto-Sangiric.
Status: This appears to be a strong piece of evidence. Note, however, that many of the languages where this word is found have also been in intense contact and have a known history of borrowing. It may be shown that this word has been widely borrowed. However, because it is a replacement, it has added strength in that borrowing cannot explain why we have not yet been able to locate reflexes of PMP *Raya in Philippine languages.

(62) *dəsdəs ‘to rub together’
Found in: BAT: Itbayaten, Ibatan; GCPH: Hanunóo
Status: violates principle 2 (also semantic mismatch)

(63) *əŋit ‘laugh’
Found in: NLZN: Kakidugen Ilongot; GCPH: Western Bukidnon, Ilianen Manobo
Status: violates principle 2

(64) *gawəd ‘betel pepper’
Found in: BAT: Yami, Itbayaten; NLZN: Ilokano, Isnag, Casiguran Dumagat, Ibaloy
Status: violates principle 1 on the grounds that PMP *Rawəd is reflected in the Philippines. violates principles 2, 4. Also, it is conspicuous that *R became *g in some of the languages that are said to reflect *gawəd.

(65) *gunay ‘movement; to move’
Found in: BAT: Yami, Ibatan; NLZN: Ilokano, Bontok, Ibaloy
Status: violates principles 2, 4

(66) *guyud ‘banana sp.’
Found in: BAT: Itbayaten, Ibatan; NLZN: Isnag
Status: violates principles 2, 4

(67) *hadawiq ‘far, distant’
Found in: BAT: Yami, Itbayaten, Ivatan; NLZN: Bontok, Kankanaey, Ifugaw
Status: violates principles 2, 3, 4. Note that two words are found in the Philippines for ‘far’, *hadawiq and *adayuq, with no clear distinction between the two. However, it is also true that PMP *zauq is found nowhere in the Philippines. One of the two candidates for ‘far’, *adayuq or *hadawiq, might be a replacement innovation, and, thus, carry more weight as subgrouping evidence.

(68) *hayəp ‘animal’
Found in: CLZN: Bolinao; NLZN: Pangasinan; GCPH: Tagalog, Bikol, Hanunóo, Aklanon, Hiligaynon, Palawan Batak, Cebuano, Mansaka.
Status: violates principles 2, 3, 4. This word is mostly found in GCPH languages, with only two examples, Bolinao and Pangasinan, found outside GCPH.

(69) *hapun ‘to roost, as chickens’
Found in: BAT: Itbayaten, Ibatan; NLZN: Ilokano, Isnag, Casiguran Dumagat; CLZN: Kapampangan; GCPH: Tagalog, Bikol, Aklanon, Palawan Batak, Cebuano, Maranao, Mansaka, Manobo
Status: violates principles 3, 4: Batanic languages should preserve initial *h-.

(70) *hələk ‘sleep’
Found in: NLZN: Kankanaey, Ifugaw; GCPH: Tagalog, Kalamian Tagbanwa, Palawan Batak.
Status: violates principle 1 on the grounds that PMP *tiduR and *tuduR ‘sleep’ are found throughout the Philippines. Violates principles 2, 4.

(71) *hilək ‘rudderfish: *Kyphosus cinerascens*’
Found in: BAT: Yami, Itbayaten; NLZN: Ilokano
Status: violates principles 2, 4

(72) *ipus ‘tail’
Found in: BAT: Itbayaten; NLZN: Ilokano, Isnag, Itawis, Bontok, Casiguran Dumagat; MIN: Proto-Minahasan; GCPH: Mongondow
Status: violates principle 1 on the grounds that PMP *ikuR is reflected throughout the Philippines.

(73) *ka- ‘stative prefix marking high degree of some quality; extremely; superlative marker’
Found in: BAT: Yami; NLZN: Ilokano; CLZN: Kapampangan; GCPH: Aklanon, Hiligaynon, Cebuano
Status: This is probably a retention. A *kə-* prefix is also used in Kanowit (a language of Borneo), where all prepenultimate vowels reduced to schwa: for example, *pədeh* ‘painful’ but *toʔoh kə-pədeh!* ‘very painful!’ and *ləboʔ ñiduh mənibaʔ səno* ‘that house is truly short’ but *ñiduh nam rəgaw toʔoh kə-mənibaʔ* ‘he is not tall, (he) is very short.’ These examples may show that *ka- was used in this way as early as PMP.

(74) *kawkaw ‘dig a hole by scratching’
Found in: BAT: Itbayaten; NLZN: Bontok
Status: violates principles 2, 3, 4: reflexes of *-aw are irregular.

Table 6 summarizes the item-by-item list presented above. As the table makes clear, the vast majority of proposed innovations violate one or more of the principles used to judge lexical evidence. Nearly three fourths of the examples are restricted to only two microgroups, and 84 percent are restricted to Batanic, Northern Luzon, Central Luzon, and Greater Central Philippines, groups that are concentrated in a geographically contiguous area with an established history of contact. Not listed on the table is the fact that 15 (30 percent) of the lexemes are found in only two or three *languages*. At this point, it is reasonable to predict that the PPH reconstructions currently found in the *Austronesian comparative dictionary* will have similar issues with quality. Two of the words from the above list, however, are of high quality. They each appear in five microgroups and are replacement innovations. They are discussed in more detail below.

TABLE 6. SUMMARY OF PPH EVIDENCE

Microgroups where innovations are found	No. lexemes in the microgroups
BAT, NLZN	17
BAT, GCPH	8
NLZN, GCPH	8
BAT, NLZN, CLZN, GCPH	3
BAT, NLZN, GCPH	3
GCPH, SAN	2
NLZN, BIL	1
BAT, CLZN	1
NLZN, GCPH, MIN	1
BAT, GCPH, KAL	1
BAT, NLZN, GCPH, BIL, MIN	1
BAT, NLZN, CLZN, GCPH, SAN	1
CLZN, NLZN, GCPH	1
BAT, NLZN, CLZN, GCPH	1
BAT, NLZN, MIN, GCPH	1

The innovated form *dakə́l replaced *Raya as ‘big’, and reflexes of *Raya have not been located in Philippine languages. PMP *bulu ‘body hair; feathers; fur; down; floss on plant stems; color; type; kind’ underwent semantic narrowing, and lost the meaning ‘body hair’ and ‘feathers’. PPH *bulbul ‘body hair, feathers’ was innovated in its place. These two changes are potentially powerful for subgrouping because (i) they are replacement innovations and no examples have been located where either *Raya or *bulu are retained with their full meaning, and (ii) they occur in basic vocabulary and it would be difficult to justify writing them off as loans. If *dakə́l and *bulbul are not replacement innovations, then it is necessary to explain why *Raya and *bulu (meaning ‘body hair’) have not yet been located in any Philippine language. Also, although it is legitimate to criticize these innovations as occurring only in a geographically contiguous group of languages, if we back away from the claim that PPH is legitimate in the first place, they may be evidence for a lower-level group that includes the languages where they are found, although this will still fail to explain why *Raya and *bulu ‘body hair’ have not been located in *any* Philippine language.

Another piece of lexical evidence for a Philippine subgroup is a semantic shift, which Blust (2005) points to as being of high quality. In that shift, PMP *Rumaq ‘domicile’ shifted to PPH *Rumaq ‘scabbard for a sword’. Once *Rumaq lost the meaning ‘domicile’, another word took its place. Here, PMP *balay ‘public building’ replaced *Rumaq as ‘domicile’. The latter change, ‘public building’ > ‘domicile’, is fairly common in AN languages and has little impact on subgrouping. The first, however, is unique to the Philippines, and the combination of both in dispersed subgroups suggests inheritance rather than parallel innovation. Because *Rumaq does not appear to mean ‘domicile’ in any Philippine language, it follows that the shift occurred once, in a language ancestral to all Philippine languages. In the *Austronesian comparative dictionary*, reflexes of *Rumaq ‘domicile’ as ‘scabbard for a sword’ are found in three separate groups found spread throughout central and southern Philippines:

(75) Central Luzon: Botolan
GCPH: Subanun
Binukid
W. Bukidnon
Bilic: Tiruray
Tboli
Blaan

The lexical evidence for a Philippine subgroup presents a mixed picture. The majority of the evidence proposed appears to be of low quality, with a questionable impact on the Philippine subgroup hypothesis. There are, however, a small number of lexical innovations of high quality that appear in several (but by no means all) Philippine microgroups. The impact of these lexemes on the Philippine subgroup hypothesis as a whole remains indeterminate, however, as we cannot point to any lexical innovation or group of innovations that are found throughout all Philippine microgroups. That is, there are Philippine microgroups where none of the quality evidence has been located. Until a more concise list of lexical evidence is compiled, which removes weak evidence and focuses on quality, more is needed to convincingly argue for a Philippine subgroup with lexical evidence.

