Evidence and Models of Linguistic Relations: Subgroups, Linkages, Lexical Innovations, and Borneo
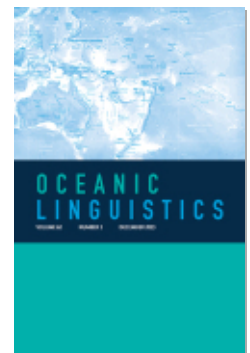
Alexander D. Smith

Evidence and Models of Linguistic Relations: Subgroups, Linkages, Lexical Innovations, and Borneo

Alexander D. Smith

NATIONAL UNIVERSITY OF SINGAPORE

Several recent studies place the languages of Borneo into one of two large groups, the Greater North Borneo subgroup and the Barito–Basap linkage. These same studies place both Greater North Borneo and Barito–Basap with the Western Indonesian subgroup, a large subgroup which is claimed to be a primary branch of Malayo-Polynesian. This paper demonstrates that the exclusively lexical evidence used to justify such subgroups is invalid as subgrouping evidence. Instead, it is shown that the languages of Borneo developed a small number of Bornean-only lexical items through contact, borrowing, and early innovations within the first Proto-Malayo-Polynesian-speaking settlers of the island. To support these claims, a detailed description of both the methods of lexical innovation evaluation as well as the types of linguistic relations that such lexical innovations support is undertaken in this paper. A new standard for the use of lexical evidence in subgrouping arguments is established, with wide-ranging implications for not only the classification of Bornean languages but of western Malayo-Polynesian languages in general.

Keywords: Historical; Subgrouping; Linkage; Borneo

1. INTRODUCTION.[1] Recent trends in Malayo-Polynesian (MP) subgrouping have seen a shift away from the traditional binary-split model of Blust (1977, 1993) toward a less-hierarchically structured, more rake-like representation (Adelaar 2005; Ross 2005; Smith 2017a). These more recent proposals may be described as "splitting," rather than "lumping" approaches and are concerned mostly with higher-order divisions, those that deal with the immediate descendants of Proto-MP (PMP). Such changes are not arbitrary and have been motivated both by a reevaluation of available evidence and by a general reconciliation of linguistic subgrouping with the rapid movement of

---

Austronesian (AN)-speaking people into Island Southeast Asia, as such rapid expansions favor rake-like, rather than nested, family trees (Smith 2017a).

The shift from lumping to splitting at higher-order divisions has mostly taken place within a theory of linguistic relations that remains heavily reliant on the traditional subgrouping model (one with discrete proto-languages, innovations that originate from those proto-languages, and clear genealogical divisions). Even in the more recent proposal of Smith (2017a), a now rake-like MP family tree with nine equidistant divisions is mostly comprised of traditional subgroups.[2] Some of these first-order divisions encompass hundreds of languages yet rely exclusively or near exclusively on lexical evidence, including such large subgroups as the Philippine and Western Indonesian (WIN) subgroups.

The reliance on lexical evidence in the formation of these large first-order subgroups has generated recent controversy, especially with regard to the validity of the Philippine subgroup (Recent defenses of the Philippine subgroup include Blust 2019; Blust 2020; Zorc 2020, while recent skepticisms include Liao 2020; Reid 2020; Ross 2020; Smith 2017a.) The other major lexically defined first-order MP subgroup, Western-Indonesian, has received less dedicated attention in published literature, but it is nevertheless similar to the Philippine subgroup in its near total reliance on lexical innovations as core evidence.

The current subgrouping situation clearly needs to be reevaluated, and this paper attempts to address these issues from the perspective of Bornean higher-order subgrouping in the context of the first arrival of MP-speaking migrants from the Philippines. This paper therefore argues for a less structured subgrouping model in Borneo due to less emphasis being placed on lexical evidence. It argues that the large, lexically defined subgroups of Greater North Borneo (GNB) and WIN are not valid subgroups but rather early zones of contact and lexical diffusion between languages and dialects. This argument is based on three main theses about subgrouping and lexical evidence, with a special focus on AN languages in Insular Southeast Asia (ISEA):

1. Models of linguistic relations must incorporate multiple types of relatedness that include descendance from a common ancestor, slow differentiation in a network, and contact. In Borneo, the following are observed:
   a. Innovation-defined subgroups (descending from an exclusive common ancestor).
   b. Innovation-defined linkages (descending from a dialect network).
   c. Innovation-defined zones of lexical diffusion (contact/borrowing).
   d. Innovation-defined nonexclusive lexical groups (inheritance from a nonexclusive ancestor, that is, borrowing or innovation within a PMP-speaking settler community).

---

2. PAN, AN = (Proto-)Austronesian. PMP, MP = (Proto-)Malayo-Polynesian. PWIN, WIN = (Proto-)Western Indonesian. PGNB, GNB = (Proto-)Greater North Borneo. PBB, BB = (Proto-)Barito–Basap.

2. Lexical innovations, which are heavily relied upon in Bornean higher-order subgrouping, must be divided into different types according to their relative strength (listed below in order from strongest to weakest):

    a. Exclusively shared lexical replacement innovations.

    b. Novel concept lexical innovations.

    c. Homonymic lexical innovations.

3. The evaluation of lexical innovations as evidence for a subgrouping argument must focus on the quality of the lexical innovations rather than the quantity. Exclusively shared lexical replacement innovations may form part of a strong argument for a traditional innovation-defined subgroup. However, neither novel concept lexical innovations nor homonymic lexical innovations, no matter their quantity, provide strong evidence for a traditional innovation-defined subgroup. Their main use in determining linguistic relations is rather the identification of lexical diffusion areas and early lexical innovation spread.

These theses are explored below in the context of the initial settlement of Borneo by PMP-speaking people and the development of their language into the current Bornean linguistic landscape, beginning with a review of the current model of linguistic relations in Borneo in section 2. Section 3 gives a detailed description of subgroup and linkage models, the way that lexical innovations interact with these models, and the different quality levels of lexical innovation. Section 4 discusses the GNB subgroup and argues that it is not a traditional subgroup but an innovation-defined zone of lexical diffusion. Section 5 applies a similar argument to the WIN subgroup and argues that it is an innovation-defined, nonexclusive lexical group restricted only to the languages of Borneo. Section 6 ties up loose ends within Bornean subgrouping by addressing the North Borneo subgroup. Here, a more traditional subgrouping argument is made against North Borneo due to issues with the proposed phonological evidence. A final statement and presentation of a new model for Bornean and MP higher-order linguistic relations are presented in section 7. That model has seven subgroups: Northeast Sabah, Southwest Sabah, North Sarawak, Central Sarawak, Kayanic, Malayic, and Land Dayak, plus the Barito–Basap (BB) linkage, descending directly from PMP with no intervening node. Rather than forming a subgroup, the languages form a lexically defined group that inherited certain Bornean-specific lexical items but did not descend from a single "Proto-Bornean" language.

2. SUBGROUPING AND THE HISTORY OF AN BORNEO. In this section, relevant background is provided on both previous proposals for subgrouping in Borneo as well as the likely settlement patterns of early AN settlers in Borneo. Recent subgrouping proposals tend to argue for subgroups which encompass large portions of the island, whereas the settlement of Borneo is

assumed to have taken place from the north during the rapid expansion of AN-speaking people into ISEA.

2.1. BORNEAN SUBGROUPING. As already mentioned, this study is concerned with subgroups previously proposed in Blust (2010) and later Smith (2017a,b, 2018) that include all AN languages of Borneo, except for the Tamanic languages, under the WIN subgroup, which also includes languages to the west and southwest of Borneo such as Javanese, Sasak, Lampungic, and Chamic but excludes Batak and Sumatran Barrier Island languages. WIN is considered a primary branch of PMP by its proponents. A tree representation is given below in figure 1. The non-Bornean WIN languages are listed as *External Languages*, although they do not form a single subgroup and their interrelatedness is indeterminate. Of the Bornean languages, two main divisions have been proposed: GNB and BB. BB is considered a linkage by Smith (2018) and includes the Basap varieties of north-eastern Borneo, Barito languages including Malagasy, and Sama–Bajaw. GNB was first proposed by Blust (2010) and contains all remaining indigenous languages of Borneo.

The inclusion of *external languages* in WIN is problematic, because WIN is primarily a lexically defined subgroup and most of the lexical evidence is absent from languages outside of Borneo. In addition, although Smith (2017a) provides a small list of phonological innovations, Adelaar (2023) demonstrates that many of the languages spoken to the west of Borneo lack these innovations and may ultimately not form a subgroup with the languages of Borneo.

Of the two major divisions in Borneo, this paper focuses mostly on the GNB group (although BB is also discussed). A tree representation following Smith (2017b) is given in figure 2. There are five proposed branches of GNB, including North Borneo, Kayanic (Kayan, Modang, Segai), Central Sarawak

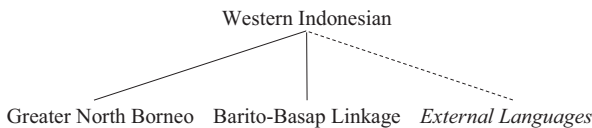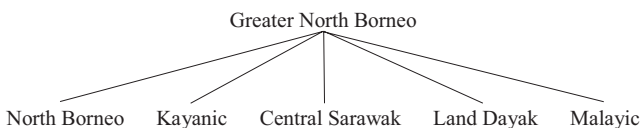**FIGURE 1. WIN AND FIRST-ORDER SUBGROUPS (BLUST 2010; SMITH 2017a,b).**

FIGURE 2. GNB INTERNAL SUBGROUPING (SMITH 2017b).

(Melanau, Kajang, Punan, Upper-Kapuas–Mahakam[3]), Land Dayak (Bidayuh, Benyadu-Bekati, Southern Land Dayak), and Malayic (Malay, Iban, Kendayan, and others).

Following Smith (2018), a linkage relationship is assumed for Barito, which includes Sama–Bajaw languages. The Basap group is only distantly related to Barito but is thought to be the Northern extreme of what was once a large dialect network that stretched from the Barito River to North Kalimantan. The linkage relationship is shown in figure 3 in a simplified wave diagram.

Finally, the internal divisions of the North Borneo group are shown below in figure 4, following the proposal from Blust (2010). There are three divisions: Southwest Sabah, Northeast Sabah, and North Sarawak. Individual language groups are listed in the tree. Southwest Sabah internal subgrouping mostly follows Lobel (2013, 2016).