4.2.4 The phonological evidence for PPH. Despite claims to the contrary, the phonology of PPH is not identical to PMP as it is currently reconstructed. All Philippine languages have shifted PMP *z to *d and merged *c with *s (Charles 1974:480).[12] The second of these changes is very common (nearly universal, in fact) and not everyone agrees on reconstructing *c in the first place. The first, however, is slightly more significant. Although it is true that all Philippine languages for which relevant data are available show a shift of *z to *d, because Kapampangan retains *ñ and languages do not typically have more nasal places of articulation than obstruents (Ferguson 1966, although Kapampangan itself runs counter to this claim), then PPH itself probably retained *z as *z. If one accepts the shift of *z to *d as evidence for Proto-Philippines, however, it raises other concerns: namely, the change *z > *d is not uncommon. A similar change is found in many independent cases in MP, and one must question if it arose through independent parallel innovation in the Philippines as well.

Philippine accent has also received some attention, and if one reconstructs accent to PPH on the assumption that it appears in multiple primary branches, then it too will form evidence for a Philippine subgroup. There are two problems with this hypothesis, however. First, there is a lack of agreement regarding the position of accent in Austronesian (Dahl 1981, Ross 1992, and Zorc 1983 have proposed reconstructing distinctive accent to higher nodes); and second, the logic for reconstructing an accent system to PPH is circular, as it requires one to accept the validity of the Philippine subgroup in order to make the reconstruction in the first place, since multiple Philippine subgroups show no signs of

12. Although Charles claims that *z became *d, the data do not appear to back this up, because if *z had become *d at the PPH level, it would have merged with *d from PMP *d. However, multiple languages have different reflexes of *z and *d: for example, Itbayaten *doha* 'two' from *duha but *rahan* 'road, path' from *zalan, and Kakilingan Sambal *olaŋ* 'shrimp' from *qudaŋ but *odan* 'rain' from *quzan.

an accent system.[13] Ultimately, accent cannot be used to argue for a Philippine group, as there is still disagreement about its overall placement in Austronesian.

4.2.5 Conclusion. I remain skeptical of the lexical evidence used to define the Philippine group. This is mostly because, although a very large number of innovations have been reported, little time has been dedicated to discussing the quality of each of these innovations. Historical linguistics is not a numbers game, it is the qualitative analysis of evidence, and even if hundreds of lexical innovations can be identified, it is the quality of those innovations that makes them count. The majority of the lexical evidence presented so far is of low quality, and high-quality innovations are mostly located in a small number of subgroups, in the northern and central Philippines. Few of the lexical innovations are found throughout the Philippines in diverse subgroups. There is no strong phonological evidence for a Philippine subgroup; the Proto-Philippine sound system is nearly identical to that of Proto-Malayo-Polynesian. If one accepts *z > *d as a subgroup-defining innovation, it will ultimately be criticized as being of too low quality.

However, skepticism of a proposal is not the same as having a viable alternative. There are 15 widely agreed upon microgroups in the Philippines, which are reduced to 9 if one accepts the Greater Central Philippine hypothesis. While the evidence for a Philippine subgroup is not convincing, neither is an argument that the nine microgroups each constitutes a primary branch of MP. Also, although much of the Philippine lexical evidence is troublesome, the presence of innovations like *Raya ‘big’ > *dakə́l ‘big’, *bulu ‘body hair’ > *bulu ‘floss’/*bulbul ‘body hair’, and *Rumaq ‘domicile’ > *Rumaq ‘scabbard’ forces one to consider the possibility that these words were inherited from a common ancestor (at least in the languages where they are located). No alternative model will be proposed here, as the question of the ultimate position of the Philippine languages will take time to answer. However, as indicated earlier, the linkage model has a great potential to explain why the Philippine languages appear to be more closely related to each other than to other AN languages, and also why the evidence for Proto-Philippines remains so inconclusive. To argue for a linkage model, however, will take a great deal of space, and deserves a dedicated discussion. In this paper, I instead list Philippines in italics to indicate that it may not be a subgroup at all, due to issues with the lexical evidence.

4.3 CELEBIC. Historical research in Sulawesi has succeeded in delineating subgroup boundaries within Sulawesi and its offshore islands. There are several studies that have argued for genetic relationships within subgroups on the island (see Sneddon 1993), but until relatively recently, few that argue for a larger Celebic subgroup. Mead (2003) recognizes ten individual subgroups in Sulawesi, summarized in table 7.

Of these subgroups, the first three, Sangiric, Minahasan, and Gorontalo-Mongondow, do not seem to subgroup with languages of Sulawesi. Rather, they represent the southern

13. In fact, many Philippine languages appear to support the hypothesis that PMP had regular penultimate stress with automatic gemination of stops after stressed penultimate schwa, even languages with modern accent systems. For example, Ilokano *həbás ‘evaporate’ > *ebbés* ‘subside; decrease (of rivers)’ but *sabay ‘do something with others’> *sabáy* ‘to do together’, Isnag *təbiq ‘split; section of betel nut’ > *tabbí* ‘section of a betel nut’ but *tuba ‘derris root’ > *tuba* ‘derris root’, Maranao *təbuh ‘sugar cane’ > *təbo* ‘sugar cane’ but *tuba ‘derris root’ > *toa* ‘derris root’, and Tboli *pajay > *halay* ‘rice’ but *hapəjiq > *hədek* ‘sting’.

TABLE 7. TEN LINGUISTIC GROUPS IN SULAWESI

Bungku-Tolaki	Gorontalo-Mongondow
Kaili-Pamona	Minahasan
Muna-Buton	Sangiric
South Sulawesi	Suluan-Banggai
Tomini-Tolitoli	Wotu-Wolio

extreme of the Philippine languages, part of a past southern expansion of Philippine language speakers into northern Sulawesi. Blust (1991) included Gorontalo-Mongondow in his Greater Central Philippine subgroup. From what remains, Mead (2003) argues for a Celebic supergroup that includes Tomini-Tolitoli, Kaili-Pamona, Suluan-Banggai, Bungku-Tolaki, Muna-Buton, and Wotu-Wolio, but not South Sulawesi. The evidence includes the phonological innovations in table 8.

TABLE 8. PROTO-CELEBIC SOUND CHANGES

a) *C_1C_2 > *C_2
b) *h > Ø
c) *d > *r
d) *-ay > *e
e) *-aw > *o
f) *j > *y

These sound changes are not of equal value. Most, in fact, are of rather low quality and would not on their own constitute strong subgrouping evidence. The final change, *j to *y, however, is more significant for subgrouping. Each of the above sound changes is discussed briefly below.

(a) C_1C_2 > C_2. This change concerns the reduction of consonant clusters, where the first consonant was deleted, and the second retained. It only occurred with nonnasal clusters. Clusters of nasal+obstruent were not reduced. This change is nearly universal in Island Southeast Asia, although in North Borneo languages, C_1C_2 became C_2C_2, resulting in a geminate consonant that, in some cases, may superficially appear to parallel the Celebic sound change. At any rate, this is not strong subgrouping evidence.

(b) *h > Ø. Except for some Philippine languages (Batanic), and possibly some languages in Borneo,[14] this sound change is universal in Malayo-Polynesian and cannot on its own provide strong evidence for a Celebic group.

(c) *d > *r. Again, *d > *r* is common, particularly in medial position. This sound change, however, affected *d in all positions, which, although still found in other language groups, is less common.

(d/e) *-ay > *e and *-aw > *o. The coalescence of the low vowel *a with the high-front/high-back offglide in final position is also quite common in Malayo-Polynesian. Even if this can be reconstructed to Proto-Celebic, it alone cannot provide strong evidence for a Celebic subgroup, nor can the possibility that it arose through multiple parallel innovations be ruled out.

14. In Lebo' Vo' Kenyah, for example, PMP *lahud 'toward the sea' is reflected as *laʔot* 'toward the river'. In this part of Borneo, when *h occurred between vowels of different height, it is typically reflected as a glottal stop.

(f) *j > *y. Of the six sound changes in table 8, one stands out as particularly strong, the change of PAN *j to *y. Blust (2013:578) lists 18 different reflexes of *j, and *y* is listed for only one language, Seediq of Taiwan, and here *j only became *y* in intervocalic position. In final position, *j became Seediq *c*. Earlier, it was noted that Moken also reflects *j as *y*, but like Seediq, only in intervocalic position; *j became *t* word-finally in Moken. There is, thus, good reason to assume that the change *j > *y in Celebic occurred only once and that this change supports a Celebic subgroup. The other five changes are more common, and, thus, provide only weak evidence for a Celebic subgroup.

4.4 SOUTH SULAWESI. South Sulawesi includes the languages of southwestern Sulawesi plus the Tamanic languages of interior Borneo (Adelaar 1994, 1995). Like Celebic, the strongest piece of phonological evidence for a South Sulawesi subgroup involves reflexes of PMP *j, a voiced palatalized velar stop [g^{y}]. In Proto-South Sulawesi (PSS), Mills (1975) reconstructs *z from PMP *j, a change not found in other languages in western Indonesia, although he does not discuss its value for subgrouping. PSS *z then devoiced in some South Sulawesi languages, including Buginese and Tamanic. This remains the strongest piece of evidence for South Sulawesi, as well as for grouping South Sulawesi with Tamanic.

Some South Sulawesi languages appear to reflect a change *j > *r*, for example Makasarese *pare* 'field rice' from PMP *pajay. These words, however, reflect the intermediate stage *z (Mills 1975); thus, PMP *pajay became PSS *paze before becoming Makasarese *pare*. According to this history, they do not constitute an independent example of *j > *d, as they arose through an intermediate stage *z that then became Buginese and Tamanic *s*, Makasarese and other South Sulawesi *r*.

In addition to the phonological evidence, Adelaar (1994:12–20) includes a list of irregular sound changes and lexical replacement innovations that support a South Sulawesi subgroup that includes Tamanic. Given the reconstruction of PSS *z from *j, irregular sound changes and lexical innovations, the existence of a South Sulawesi subgroup is fairly well established.

Because Sulawesi is home to two apparently unrelated subgroups, some may feel the need to combine both into a larger Sulawesi subgroup, with South Sulawesi and Celebic representing two primary branches. Although it is not necessary that a single subgroup be proposed to unite South Sulawesi and Celebic, reflexes of PMP *j leave open the possibility that these two groups might be combined. This possibility centers around a tendency for *y in intervocalic position to strengthen, producing *z*. This change is well attested throughout the AN family (for example, the language name Kadazan derives from *ka-daya-an), and if *j first became *y, and later *y strengthened to *z, it may provide a plausible scenario through which South Sulawesi and Celebic may be combined. However, it is important to note that while PMP *j became PSS *z in intervocalic position in PSS, it is reconstructed as *t in final position (Mills 1975), while reflexes of PMP *y in intervocalic position remain *y* in South Sulawesi (Makasarese *uyuŋ* < *uyun 'bundle of long objects'). Thus, although it is tempting to combine Celebic and South Sulawesi on the grounds that *j became *y, which then became *z, the history of these languages argues against this scenario. It is necessary, then, to separate South Sulawesi

and Celebic. If the two are separated, considering the uniqueness of the change PMP *j > *z and the lack of any other language with which South Sulawesi might subgroup, it is necessary to list South Sulawesi as a primary branch of MP.

4.5 CHAMORRO. Blust (2000:103) examines three views on the linguistic position of Chamorro within Austronesian. These are: (i) Chamorro is most closely related to the languages of the Philippines (essentially making Chamorro a Philippine language), (ii) Chamorro is most closely related to an unnamed language or group of languages in Indonesia, and (iii) Chamorro is not closely related to any modern Austronesian language. This paper supports the third view, that Chamorro is not closely related to any language, but with the clarification that it constitutes its own primary branch of MP.

Different arguments have placed Chamorro either within the Philippine group broadly, or have claimed a special relationship between Chamorro and one or two specific Philippine languages (Safford 1909; Topping 1973; Voegelin and Voegelin 1977; Zobel 2002), with Zobel placing Chamorro with Palauan in a Nuclear Malayo-Polynesian group originating from the southern Philippines. Such attempts at placing Chamorro within the Philippines are not widely supported, however, with Starosta and Pagotto (1991), Reid (2002), and Blust (2000) all rejecting proposals that group Chamorro with Philippine languages on morphological or morphosyntactic grounds. These earlier studies tend to point to a general similarity of Chamorro verbal morphology with that of Philippine languages. However, there is no basis for positing a subgrouping relationship between Chamorro and the Philippine languages simply due to a retention of PAN and PMP morphology.

Other arguments have assigned Chamorro to various subgroups within MP (Starosta 1995 even argues that Chamorro split off from Formosan languages, removing it from MP). All of these arguments are addressed in Blust (1999, 2000), so it is unnecessary to go into a detailed criticism here. To be clear, however, Chamorro shows unmistakable signs of inclusion in Malayo-Polynesian. This includes evidence such as *S-metathesis (Blust 1999:56), pronominal and lexical innovations, and all of the mergers that define PMP (*t/*C >*t and *n/*N >*n). Additionally, Chamorro reflects PMP *j with *ʔ*, a change found nowhere else in the Austronesian family. Based on this evidence, I take the position of Blust (2000:104) that "phonological, lexical, and morphosyntactic evidence provide no clear or widely accepted basis for assigning Chamorro to any subgroup lower than Malayo-Polynesian." This view is seconded by Reid (2002:87), who states that "Chamorro is probably a first order branch of Proto-Extra Formosan," where Extra-Formosan is comparable to Malayo-Polynesian. Reid, in the same paragraph, also endorses the view that, while Chamorro is not a Philippine language, the pre-Chamorro probably sailed to the Marianas from the Northern Philippines (a claim that is further supported by a regular reflex of *baRius 'typhoon', *pakyo*). Given the available evidence, Chamorro is best thought of as the only member of a primary branch of PMP. This view, based on linguistic grounds, is supported by archaeological evidence that dates the arrival of a Neolithic Austronesian culture on the Marianas around 3,500 BP (Craib 1993; Butler 1994; Rainbird 1994; Amesbury and Hunter-Anderson 1996) and core samples that show a significant

increase in charcoal particles as early as 4,300 BP (Kirch 2002:172), during roughly the same time that MP speakers were spreading into western Island Southeast Asia.

4.6 PALAUAN. In many respects, the history of Palauan remains a mystery. The complex historical phonology of Palauan, as detailed by Blust (2009b), contains no sound changes that straightforwardly link Palauan to any other Austronesian language, and several that make Palauan unique. These changes are too numerous and demand too much detail for a complete review here, but the basic sound correspondences are reprinted in appendix 2. While linguistic evidence points only to isolation, there are facts that may aid the search for Palauan origins. The first deals with the date of initial settlement. Archaeological records in Palau, based on the dating of pottery shards, place the first arrival of human population at about 2,000 BP. These dates, however, do not fit the linguistic data. If speakers of pre-Palauan reached Palau only 2,000 years ago, one would not expect the language to have changed so completely. Further, if Palauan originated in Island Southeast Asia, a time depth of 2,000 years would result in a comfortable nesting of Palauan within one of the established subgroups in this area. Sulawesi, for example, was settled at least 4,000 years ago. If pre-Palauan was spoken in Sulawesi and if it left only 2,000 years ago, early sound changes that define the Celebic or South Sulawesi subgroups should be present in modern Palauan, yet they are not. The best explanation for why Palauan shows no linguistic affinities with other Austronesian languages is an early split, far earlier than 2,000 BP. Kirch (2002) citing research by Ward, Athens, and Hotton (1998) offers evidence that initial settlement of Palau occurred earlier than the archaeological record shows. Sediment cores from Palau show significant increases in pollen and charcoal particles between 3,100 and 4,200 BP. Pollen and charcoal indicate early agriculture, introduced plants, and forest clearing. These dates, if correct, complement the linguistic data and indicate a likely settlement of Palau that coincides with the settlement of Borneo, Sumatra, Sulawesi, Timor, the Marianas, and Halmahera.