With the exception of North Borneo, the major higher-order subgroups, including GNB, BB, and WIN, have only lexical evidence as support. The lower-level subgroups, those other than GNB, BB, and WIN, are supported
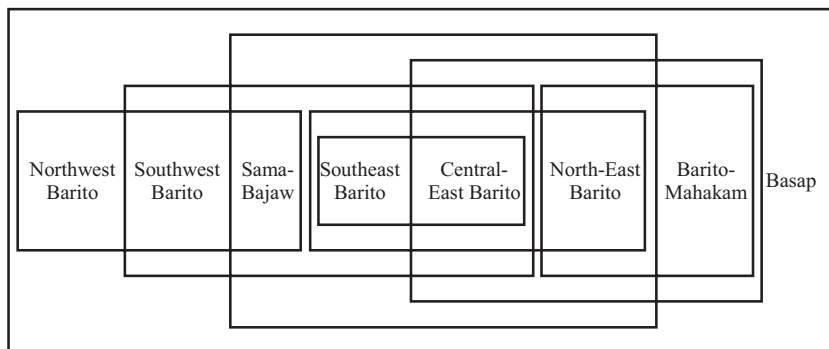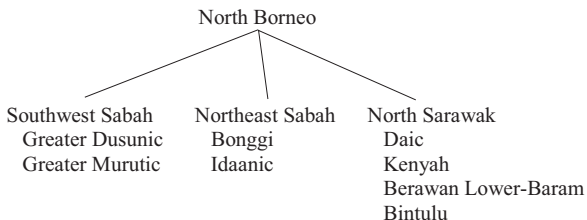
FIGURE 3. BB LINKAGE (SMITH 2018).

FIGURE 4. NORTH BORNEO INTERNAL SUBGROUPING (BLUST 2010).

---

3. I use the term Upper-Kapuas–Mahakam to replace the commonly used exonym "Müller-Schwaner" which appears in most earlier publications.

by a wider range of evidence, including regular phonological changes, shared irregular changes, and morphological evidence.

2.2. SOME PREVIOUS PROPOSALS. The arguments made in the remainder of this paper share certain similarities with previous proposals which sought to dismantle the internal structure of MP, namely those from Ross (2005) and Adelaar (2005). These proposals do not reference WIN but rather focus on the larger issue of the Western MP subgroup, a group that includes all MP languages not formally included in Central-Eastern MP. Ross does not go into much detail but speculates that Western MP might consist of "twenty-five to thirty groups." Adelaar proposes twenty-three subgroups within Western MP, which include the following subgroups relevant to the present discussion: Malayo-Sumbawan, North-Bornean, Kayanic, Land Dayak, East Barito, Barito–Mahakam, and West Barito. Adelaar does not propose any specific method of organizing these subgroups, but the implication is that Western MP is most likely comprised of multiple primary branches of MP. A further implication is that Bornean subgroups, like subgroups in other parts of the Western MP area, likely do not form an exclusive subgroup with one another.

Although these publications predate Blust and Smith's GNB and WIN subgroups, they remain relevant since they both argued for a much less structured MP family tree. Both authors' proposals are fundamentally incompatible with GNB and WIN as traditional subgroups. This paper is in the spirit of these earlier publications but differs in offering a specific proposal for classifying the languages of Borneo. This paper does not do away with the concept of a shared history between the languages of Borneo, but it does reclassify that history as one of contact and borrowing rather than one of inheritance.

2.3. THE SETTLEMENT OF BORNEO BY AN-SPEAKING PEOPLE. A major contribution of the *Greater North Borneo Hypothesis* (Blust 2010) was the acknowledgment that geographical boundaries affect linguistic classification. Blust hypothesizes that the first AN settlers to reach Borneo did so via the Philippines, likely moving to the northern tip of Borneo via Palawan. Once they reached the northern tip, the group split, with some following a western route around the part of the island facing the South China Sea and another following an eastern route around the opposite side of the island. The proposed linguistic division between GNB and BB, according to Blust, is the natural result of this geographical division. A full summary of the assumed migration routes, their linguistic properties, and their chronology is given below with reference to the paths plotted in figure 5.

A   The settling population that reached northern Borneo is assumed to have spoken a discrete language, Proto-WIN (PWIN) (Blust 2010). This group split into two or more groups that traveled along the western and eastern coasts of Borneo.

FIGURE 5. ASSUMED INITIAL SETTELMENT PATTERNS OF BORNEO
AND WESTERN INDONESIA BY AN-SPEAKING PEOPLE (BLUST 2010).
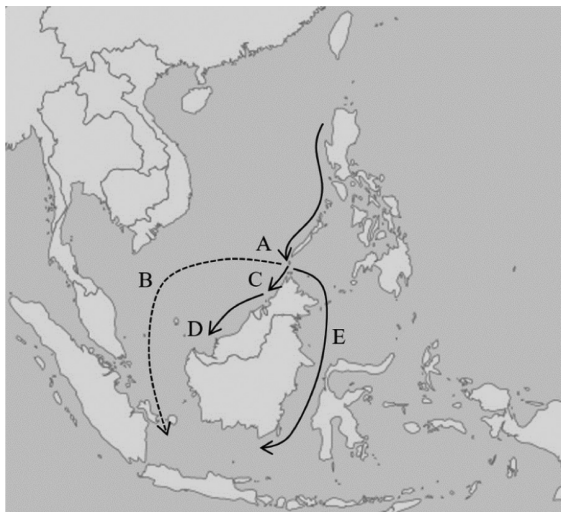
B   Some groups continued around Borneo and eventually settled the Greater
     Sunda Islands to the Southwest (not necessarily taking the path indicated
     by a dotted line).
C   The group that traveled along the western coast must have at some point
     developed as a single linguistic community. Blust (2010) suggests that
     this group eventually gave rise to a discrete Proto-GNB (PGNB)
     language.
D   After the development of PGNB, Blust posits that its speakers would
     have continued to spread out eventually coming to dominate the
     entire western coast of Borneo. Subgroups within GNB developed
     along the many large river systems and speakers of these languages
     moved inward and settled the deep interior of the island (Smith and
     Rama 2022).
E   On the other side of the island, a discrete Proto-BB language did not
     form, but rather a long dialect network formed stretching along the
     entire eastern coast of Borneo, eventually giving rise to the BB
     linkage (Smith 2018).