Beyond the archaeological evidence, however, linguistic evidence also supports the view that Palauan, like Chamorro, does not subgroup closely with any known AN language. Below, Palauan is compared to several major MP subgroups in an attempt to show that it is indeed an isolate within MP.

4.6.1 Palauan and CEMP. Palauan does not seem to belong to the CEMP subgroup. Although there are precious few phonological changes that distinguish PCEMP from PMP, Blust (1993) points out that PCEMP reduced consonant clusters in reduplicated monosyllables. Palauan did not reduce consonant clusters, but separated them by inserting a schwa,[15] as in PMP *bəjbəj 'wrap around; bind' > Palauan *bəsébəs* 'bark of tree for tying'. An apparent exception to reduplicated consonant cluster reduction in PCEMP is 'dark', which Blust reconstructs as PMP *dəmdəm > PCEMP *dəndəm with nasal place assimilation. Palauan reflects the PMP form as *kə-rəmérəm* with schwa insertion and no place assimilation. In the same paper, several CEMP lexical innovations are also listed, none of which are shared with Palauan. These include PCEMP *paniŋ/*panin 'bait' where Palauan reflects PMP *paən as *wáwl*, and PCEMP *kanzupay 'rat' where

15. Schwa insertion in this environment is rare in itself, and adds further support for Palauan as a primary branch.

Palauan reflects PMP *balabaw as *byáb*. The fact that Palauan lacks the innovations that define CEMP is significant. Taken at face value, it rules out inclusion of Palauan in a very large subgroup and significantly narrows the possible homeland.

4.6.2 Palauan and the Philippines. Palauan does not seem to subgroup with Philippine languages. None of the exclusive lexical innovations that Blust and Zorc used to define a Philippine group are found in Palauan, nor do any Palauan lexical innovations appear in Philippine languages (when compared to the word lists in Reid 1971). Although Palauan may not subgroup with Philippine languages, this does not necessarily rule out the Philippines as a possible homeland. Pre-Chamorro was likely spoken in the northern Philippines. Evidence is found in the native word for 'typhoon', *pakyo* which reflects PAN *baRius 'typhoon' and suggests a homeland within the typhoon belt (Blust 2000:106). Chamorro, like Palauan, shows no close affinity to any Austronesian language despite fairly sound evidence for a homeland in an area where Austronesian languages abound. It is possible that the expansion of an ancient group in the Philippines expelled pre-Chamorro speaking peoples from the north, and later, pre-Palauan speakers from the south. Under this scenario, Palauan and Chamorro both represent the only surviving member of different primary branches of MP.

4.6.3 Palauan and Sulawesi. The Celebic subgroup, as discussed earlier, is phonologically defined by several sound changes, the most important being *j > *y. Palauan reflects *j as *s* or *k* when in the onset of a final syllable (Blust 2009b) and, thus, cannot be placed within Celebic. In Proto-South Sulawesi, *j was likely reflected as *z, which also argues against placing pre-Palauan in Sulawesi. It is fairly clear, from the available comparative materials, that Palauan does not form a subgroup with any known AN language.

4.6.4 Palauan as an isolate within Malayo-Polynesian. Blust (2009b) provides an in-depth overview of Palauan historical phonology, where it is clear that Palauan has been developing in isolation for some time, with numerous defining sound changes. Lexical evidence only strengthens the argument for Palauan isolation. A search of basic vocabulary terms in McManus (1977) provides 57 apparent exclusive lexical innovations in Palauan (where 167 total words were investigated and reconstructed to pre-Palauan). It provides additional evidence that Palauan has had a long history of isolated development. Out of 167 words, only 63 were obvious retentions; 104, or 67 percent, were either innovations, borrowings, or historically opaque retentions. Because this number will likely be revised as the list is analyzed more precisely, it may be safer to say that between 60 and 70 percent of the Palauan basic vocabulary consists of lexical replacements of some type. This suggests that Palauan has been in isolation long enough for up to 70 percent of its vocabulary to be replaced with words that have no cognates in any other MP language. A list of the most important innovations and pre-Palauan reconstructions is located in appendix 2. Thus, Palauan, like Chamorro, is listed as an isolate within MP; the only member of a Palauan primary branch.

4.7 CENTRAL-EASTERN MALAYO-POLYNESIAN. This paper is meant to deal primarily with the validity of a WMP subgroup, the subgrouping of WMP languages, and the implications of removing WMP on the higher-order subgrouping of MP.

Therefore, less time will be spent defending a CEMP subgroup of languages. Blust (1983–84, 1993, 1999, 2009a, 2012) has for some time argued that all languages that were not included in WMP form a separate group, Central-Eastern Malayo-Polynesian, itself a primary branch of Malayo-Polynesian. Evidence for this subgroup includes regular phonological innovations, irregular innovations that are shared by all member languages, and lexical innovations—particularly those dealing with the innovation of terms referring to marsupials, a type of animal that Austronesian speakers would have been unfamiliar with when they first crossed the Wallace Line.

As recently as 2008,[16] however, Donohue and Grimes have argued that CEMP itself is invalid, on the grounds that the innovations that define the group are not found in every language in the subgroup, and are also found in languages that lie outside of CEMP. The majority of their argument is made from the perspective of CMP, which Blust (1993) has already stated probably forms a linkage, not an innovation-defined subgroup. In short, Donohue and Grimes support the view that CMP contains multiple primary branches of MP, with Eastern Malayo-Polynesian remaining a valid subgroup, itself yet another primary branch of MP (this proposal in itself resembles Blust's original 1977 proposal). Blust (2009a) replied in defense of CEMP. He pointed out that while Donohue and Grimes attacked CMP because of the incomplete attestation of sound changes in member languages, Blust himself never claimed that the proposed innovations in CMP would define a traditional subgroup. Perhaps the most important proposal from the Donohue and Grimes paper, however, is that WMP languages of Southern Sulawesi share features of CMP languages and, therefore, should be grouped with CMP, rather than WMP. To be clear, while it is beyond the scope of this paper to defend CMP, languages that belong to the Celebic group reflect the subgroup-defining change *j > *y, and, thus, are considered Celebic, not part of any CMP group.

Schapper (2011) also argues against a key part of Blust's CEMP evidence, by claiming that PCEMP *kandoRa 'cuscus' is instead restricted to only EMP, and that PCEMP *mans(aə)r, while still assignable to PCEMP, also meant 'cuscus', and not 'bandicoot'. Blust (2012) replied to Schapper, but because Schapper's claims do not directly challenge the validity of CEMP, they do not demand an in-depth review here.

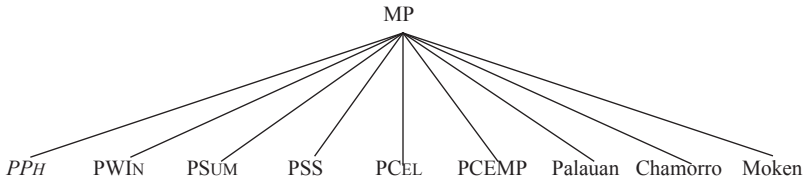
Although CMP continues to pose problems specifically because of the incomplete dispersal of relevant sound changes throughout the subgroup, the question of the validity of CMP is best discussed separately. Also, even if CMP is said to consist of several smaller subgroups with no special relationship to each other (if CMP is to be done away with), it remains to be seen if these languages would be assigned directly to MP as Donohue and Grimes suggest, or if they are still best considered descendants of PCEMP. This paper does not seek to question the validity of CEMP, and keeps with Blust (1983–84, 1993, 1999, 2009a, 2012) in listing it as a primary branch of MP.

4.8 PUTTING THE EVIDENCE TOGETHER. Island Southeast Asia was settled from the northern Philippines in what appears to have been a single population expansion. The Western Malayo-Polynesian group, which has traditionally contained all non-CEMP languages of Island Southeast Asia, cannot possibly have existed, as this

16. Ross (1995), Grimes et al. (1997), and Adelaar (2005a) also question CEMP or CMP, but Donohue and Grimes (2008) represents the most thorough argument against its validity.

population expansion occurred too rapidly over too great an area. The settlement patterns of this expanding population and the linguistic evidence summarized throughout this section support a rake model of the type shown in figure 10. Here, at least nine primary branches of MP are proposed, with the Philippines in italics to indicate that they may represent more than one primary subgroup or perhaps, an innovation-defined linkage.

FIGURE 10. MALAYO-POLYNESIAN HIGHER-ORDER SUBGROUPING



5. BEYOND SUBGROUPING: IMPLICATIONS FOR PMP LEXICAL RECONSTRUCTION. After WMP is removed from the AN family tree, a situation arises where the large number of PWMP reconstructions suddenly become eligible for reconstruction to PMP. The issue at hand is the fate of the more than 6,750 PWMP reconstructions listed in the *Austronesian comparative dictionary* (Blust and Trussel ongoing). An immediate problem faced by any researcher in this area will be justifying PMP reconstructions with evidence from primary branches in what used to be the WMP subgroup. Some of these primary branches, particularly Western Indonesian, Sumatran, Celebic, South Sulawesi, and the languages of the Philippines, are in an area of widespread contact. Malay has played a particularly important role in this area, and is responsible for lexical borrowings throughout these subgroups. Moreover, many of the PWMP reconstructions in Blust and Trussel (ongoing) are already listed as PMP reconstructions. For example, a quick glance at the list of PWMP reconstructions reveals words like PWMP *anak 'child', *babuy 'pig', and *daRaq 'blood', which are also listed at the PAN and PMP levels. It is, thus, not possible to directly transfer all 6,750 PWMP reconstructions to PMP. Time must be taken to go through the list, and make qualitative judgments about each form. Such a task is obviously beyond the scope of this paper, but it is something that must be done as the model for the breakup and diversification of PMP continues to change.

6. CONCLUSION. In this paper, I have attempted to demonstrate how a binary branching model of Malayo-Polynesian higher-order subgrouping fails to capture the historical facts about AN migration into Island Southeast Asia and the Pacific. It was argued that PMP spread too quickly over too great a geographical area for only two primary branches to have developed. This view is supported by countless statements in the literature that explicitly or implicitly remove WMP from the Austronesian family tree. In its place, seven primary branches were proposed: Western Indonesian, Sumatran, Celebic, South Sulawesi, Moken, Chamorro, and Palauan. Also, the languages of the Philippines have been grouped together in an italicized *Philippine* group, but it is understood that the languages of the Philippines may represent multiple, yet unnamed, primary branches of

Malayo-Polynesian or that they may better be organized into a linkage, not a subgroup. Central-Eastern Malayo-Polynesian was preserved as an eighth primary branch.

Because the settlement of Island Southeast Asia took place over too short a time for only one subgroup to represent all languages, it was necessary to take a fresh look at the westernmost languages of Island Southeast Asia. In western Indonesia, it was shown that at least the islands of Borneo and Sumatra were settled by groups whose closest common ancestor is PMP itself, with Moken representing a third group of unknown origin. This runs contrary to claims in Blust (2010) that all languages of western Indonesia—excluding the languages of Sulawesi—belong to a single subgroup, but it complements the archeological evidence that shows that forest clearing took place on Sumatra as early as any other location in Island Southeast Asia. The presence of a discrete group of languages on Sumatra does, however, lend support to Blust's claim that the geography of Island Southeast Asia played a major role in the development of linguistic subgroups (Blust 2010:47–48).

A new avenue of lexical reconstruction was revealed, where primary branches of MP in what used to be the WMP subgroup may now provide evidence for PMP reconstructions without CEMP evidence. However, because of the realities of contact among former WMP languages, great care must be taken when reconstructing new PMP forms to avoid the reconstruction of borrowed words.

APPENDIX 1. REFLEXES OF *-j- AND *-d- IN "WESTERN" MALAYO-POLYNESIAN LANGUAGES. (Data from Blust and Trussel ongoing)

Geographical location	Language	*-j- reflex	*-d- reflex	Subgroup
	Itbayaten	r	r	Batanic
	Ilokano	g	d	
	Atta	g	d/r	Northern Luzon
	Ifugaw	g	d	
Philippines (Sangir and Tontemboan are spoken in northern Sulawesi)	Bikol	d/r	d/r	
	Subanun	l/d	l/d	GCPH
	Hiligaynon	l	l	
	Tboli	l/d	l/d	Bilic
	Sangir	r/d	r/d	Sangiric
	Kalamian Tagbanwa	r	r	Kalamian
	Tontemboan	r	r	Minahasan
Borneo	Kelabit	d	d	Western Indonesian
	Lepo Tau	d/t	d/t	
	Kayan	r	r	
	Aoheng	r	r	
	Kanowit	d	d	
	Maanyan	r	r	
Java/Bali/Lombok/ Sumbawa	Sundanese	r	r	
	Javanese	r/d	r/d	
	Balinese	r/d	r/d	
	Sasak	r/d	r/d	
	Sumbawa	d	d	
Sumatra	Rejang	d	d	
	Lampung	x	x	

	Enggano	h	d	
	Nasal	g/h	d	
Sumatra	Simeulue	x	d	Sumatran
	Mentawai	g	r	
	Nias	x	d	
western Thailand	Moken	y	d	Moken
	Uma	j [ʤ]	r	
	Muna	Ø	r	Celebic
Sulawesi	Banggai	y	d	
	Buginese	s	r	South Sulawesi
	Embaloh	s	r	
Palau	Palauan	s/k	r	Palauan
Mariana Islands	Chamorro	ʔ	h	Chamorro

APPENDIX 2. PALAUAN LEXICAL REPLACEMENT INNOVATIONS AND PRE-PALAUAN RECONSTRUCTIONS

The following list of Palauan-only lexical innovations does not on its own provide evidence that Palauan is a primary branch of MP. However, it does make certain points about the history of Palauan. The long list of innovations that have no counterparts in any other AN languages adds evidence to the hypothesis that Palauan has been developing in isolation for some time. In particular, it agrees more with the pollen samples suggesting that the first arrival of Neolithic cultures occurred closer to 4,000 BP. The list is preceded by a table, based on the work in Blust (2009b), that identifies reflexes of PMP phonemes in Palauan, which aids in the reconstruction of pre-Palauan lexicon. The main goal in reconstructing pre-Palauan is to aid in the search for cognates of apparent Palauan lexical innovations in other phonologically more conservative MP languages. Note that, in the list below, I also indicate where a lexical replacement has taken place: for example, PMP *qulu > Ø shows that *qulu 'head' was lost. Lexical replacement innovations are the most powerful type of lexical evidence, and noting them here is necessary. Also, because Palauan reflects a number of mergers, there are often multiple possible reconstructions for a single lexeme. I have listed all possible reconstructions separately. For example, I list *bərVt, *bədVt, *bəyVt, *bəRVt, *bəzVt rather than *bə[rdyRz]Vt as the possible reconstructions of 'lips'.

Reflexes of PMP phonemes in Palauan are as follows:

PMP	Palauan	PMP	Palauan
*í	í	*q	ʔ
*i̱	ə; Ø / _#	*b	b
*ú	u	*d	r
*u̱	ə; Ø / _#	*z	r; l / _l
*á	a	*j	s; k / _(v)C#
*a̱	ə; Ø / _#	*ṉ	l
*ə́	o; e	*ñ	l
*ə	ə ~ Ø	*m	m
*-ay	-Ø	*ŋ	ŋ; Ø / C(v)_#
*-aw	-Ø	*s	t
*-uy	-Ø	*h	Ø
*p	w	*l	y
*pa-	o-	*r	r
*-ap	-o	*R	s; r / _{t,s}
*-əp	-o	*w	w
*t	ð; t / b(v)_; t / [+stop, -voi](v)_#	*y	r
*k	k		

‘head’
PMP *qulu > Ø
Pre-Palauan *bVtun(V) > Palauan *bəðúl*

‘skin’
PMP *kulit > Ø
Pre-Palauan *butV-n > Palauan *buðél*
cf: *məlábð* < *maŋ-tabutV[17] ‘to skin’
In modern Palauan, stress typically shifts to the final syllable when suffixed with the possessive final *l*. McManus (1977) lists this word as obligatorily possessed.