Archaeological and linguistic evidence suggest that the initial settlement of
Island Southeast Asia by the AN people was a product of rapid expansion,
which I will refer to as the rapid expansion hypothesis or the rapid expansion
model. Evidence supporting the rapid expansion hypothesis comes from early
signs of Neolithic presence in ISEA that closely follow the initial settlement of
the northern Philippines (see Bellwood 2007, for an overview). Evidence for
AN arrival all appears at around the same time, 3,500–4,000 years ago, again
suggesting rapid expansion and settlement. The linguistic evidence agrees with

the rapid expansion hypothesis; PMP split into multiple branches in a rake-like structure with little "nesting," indicative of a rapid expansion resulting in the dispersal of the community before there was time to accrue linguistic innovations (Smith 2017a).

The formal split between GNB and BB is explicitly attributed to the initial settlement indicated in figure 5 by Blust (2010:45). The original hypothesis therefore implies that the split occurred during the rapid expansion of AN-speaking people into ISEA. The homeland of PWIN is not known, nor is the homeland of PGNB, although the timeframe for such languages to have developed is exceptionally short. The likelihood that a discrete PGNB language developed somewhere on the northern coast of Borneo during the rapid expansion is low. A later expansion of PGNB-speaking people from their currently unknown homeland after the rapid expansion might explain how the language was able to develop, but there is no clear center of such a dispersal. Blust argued that the lexical innovations provide evidence for this expansion, but as this paper argues, the lexical evidence is not as strong as originally believed.

3. SUBGROUPING MODELS AND LEXICAL EVIDENCE.

3.1. SUBGROUPING MODLES. One of the focuses of this paper is to argue for a more diverse representation of linguistic relations in the AN subgrouping. Ultimately, the AN language family, even at higher-order divisions once thought to consist of traditional subgroups, must be thought of as consisting of many different types of linguistic relations, including traditional innovation-defined subgroups, innovation-defined linkages, and various types of contact areas. A contact area may consist of languages with clearly different histories (maybe even unrelated) that have exchanged structural properties through clearly identifiable contact as well as more closely related languages that have undergone an extended history of mutual development but which nevertheless do not descend from an exclusive common ancestor. This latter type of contact area is argued below to accurately describe the linguistic relations of subgroups in Borneo.

3.1.1. Innovation-defined subgroups. Innovation-defined subgroups are the "classical" subgroup of comparative linguistics. It is not necessary to go into much detail on what makes an innovation-defined subgroup since they are the quintessential linguistic relation. Innovation-defined subgroups require *exclusively shared innovations* and are assumed to have a *single exclusive proto-language* from which modern languages descend. Those innovations are most commonly phonological or morphological innovations. Lexical innovations may also indicate an innovation-defined subgroup, but lexical innovations are less reliable and require careful consideration (more below in section 3.2).

3.1.2. Innovation-defined linkages. Innovation-defined linkages are similar to subgroups in that they require exclusively shared innovations. They differ

TABLE 1. STEP-LADDER LIKE DISTRIBUTION OF SOUND CHANGES
IN A LINKAGE (SMITH 2018).

	NwB	SwB	Yakan	SᴇB	C-EB	NᴇB	Tunjung
*R > h	1	1	1	0	0	0	1
*ə > e	0	1	1	1	1	0	0
*z > *d > (r)	0	0	1	1	1	1	0
*-R > j	0	0	0	1	1	1	0
*-b > w	0	0	0	0	1	1	1
*-d > r	0	1	0	0	0	1	1
*-l > r	0	0	0	0	0	1	1
*d- > r	0	0	0	1	1	0	1
*b- > w	0	0	0	1	1	0	0

in the distribution of those innovations throughout the linkage. Pioneering work on linkages is found in Ross (1988), where the foundational principles of linkage relationships are laid out with a focus on Proto-Oceanic and subsequent developments within Oceanic. A prominent feature of linkages is that innovations are found spread throughout the subgroup, but each innovation is not necessarily found in every member language and cannot be said to have been inherited from a common ancestor (François 2014; Heggarty, Maguire, and April 2010). The innovations in a linkage are distributed in such a way that no nonarbitrary line may be drawn where one language is separated from another. This situation arises when innovations spread "horizontally" from one community to another, influenced by geographical closeness, social interactions, and cross-community relations (Milroy and Milroy 1985; Ross 1997). Some linkages have an additional attribute where the innovations form a "step-ladder" distribution when plotted on a table. Smith (2018), for example, provides an example reprinted in table 1, as a demonstration of the step-ladder-like distribution of innovations in the Greater Barito Linkage. Note that the order of languages is not arbitrary, and the step-ladder distribution follows the geographical position of languages.

Although the innovations of a linkage are not found in all member languages, the languages are still considered to be more closely related to one another than to any nonmember language. They may descend from a dialect network in much the same way that languages of a subgroup descend from a single ancestor language. In fact, a single language may first diversify into a complex dialect network and, over time, develop from that network further into an innovation-defined linkage.