‘lips’
PMP *bibiR > Ø
Pre-Palauan *bərVt, *bə́dVt, *bə́yVt, *bə́RVt, *bə́zVt > Palauan *bərð-él*, *bərðáwl*
There is a clear disyllabic base form, *bərð*. Final *-awl* in *berðawl* is inexplicable, but the possessed form *bərð-él* reveals the base.

‘tooth’
PMP *ŋipən > Ø
Pre-Palauan *píŋən > Palauan *wíŋəl*
This innovation involves metathesis rather than lexical replacement. Nevertheless, it is unique to Palauan.

‘tongue’
PMP *zəlaq > Ø
Pre-Palauan *qúra, *qúda, *qúya, *qúza > Palauan *ʔur, ʔurál* (possessed form)
The possessed form *ʔurál* supports reconstructing final *-a*.

‘yawn’
PMP *huab > Ø
Pre-Palauan *wajV, *waRV > Palauan *was*

‘body hair’
PMP *bulu > Ø
Pre-Palauan *bújuq, *búRuq > *búsəʔ*, *bsúʔəl*
The addition of verbal morphology and subsequent stress shift in *bsúʔ-əl* supports the reconstruction of *u in the ultimate syllable.

‘neck’
PMP *liqər > Ø
Pre-Palauan *qíkVn, *qíjVn > Palauan *ʔíkl*

‘nape’
PMP *batuk, *təŋuk > Ø
Pre-Palauan *paŋ-bətá-an > Palauan *omðáəl*, *oméð* ‘to catch, grab, pick up’, *milléð*, *með*.
According to McManus, the word for ‘nape’ is related to the verb ‘to catch, carry, pick up’, perhaps by association with how a mother animal might carry her offspring by grasping the nape. This supports the reconstruction of a morphologically complex protoform, with **bəta* as the base. Stative forms such as *bléð* support reconstructing initial *b*.

‘bone’
PMP **tuqəlaŋ* > *ðəʔóyl* ‘backbone’
Pre-Palauan *qVlúlVR, *qVlúlVj > Palauan *ʔyúys* ‘bone’, *ʔyuəsél* (possessed form)

17. A capital V represents a vowel of unknown quality that can nevertheless be reconstructed for pre-Palauan. However, in final position, V may indicate either a deleted final vowel, a deleted final vowel-glide sequence, or a deleted final vowel-*ŋ* sequence. It is important to keep this in mind while reading through the list.

Although Palauan retains a reflex of PMP **tuqəlaŋ*, its meaning has narrowed to specifically 'back bone'. The broader category 'bone' is represented by the innovation *ʔyúys*. Also, in Palauan, it appears that final *ŋ was irregularly reflected as pre-Palauan *n, which became modern Palauan *l*.

'back'

PMP *likud > Ø
Pre-Palauan *wúnVk, *wúñVk, *púnVk, *púñVk > Palauan *wúlk*

'heart'

PMP *pusuq > Ø
Pre-Palauan *rəŋu, *də́ŋu, *zə́ŋu > Palauan *reŋ*, *rəŋúl* (possessed form)
The possessed form supports reconstructing final *u* in the Pre-Palauan word.

'elbow'

PMP *siku > Ø
Pre-Palauan *buku-n a qalíma > Palauan *bkúl a ʔím*
The individual parts of this word are not innovations. *buku and *qalima are fairly widely attested. However, it is interesting that the PMP word for elbow, **siku*, was replaced by what is essentially a phrase, 'joint of the arm'.

'flesh'

PMP *həsi, *isiʔ > Ø
Pre-Palauan Rə́qV > Palauan *seʔ*

'urine'

PMP *miqmiq, *iḥiq, *sibu > Ø
Pre-Palauan *qVmə́qim > Palauan *ʔəmóʔəm*; *oŋ-əməʔím-əl* < *paŋ-qVmə́qim-an 'urinal'
The suffixed form *oŋ-əməʔím-əl* supports reconstructing *i* in the ultimate syllable.

'head hair'

PMP *buhək > Ø
Pre-Palauan *qulu > Palauan *ʔuy*
This reconstruction is potentially quite interesting. If accurate, this is, as far as I can tell, the only case of a semantic shift from *qulu* 'head' to *qulu* 'head hair'.

'cheek'

PMP *piŋi, *pipi > Ø
Pre-Palauan *pasáŋV > Palauan *otáŋ*

'bird'

PMP *manuk-manuk > Ø
Pre-Palauan **Rə́bVk-an qayam* > Palauan *suébək əl ʔarm* '(lit.) flying animal'
Inexplicable *u* in *suebek*, but compare *olsébək* < *pan-sebVk 'to make fly', *sobəkaŋ* 'starting to fly', *sébək reŋ* 'worry, (lit.) flying heart' and *sebəsəbək* (redup.)
Like *bkul a ʔim*, this is a puzzling case where, for no discernible reason, the single word for 'bird' was replaced by a phrase 'flying animal'. Note that pre-Palauan **qayam* reflects PAN **qayam* 'domesticated animal'. *suébək* reflects PMP *Rəbək, according to the *Austronesian comparative dictionary*, but is only attested outside of CEMP in Palauan and Chamorro.

'crocodile'

PMP *buqaya > Ø
Pre-Palauan *lujV, *luRV > Palauan *yus*
Crocodiles live on Palau and are native, so the loss of *buqaya cannot readily be explained.

'beetle'

PMP *abuqaŋ (very limited distribution) > Ø
Pre-Palauan *qVbərVt-bərVt > Palauan *ʔəbərðórəð*

‘hundred’
PMP *Ratus
Pre-Palauan *taRus > Palauan *ðart*
This does not constitute a replacement. However, the metathesis shown here appears to be unique among all Austronesian languages.

‘loincloth’
PMP *bahaR > Ø
Pre-Palauan *?-Rakəd, *?-Rakər > Palauan *usakər*
The initial *u* in *usakər* is inexplicable.

‘canoe’
PMP *qabaŋ > Ø
Pre-Palauan *mVnálV > Palauan *mlá́y*

‘when’
PMP *ijaŋ > Ø
Pre-Palauan *palV ŋaraŋV > Palauan *oiŋará́ŋ*

‘where’
PMP *pai, *inu > Ø
Pre-Palauan *kərV, *kədV, *kəyV, *kəzV > Palauan **kər*

‘what’
PMP *apa, *a-nu > Ø
Pre-Palauan *ŋará́ŋV, *ŋadá́ŋV > Palauan *ŋará́ŋ*

‘or’
PMP *?
Pre-Palauan *a nVqú́bV > Palauan *a ləʔú́b*

‘fruit’
PMP *buaq > Ø
Pre-Palauan *rə́tVq, *də́tVq, *zə́tVq > Palauan *ró́ðəʔ*

‘bamboo’
PMP *buluq > Ø
Pre-Palauan *qVRənV > Palauan *ʔəsél*

‘thorn’
PMP *duRi >Ø
Pre-Palauan *səqVt > Palauan *tóʔəð*

‘land’
PMP *tanəq > Ø
Pre-Palauan *qusVm > Palauan **ʔutəm*

‘grass’
PMP *udu, *dukut > Ø
Pre-Palauan *qú́tVn > Palauan *ʔú́ðəl*

‘saltwater’
PMP *tasik > Ø
Pre-Palauan *tapVb, *tawVb > Palauan *ðawb* ‘ocean, salt water’

‘thick’
PMP *kapal > Ø
Pre-Palauan *pakVr-pakVr, *pakVd-pakVd > Palauan *okró́kr* ‘thick, of flat objects’
Pre-Palauan *kVtənVR, *kVtənVj > Palauan *kəðó́ls* ‘thick, of round objects’

‘thin’
PMP *-ipis, *-əpis > Ø
Pre-Palauan *qVsVŋálVt > Palauan *ʔətŋayð*

‘left’
PMP *ka-wiRi > Ø
Pre-Palauan *kasurV, *kasudV, *kasuzV, *kasuyV > Palauan *katúr*

‘right’
PMP *taqu, *ka-wanan > Ø
Pre-Palauan *katikVm > Palauan *kaðíkm*

‘near’
PMP *adani, *hamper > Ø
Pre-Palauan *k<um>əlVt > Palauan *kméyð*
The original base was *kəlVt, which is reflected by *kéyð* ‘nearness’.