3.2. LEXICAL INNOVATIONS AND SUBGROUPING. The WIN and GNB subgroups, as they are currently proposed, do not have any exclusively shared phonological or morphological innovations. Instead, Blust (2010) and later Smith (2017a,b) argue for these subgroups with only lexical evidence. These proposals assumed that the presence of a lexical innovation in only some

of the member languages of GNB supported the GNB hypothesis because (i) contact was not considered a reasonable explanation due to regular sound correspondences and (ii) the presence of an innovation in only two members still warranted its reconstruction to the proto-language, and once an innovation is reconstructed to the proto-language it then becomes part of the total lexical package that is used to justify the existence of that proto-language. This approach is clearly flawed, since it relies on assertions about the history of individual lexical items that are not supported by the evidence itself. This section therefore proposes a more cautious approach to the use of lexical innovations as a means to study linguistic relations and histories.

3.2.1. Distribution of lexical innovations. Linguists must be vigilant to avoid the mistake of giving too much subgrouping power to sparsely attested lexical innovations. The distribution of lexical innovations, for example, was left mostly undiscussed in Blust (2010) and Smith (2017a,b), even though distribution may contain information critical to accurate analysis. Lexical items are the most readily borrowed part of any language, and lexical diffusion between languages or dialects in contact occurs regularly. If the lexical innovations for a proposed subgroup are distributed in a way that resembles a linkage relationship, then horizontal spread between languages or dialects in contact is the most likely explanation for the observed distribution. Lexical items that diffuse in a network are also quite difficult to place temporally if no phonological evidence exists to help place them, since lexical borrowing can easily occur between dialects, closely related languages, or distantly related languages without leaving a phonological clue as to their status. For this reason, lexical innovations that are distributed in a linkage-like distribution cannot be assumed to have been inherited from a proto-language. All that can be said about such innovations is that they spread from one community to another during a time before major subgroup-defining phonological innovations had occurred, but we cannot even state if that spread occurred between dialects or languages.

Lexical innovations may also be distributed more randomly within a proposed subgroup and may be robustly attested. The analysis of such a distribution depends on quality, but if phonologically regular lexical innovations are found with no clear linkage-like distribution, then it becomes more likely that they were inherited from some common ancestor. However, without a complementary list of exclusive phonological or morphological innovations and no high-quality replacement innovations, it is still not possible to separate the ancestor of this lexically defined group from the next-highest proto-language. In the context of Borneo, it is not possible to show that the common ancestor of Bornean languages was anything other than PMP itself, even with the existence of a Borneo-only lexicon. A thought experiment dealing with the initial settlement of Borneo by PMP-speaking people may better explain this history.

Imagine speakers of PMP move into the western portion of ISEA from the Philippines. Some groups settle in Borneo, but these groups do not speak a special "Proto-Bornean" language. Rather, they speak PMP. During the first few

years after initial settlement, these PMP speakers would inevitably develop new terminology relating to their new environment, including many of the lexical innovations that we find in Bornean languages today. Eventually, these people would continue to settle the entire coast of Borneo, as is expected of PMP-speaking people taking part in the rapid expansion. They take along with them their version of PMP, now infused with additional novel concept innovations specific to Borneo but still PMP. Fast forward several thousand years, and we have today's situation, an island inhabited by the descendants of these first AN settlers. How should we model the linguistic relations of these languages? They all descend from speakers of PMP and did not begin to diversify until long after the initial settlement of Borneo. Such conditions would suggest that each of the major subgroups within Borneo represents a primary branch of PMP with no intervening nodes. However, after initial settlement while still speaking PMP, certain Bornean-specific terms were introduced to the Bornean variety of PMP, resulting in several Borneo-specific vocabulary items. The ancestor of modern-day Bornean languages is still PMP, but lexically speaking, it is not the same PMP that was spoken by people in Sumatra, the Philippines, or Java. Focusing on these facts, one may be inclined to posit a Bornean subgroup which consists only of the Bornean languages that held on to these lexical innovations. Neither solution is perfect, since the hypothetical "Proto-Bornean" language is just PMP with a few extra lexemes, but claiming that each subgroup is simply a primary branch of PMP ignores the shared history of Bornean languages as observed in an exclusive Bornean lexicon. In this case, it is necessary to distinguish between exclusive and nonexclusive common ancestors. An exclusive common ancestor may receive a node on a family tree, whereas a nonexclusive common ancestor (like the language of the PMP-speaking first settlers of Borneo) does not.

3.2.2. Quality of lexical evidence. Another important aspect of lexical evidence is quality. As Smith (2017a) notes in his argument against the Philippine subgroup, lexical innovations, like phonological innovations, must also be evaluated as "stronger" or "weaker" depending on their quality. Low-quality lexical innovations, even those with regular sound correspondences, cannot be assumed to have originated in a common ancestor.

The most powerful type of lexical evidence is exclusively shared lexical replacement innovations. These types of innovations occur when a well-supported reconstruction is replaced by an innovated word in a daughter language. An example of such an innovation is Proto-Kayanic *ŋad 'gills', which replaced PMP *hasaŋ (Smith 2017b). The innovation *ŋad is exclusive to Kayanic since it appears nowhere else in AN. It is a replacement innovation because the reflexes of PMP *hasaŋ are absent from Kayanic. The best interpretation of these facts is that there was a lexical replacement event which occurred in a common ancestor of Kayanic languages, Proto-Kayanic. That lexical replacement was then inherited in the daughter languages in much the same way that a sound change is inherited.