‘far’
PMP *zauq > Ø
Pre-Palauan *qVrəlVt, *qVdəlVt, *qVyəlVt, *qVzəlVt > Palauan *ʔəróyð*

‘wet’
PMP *besaq, *baseq > Ø
Pre-Palauan *kV-RəmVk-Rə́mVk > Palauan *kəsəməsémək*
Pre-Palauan *tVkímVR, *tVkíməj > Palauan *ðəkíməs*

‘black’
PMP *ma-qitəm > Ø
Pre-Palauan *qV-tənVk-tənVk > Palauan *ʔəðələkélək*
McManus does not indicate that this word is synchronically reduplicated. There is no *ʔəðələk* or *ðələk* listed in the dictionary. It must be assumed that the obvious reduplication in the reconstructed proto-form was lost or that it was borrowed.

‘cry’
PMP *taŋis > Ø
Pre-Palauan *naŋVn > Palauan *láŋəl*

REFERENCES

Adelaar, K. Alexander. 1994. The classification of the Tamanic languages (West Kalimantan). In *Language contact and change in the Austronesian world*, ed. by Tom Dutton and Darrell T. Tryon, 1–41. Trends in Linguistics Studies and Monographs 77. Berlin: Mouton de Gruyter.

———. 1995. Borneo as a cross-roads for comparative Austronesian linguistics. In *The Austronesians: History and comparative perspectives*, ed. by Peter Bellwood, James J. Fox, and Darrell Tryon, 75–95. Canberra: Research School of Pacific and Asian Studies, Australian National University.

———. 2005. The Austronesian languages of Asia and Madagascar: A historical perspective. In *The Austronesian languages of Asia and Madagascar*, ed. by K. Alexander Adelaar and Nikolaus Himmelmann, 1–43. Routledge Language Family Series 7. London: Routledge.

Amesbury, J., D. R. Moore, and R. L. Hunter-Anderson. 1996. Sea level change in the Marianas: Recent excavations at Chalan Piao, Saipan, Micronesia. *Bulletin of the Indo-Pacific Prehistory Association* 15:53–69.

Anderbeck, Karl, and Herdian Aprilani. 2013. *The improbable language: Survey report on the Nasal language of Bengkulu, Sumatra.* SIL Electronic Survey Report 2013-012, August 2013. SIL International.

Bellwood, Peter. 1984-85. A hypothesis for Austronesian origins. *Asian Perspectives* 26:107–17.

———. 1988. *Archaeological research in south-eastern Sabah*. Kota Kinabalu: Sabah Museum and State Archives.

———. 1989. Archaeological investigations at Bukit Tengkorak and Segarong, south-eastern Sabah. *Bulletin of the Indo-Pacific Prehistory Association* 9, 122–62.

———. 2007. *Prehistory of the Indo-Malaysian Archipelago*, revised ed. Canberra: Australian National University E Press.

Blake, Frank R. 1902. Study of Philippine languages at Johns Hopkins University. *American Anthropologist* 4:793–94.

———. 1920. A bibliography of the Philippine languages. *Journal of the American Oriental Society* 40: 25–70.

Blumentritt, Ferdinand. 1899. List of the native tribes of the Philippines and of the languages spoken by them. Translated with an introduction and notes by O. T. Mason. *Annual Report of the Smithsonian Institution for 1899*, 527–47. Washington, D.C.: Smithsonian Institution.

Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *Working Papers in Linguistics* 9(2):1–15. Honolulu: Department of Linguistics, University of Hawai'i.

———. 1982. The linguistic value of the Wallace Line. *Bijdragen tot de Taal-, Land-, en Volkenkunde* 138:231–50.

———. 1983–84. More on the position of the languages of eastern Indonesia. *Oceanic Linguistics* 22/23:1–28.

———. 1984–85. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26:45–67.

———. 1991. The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30:73–129.

———. 1995. The position of the Formosan languages: Method and theory in Austronesian comparative linguistics. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng, 585–650. Symposium Series of the Institute of History and Philology, Academia Sinica, No. 3. Taipei: Academia Sinica

———. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul J. K. Li, 31–94. Symposium Series of the Institute of Linguistics, Academia Sinica 1. Taipei: Academia Sinica.

———. 2000. Chamorro historical phonology. *Oceanic Linguistics* 39:83–122.

———. 2004. *t to *k*: An Austronesian sound change revisited. *Oceanic Linguistics* 43:365–410.

———. 2005. The linguistic macrohistory of the Philippines: Some speculations. In *Current issues in Philippine linguistics and anthropology: Parangal kay Lawrence A. Reid*, ed. by Hsiu-chuan Liao and Carl R. Galves Rubino, 31–68. Manila: Linguistic Society of the Philippines and SIL International.

———. 2009a. The position of the languages of eastern Indonesia: A reply to Donohue and Grimes. *Oceanic Linguistics* 48:36–77.

———. 2009b. Palauan historical phonology: Whence the intrusive velar nasal? *Oceanic Linguistics* 48:307–46.

———. 2010. The Greater North Borneo hypothesis. *Oceanic Linguistics* 49:44–118.

———. 2012. The marsupials strike back: A reply to Schapper (2011). *Oceanic Linguistics* 51:261–77.

———. 2013. *The Austronesian languages*, 2nd ed. Canberra: Pacific Linguistics.
———. 2014. Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics* 53:300–91.
Blust, Robert, and Stephen Trussel. Ongoing. *Austronesian comparative dictionary*. Online: http://www.trussel2.com/*Austronesian Comparative Dictionary*.
Bulbeck, David. 1992. A tale of two kingdoms: The historical archaeology of Gowa and Tallok, South Sulawesi, Indonesia. PhD thesis, Australian National University.
Butler, B. 1994. Early prehistoric settlements in the Mariana Islands. *Man and Culture in Oceania* 10:15–38.
Charles, Mathew. 1974. Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. *Oceanic Linguistics* 13:457–509.
Craib, J. 1993. Early occupation at Unai Chulu, Tinian. *Bulletin of the Indo-Pacific Prehistory Association* 13:116–34.
Dahl, Otto Chr. 1981. *Early phonetic and phonemic changes in Austronesian*. Oslo: Instituttet for Sammenlignende Kulturforskning.
Donohue, Mark, and Charles E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia and East Timor. *Oceanic Linguistics* 47:114–58.
Dyen, Isidore. 1962. The lexicostatistical classification of the Austronesian languages. *Language* 36:38–46.
———. 1965. *A lexicostatistical classification of the Austronesian languages*. Indiana University Publications in Anthropology and Linguistics, and Memoir 19 of the International Journal of American Linguistics. Baltimore: The Waverly Press.
———. 1995. Borrowing and inheritance in Austronesianistics. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng, 455–520. Symposium Series of the Institute of History and Philology 3. Taipei: Academia Sinica.
Ferguson, Charles. 1966. Assumptions about nasals: A sample study in phonological universals. 2nd ed. In *Universals of language*, ed. by Joseph H. Greenberg, 53–60. Cambridge, MA: MIT Press.
Flenley, John. 1988. Palynological evidence for land use changes in South-East Asia. *Journal of Biogeography* 15:185–97.
François, Alexandre. 2014. Trees, waves, and linkages: Models of language diversification. In *The Routledge handbook of historical linguistics,* ed. by Claire Bowern and Bethwyn Evans, 161–89. London and New York: Routledge.
Glover, Ian. 1976. Ulu Leang cave, Maros: A preliminary sequence of post-Pleistocene cultural development in South Sulawesi. *Archipel* 11:113–54.
———. 1977a. The late stone age in eastern Indonesia. *World Archaeology* 9:42–61.
———. 1977b. Prehistoric plant remains from Southeast Asia, with special reference to rice. In *South Asian archaeology*, ed. by M. Taddei, 7–37. Naples: Istituto Universitario Orientale.
Grimes, Charles E., Tom Therik, Barbara Dix Grimes, and Max Jacob. 1997. *A guide to the people and languages of Nusa Tenggara.* Paradigma B–1. Kupang: Artha Wacana Press.
Hudson, Alfred B. 1978. Linguistic relations among Bornean peoples with special reference to Sarawak: An interim report. *Studies in Third World Societies* 3:1–44.
Kirch, Patrick. 2002. *On the road of the winds: An archaeological history of the Pacific islands*. Berkeley: University of California Press.
Liao, Hsiu-chuan. 2011. On the development of comitative verbs in Philippine languages. *Language and Linguistics* 12:205–37.
Llamzon, Teodoro A. 1975. Proto-Philippine phonology. *Archipel* 9(1):29–42.
Llamzon, Teodoro A., and Ma. Teresita Martin. 1976. A subgrouping of 100 Philippine languages. In *South-East Asian linguistic studies*, vol. 2, ed. by Nguyen Dang Liem,

141–72. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
Lobel, Jason William. 2013. Philippine and North Bornean languages: Issues in description, subgrouping, and reconstruction. PhD diss., University of Hawai'i.
———. 2016. *North Borneo sourcebook: Vocabularies and functors.* PALI Language Texts. Honolulu: University of Hawai'i Press.
Lopez, Cecilio. 1967. Origins of the Philippine languages. *Philippine Studies* 15:130–66.
McManus, Edwin. 1977. *Palauan–English dictionary.* Honolulu: University of Hawai'i Press.
Mead, David. 2003. Evidence for a Celebic supergroup. In *Issues in Austronesian historical phonology*, ed. by John Lynch, 115–41. Canberra: Pacific Linguistics.
Mills, Roger. 1975. Proto-South Sulawesi and Proto-Austronesian phonology. PhD diss., University of Michigan.
Nothofer, Bernd. 1986. The Barrier Islands languages in the Austronesian language family. In *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, ed. by Paul Geraghty, Lois Carrington, and S. A. Wurm, 87–109. Canberra: Pacific Linguistics.
Pawley, Andrew. 1999. Chasing rainbows: Implications for the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 95–138. Taipei, Taiwan: Academia Sinica.
———. 2006. Explaining the aberrant Austronesian languages of southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 115:215–58.
Paz, Consuelo J. 1981. *A reconstruction of Proto-Philippine phonemes and morphemes*. Diliman, Quezon City: Cecilio Lopez Archives and Philippine Linguistic Circle.
Rainbird, Paul. 1994. Prehistory in the northwest tropical Pacific: The Caroline, Mariana, and Marshall Islands. *Journal of World Prehistory* 8(3):293–349.
Reid, Lawrence A. 1971. *Philippine minor languages: Wordlists and phonologies*. Oceanic Linguistics Special Publication 8. Honolulu: University of Hawai'i Press.
———. 1978. Problems in the reconstruction of Proto-Philippine construction markers. In *Second International Conference on Austronesian Linguistics: Proceedings*, Fascicle I: *Western Austronesian*, ed. by S. A. Wurm and Lois Carrington, 33–66. Canberra: Pacific Linguistics.
———. 1982. The demise of Proto-Philippines. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2, ed. by Amran Halim, Lois Carrington, and S. A. Wurm, 201–16. Canberra: Pacific Linguistics.
———. 2002. Morphosyntactic evidence for the position of Chamorro in the Austronesian language family. In *Collected papers on Southeast Asian and Pacific languages*, ed. by Robert S. Bauer, 63–94.Canberra: Pacific Linguistics.
———. 2010. Palauan velar nasals and the diachronic development of PMP noun phrases: A response to Blust. *Oceanic Linguistics* 49:437–77.
Ross, Malcolm. 1992. The sound of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics* 31:23–64.
———. 1994. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary*, ed. by Darrell Tryon, 45–120. Berlin: Mouton de Gruyter.
———. 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2):1–24.
———. 2009. Proto-Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. by Alexander Adelaar and Andrew Pawley, 295–326. Canberra: Pacific Linguistics

Safford, William E. 1909. *The Chamorro language of Guam*. Washington D.C.: W. H. Lowdermilk and Co.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43:411–44.

———. 2013. The higher phylogeny of Austronesian: A response to Winter. *Oceanic Linguistics* 52:249–55.

Schapper, Antoinette. 2011. Phalanger facts: Notes on Blust's marsupial reconstructions. *Oceanic Linguistics* 50:258–72.

Smith, Alexander D. 2017a. The languages of Borneo: A comprehensive classification. PhD diss., University of Hawai'i.

———. 2017b. Merap historical phonology in the context of a central Bornean linguistic area. *Oceanic Linguistics* 56:143–80.

Sneddon, James. 1993. The drift towards final open syllables in Sulawesi languages. *Oceanic Linguistics* 32: 1–44.

Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng, 683–726. Symposium Series of the Institute of History and Philology 3. Taipei: Academia Sinica.

Starosta, Stanley, and Louise Pagotto. 1991. The grammatical genealogy of Chamorro. In *VICAL 2*, *Papers from the Fifth International Conference on Austronesian Linguistics: Western Austronesian and contact languages*, ed. by Ray Harlow and Robin Hooper, 319–348. Te Reo Special Publication. Auckland: Linguistic Society of New Zealand.

Stokhof, W. A. L., ed. 1987. *Holle Lists: Vocabularies in languages of Indonesia: Southern Sumatra* 10(2). Canberra: Pacific Linguistics.

Stuijts, Ingelise. 1993. *Late Pleistocene and Holocene vegetation of West Java*, *Indonesia*. Rotterdam: Balkema.

Thurgood, Graham.1999. *From ancient Cham to modern dialects: Two thousand years of language contact and change: With an appendix of Chamic reconstructions and loanwords*. Oceanic Linguistics Special Publication 28. Honolulu: University of Hawai'i Press.

Topping, Donald M. 1973. *Chamorro reference grammar*. Honolulu: University of Hawai'i Press.

Tryon, Darrell. 1995. Proto-Austronesian and the major Austronesian subgroups. In *The Austronesians*, ed. by Peter Bellwood, James J. Fox, and Darrell Tryon, 19–42. Canberra: The Australian National University Press.

Voegelin, C. F., and F. M. Voegelin. 1977. *Classification and index of the world's languages*. New York: Elsevier.

Ward, J. V., J. S. Athens, and C. Hotton. 1998. Holocene pollen records from Babeldaob island, Palau, Western Caroline islands. Paper presented at the Annual Meeting of the Society for American Archaeology, Seattle, March 29.

Wolff, John.1973. Verbal inflection in Proto-Austronesian. In *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, ed. by Andrew B. Gonzales, 71–91. Quezon City: Linguistic Society of the Philippines.

———.1995. The position of the Austronesian languages of Taiwan within the Austronesian group. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng, 521–83. Symposium Series of the Institute of History and Philology 3. Taipei: Academia Sinica.

Zobel, Erik. 2002. The position of Chamorro and Palauan in the Austronesian family tree: Evidence from verb morphology and morphosyntax. In *The history and typology of western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross, 405–34. Canberra: Pacific Linguistics.

Zorc, David. 1983. Proto Austronesian accent revisited. *The Philippine Journal of Linguistics* 14(1):1–24.
———. 1986. The genetic relationships of Philippine languages. In *FOCAL 2: Papers from the Fourth International Conference on Austronesian Linguistics*, ed. by Paul Geraghty, Lois Carrington, and S. A. Wurm, 147–73. Canberra: Pacific Linguistics.

smithad@hawaii.edu