Lexical replacement innovations may be split into two types: (i) new form replacement and (ii) semantic shift replacement. The example shown above of Proto-Kayanic *ŋad is an example of a new form replacement, since *ŋad has no known source. An example of semantic shift replacement is the proposed *tuzuq 'to point' > *tuzuq 'seven' shift, which is a defining innovation in GNB. The PMP word *pitu 'seven' is completely absent in GNB, assumed to have been replaced by *tuzuq. The only difference between a new form replacement innovation and a semantic shift replacement innovation is that the semantic shift replacement innovation has a known source; in this case, the source was PMP *tuzuq 'to point'. Some semantic shifts may involve universal semantic changes, and therefore may be less powerful for subgrouping than new form replacements.

Lexical replacement innovations have one weakness which diminishes their usefulness when compared to regular phonological changes: a lexical innovation may be completely obscured by further replacement, whereas a phonological change is usually detectable even if further phonological changes obscure the original sound change. An example is the proposed innovation *alud 'canoe', which was originally proposed as a GNB lexical replacement innovation. Although *alud is said to have replaced PMP *qabaŋ 'canoe', Kayanic languages reflect a different innovation, *haruk 'canoe', not *alud. Because *alud is a lexical replacement innovation, one assumes that *haruk replaced *alud, not earlier *qabaŋ 'canoe', but since lexical replacement innovations, unlike phonological changes, tend to leave no trace, there is no direct evidence that Kayanic languages ever had a reflex of *alud. One may consider the weight of other evidence which suggests that Kayanic languages may subgroup with the other languages of Borneo and therefore most likely underwent the *qabaŋ > *alud innovation, but without complete attestation of a lexical replacement innovation in all subgroups, its weight as subgrouping evidence is diminished.

A weaker type of lexical innovation is novel concept innovations, which are abundant in Borneo. These involve new terms used to describe concepts or objects not previously known to the language community. In Borneo, these are words that describe new flora and fauna that were encountered when AN speakers crossed the Wallace line between Palawan and the rest of the Philippines (as proposed by Thomas Huxley [Camerini 1993]) and reentered the Mainland Southeast Asian zone. The biggest issue with these types of innovation is that words for new concepts are often borrowed, and in the case of the early naming of plant and animal life on Borneo, it is possible that the words may have spread by means other than inheritance.

In defense of the usefulness of novel concept innovations in Borneo, one may point to the fact that there is no known donor language from where these terms may have been borrowed. It is possible that AN settlers arrived on a populated island. In fact, it is difficult to imagine that Borneo, which has archaeological evidence of human inhabitation dating back tens of thousands of years, was empty when ANs arrived (Bellwood 1988, 1989; Majid 1982). One may

therefore postulate that the names of local plants and animals were borrowed from these already-existing populations. Although it is true that words for novel concepts, plants, and animals may be easily borrowed, it is not clear if this occurred in Borneo, where there are numerous plant and animal names that refer to new species that are both well attested in all major subgroups and which demonstrate complete phonological regularity.[4] This is true even though major subgroups in Borneo sometimes have extreme differences in historical phonology. If borrowing was the main source of novel concept innovations which are phonologically regular with one another, the borrowing most likely came from a single pre-AN group to the very first Austronesians, who arrived who then passed it on to their descendants. If this is true, then the borrowing probably occurred once, with words entering the lexicon of a single language. If, however, borrowing occurred all over the island from more than one pre-AN language, one would expect that different settling populations borrowed terms for local plants and animals from different local pre-AN groups. Under that scenario, widespread cognates with regular sound correspondences in AN languages are less likely.[5]

The weakest type of lexical innovation is homonymic innovations, which occur when a new form is innovated alongside an already existing lexeme. The fact that the original word may be attested throughout the subgroup alongside the innovated lexeme means that it is much more difficult to determine if the change occurred in a proto-language or if it spread through contact. Such innovations have little value in a subgrouping argument.

An example of homonymic innovation from Blust (2010) is the proposed innovation PMP *(ba)labaw 'rat, mouse' > PGNB *(ba)labaw, PGNB *tikus 'rat, mouse'. According to Blust (2010), the innovation *tikus does not replace *(ba)labaw but rather is innovated alongside it. Thus, we may reconstruct two PGNB words for 'rat, mouse'. One may expect that these words refer to slightly

---

4. However, some Bornean vocabulary does show irregularity that may be explained by borrowing from multiple different non-AN languages. Smith (2017b) list 26 "near cognate" sets; words with similar form and meaning that can be reconstructed to numerous proto-languages but which exhibit irregularities that prevent reconstruction to a higher node. Several possible scenarios may explain these near cognates, including borrowing between AN languages, borrowing from non-AN languages, or more recent borrowing due to trade and cross-sea contact between Mainland Southeast Asia (MSEA) and ISEA. Note, however, that the list of near cognates is not exclusively novel concept innovations. In fact, only two of the 26 near cognates could be considered novel concept innovations; those meaning 'longhouse' and 'hornbill', but the remaining 24 occur mostly in basic vocabulary. So, they do not appear to have been early borrowings used to describe the natural environment of Borneo. They are better described as a later phenomenon that occurred sometime after the initial coastal settlement of Borneo by PMP-speakers.

5. The borrowed vocabulary in this scenario may be regular with one another from the perspective of the now extinct non-AN donor languages, but that "regularity" would almost certainly not also be regular from the AN perspective. For example, if non-AN languages A, B, and C have regular d : r : l correspondences, it would be a miraculous coincidence if those words were borrowed into AN languages that also had d : r : l correspondences. So, regularity within AN suggests that borrowed lexical items would have entered into the vocabulary of the PMP-speaking settlers from a single donor. Whether shared vocabulary with phonological regularity entered the Bornean lexicon through borrowing or innovation does not change their status as innovations unique to the settling community.

different types of mice, but there is no comparative evidence that allows such a difference to be reconstructed.

3.3. A NEW APPROACH TO BORNEAN LINGUISTIC RELATIONS.

An acknowledgment that lexical evidence can be weaker or stronger depending on its type and that the distribution of those innovations should impact our analysis has wide-ranging implications for subgrouping in ISEA. That, combined with the use of different subgrouping models dependent on the quality of evidence, allows for a new approach to higher-order subgrouping in Borneo. An overview of the application of this approach is provided here.

Subgroups based solely on lexical innovations should ideally be comprised of exclusively shared lexical replacement innovations. Without such innovations, we cannot be certain that the observed similarities are the result of inheritance. Large subgroups such as GNB and WIN should be judged not on the number of lexical innovations but on the quality of those innovations. In evaluating lexical evidence, one must keep in mind the following questions: Are there numerous lexical replacement innovations that support the subgroup? Are those lexical replacement innovations found throughout all proposed subgroups within the larger subgroup? Do they form a cohesive group of subgroups rather than contradictory subgroups with overlapping and fuzzy boundaries? If the answer to these questions is yes, then this suggests a well-supported subgroup with lexical innovations as the main form of evidence. On the other hand, if the evidence is primarily homonymic, is concerned with novel concept innovations, is not further supported by robust lexical replacement innovations, is not found in all member subgroups, and creates contradictory, overlapping, and unclear subgrouping boundaries, then one should consider that the evidence supports a zone of lexical diffusion rather than a subgroup or linkage.

Comparative linguists working with lexical evidence must therefore be careful about the types of proposals they make with the evidence that they have. As discussed in section 2, AN-speaking people moved quickly during their initial settlement of ISEA. The rapid expansion model is most likely to have resulted in widespread mutual intelligibility among the initial settlers over great distances, with intelligibility weakening over time, creating a complex network of languages and dialects. AN settlement did not occur without continued contact between groups. Loanwords are found throughout ISEA, likely spread back and forth between different AN groups in continuous contact over great distances. Easily identifiable examples are *kabaw/kerbaw* 'water buffalo', found as far north as Paiwan and as far east as Fiji, or *pidak/pirak* 'silver', found throughout ISEA, both from a Mon-Khmer Source (Thurgood 1999).

Would-be first-order subgroups therefore developed within the well-connected system of contact and trade between both AN and non-AN people during the period immediately after the rapid expansion. Given this history of early AN settlement in ISEA, the idea that lexical innovations were innovated in discrete, isolated proto-languages and that they did not spread through the

widespread network of contact and trade is implausible. Therefore, only the strongest type of lexical innovation, exclusively shared lexical replacement innovations, is more likely to be the product of inheritance than borrowing in this type of linguistic environment. Even so, exclusively shared lexical replacement innovations may provide contradictory results and should only be considered after careful scrutiny. Other types of lexical innovations are simply not strong enough to rule out alternatives to inheritance.

The remainder of this paper attempts to apply this approach to higher-order subgroups in Borneo, GNB, and WIN, with additional implications for higher-order subgrouping in MP.

4. GNB. Blust (2010) originally proposed GNB as a traditional subgroup defined by a set of shared lexical innovations. Smith (2017a) adds multiple additional innovations, and the current list of GNB lexical innovations stands at thirty. Those innovated lexemes are presented in table 2 with their hypothesized GNB reconstructions.

4.1. REANALYSIS OF GNB LEXICAL INNOVATIONS. In the years since Blust (2010) and Smith (2017a), the strength of this list has been somewhat diminished due to closer analysis and new data. As a result, some of the words in this list should be removed and the overall number of lexical innovations reduced. In the remainder of this section, I review two types of evidence from the list, evidence that should be rejected outright and evidence that is in doubt pending further investigation.

4.1.1. Words to be removed. In this section, I argue that certain words are almost certainly not valid as GNB innovations. The words in this list meet certain criteria for removal: (i) irregularities in sound correspondences that indicate borrowing, and (ii) presence in BB (or other subgroups), which means that the innovations, if they are considered inheritances from a common ancestor,

TABLE 2. GNB LEXICAL INNOVATIONS.

*kəraʔ	'long tailed macaque'[6]	*ñaRu	'eagle'	*cəRaʔuŋ	'sun hat'
*kuini	'mango species'	*saʔay	'frog'	*lipəs	'cockroach'
*labi	'soft shelled turtle'	*tukul	'hammer'	*saʔup	'parang handle'
*təməduR	'rhinoceros'	*kitan	'binturong'	*gaduŋ	'grue'
*kadis	'grasshopper'	*puʔan	'squirrel'	*kəlit	'small bat'
*lamin	'room of a house'	*kuju	'heron'	*lunək	'soft; mushy'
*kuyad	'long tailed macaque'	*alud	'canoe'	*tikus	'rat'
*laŋkaw	'temporary shelter'	*ajən	'fish'	*tujuʔ	'seven'
*sakay	'stranger'	*ambay	'sweetheart'	*damək	'blowpipe dart'
*guaŋ	'heart; desire'	*sulap	'temporary shelter'	*təgap	'firm; sturdy'

---

6. Two words are reconstructed for 'long tailed macaque', *kəraʔ and *kuyad.

should be reconstructed to a level above the putative GNB node. The words in this list are probably the least controversial removals.

4.1.1.1. *cəRaʔuŋ 'sun hat'. This reconstruction is proposed by Smith (2017a, b) as a supplement to Blust's evidence. It is found throughout Borneo in numerous subgroups but is absent as a native word in Barito. There is nothing wrong with the Bornean evidence, but Amis *saoŋ* 'sun hat' forces one to reconsider the status of this word as an innovation. It is possible that the -əR- element in Borneo reflects an early infix of similar shape, Proto-AN (PAN) *-aR-. A putative PAN *caquŋ may yield Amis *saoŋ*, so the status of this word as a GNB innovation is in doubt. It is also worth noting that Basap *sərauŋ* may also be a native word, although the possibility of borrowing cannot be ruled out.

4.1.1.2. *kuini 'mango species'. This word is found in Central Sarawak, Kayanic, Land Dayak, Malayic, and North Sarawak, according to Blust (2010). It therefore appears to have a wide distribution in GNB. However, most of the witnesses that Blust provides are irregular, suggesting borrowing from Malay. Smith (2019) notes that all Bidayuh languages close final vowels with -*h*, and a lack of closure indicates late-stage borrowing, in this case probably from Malay. Bidayuh *kuini* 'mango with a strong smell and fine flesh' is therefore not native. Kayan *kuini* fails to add a glottal stop to word-final position, a sound change that is ubiquitous with Kayanic languages (Blust 2002a). Kiput *kini* does not reflect word-final vowel breaking (expected *-i > -ay, Blust 2002b), and Long Terawan Berawan *kini* fails to add a word-final *h* (Burkhardt 2014). This word must therefore be thrown out as a borrowing, most likely from Malay.

4.1.1.3. *lipəs 'cockroach'. PGNB *lipəs is a proposed irregular reanalysis of PMP *ipəs 'cockroach' with the initial *l- originating from the *qali- prefix (Blust 2001). Although this fusion was originally said to have only taken place in GNB languages, reflexes of *lipəs are well attested in Basap and Barito: Ngaju *lipes* 'small cockroach', Kapuas *lipes* 'cockroach', Maanyan *lipes* 'cockroach', Dusun Witu *lipes* 'cockroach', Tawoyan *lipəs* 'cockroach', Benuaq *lipas* 'cockroach', Basap *lepəs* 'cockroach'. These words all show regular sound correspondences, so there is no reason to assume that they are cases of borrowing (If they were borrowed from Malay, for example, *ə would be irregularly reflected as *a*). The *lipəs* innovation therefore does not provide evidence for GNB since it is found in all Bornean subgroups.

4.1.1.4. *kadis 'grasshopper'. Apparent reflexes of *kadis are found in Land Dayak, Kajang, Punan Bah, Kenyah, and Berawan Lower-Baram. The Land Dayak, Kajang, and Punan reflexes of this word are all irregular with one another and need reevaluation. In Kajang, if there was a final *–s it would have blocked vowel breaking of *i, but Kajang languages do not appear to have ever had a final *-s: Sekapan *təkarəy*, Kejaman *kəlarəy*. Compare those reflexes to Sekapan and Kejaman *bəti* 'leg', from *bətis. In Punan Bah, a word-final *-s after a high vowel

does not result in vowel lowering; the high vowel is typically retained unchanged. Again, Punan Bah *beti* from *bətis provides an example of this. However, 'grasshopper' does show lowering: *təkare*, which suggests borrowing. Land Dayak reflexes of this word have a full vowel *a in penultimate position, which Smith (2019) notes is a sign either of early borrowing or coalescence of an older penultimate and antepenultimate syllable under specific conditions, and native full-vowel penults are quite rare in Land Dayak.

It is most likely that this word originates in Lowland Kenyah *kare* and was brought into Kajang through the Lowland Kenyah-Kajang zone of contact in the area around the Usun Apau plateau (Smith 2017a). Punan Bah, in turn, has a strong contact situation with Kajang. Although it is still not clear how an apparent reflex might have made its way into Land Dayak, because of the full vowel, it is not even clear that the Land Dayak word is related. All that remains are North Sarawak examples which still fail to provide evidence for a larger GNB subgroup.

4.1.1.5. *ambay 'sweetheart'. Blust (2010) lists this word as appearing in Land Dayak, Malayic, and Southwest Sabah. However, the Land Dayak word, Bidayuh *ambay* 'secondary wife, concubine; lover, sweetheart' has an irregular retention of the final diphthong. Smith (2019) points out that the inherited word final diphthongs *-ay and *-aw became -*i* and -*u* in Bidayuh. This word is therefore borrowed from Malay. The only Southwest Sabah witness, Tagol *ambay*, cannot therefore stand alone as the only non-Malayic evidence for a GNB innovation.

4.1.1.6. *alud 'canoe'. Blust (2010) and Smith (2017a, b) consider *alud a PGNB innovation, probably innovated after new canoe types were used in the transition from sea-travel to river-travel. The word is well attested throughout Borneo but also appears in BB: Kadorih *aɾut* 'canoe', Benuaq *alur* 'canoe', Tunjung *alur* 'canoe', Basap (Lebo) *alun* 'canoe'. These words are not likely borrowings. The reflexes, including Kadorih *ɾ* from *-l- and *t* from *-d, Benuaq and Tunjung *r* from *-d, and Lebo Basap *n* from -d, are all regular. Both Kayan and Upper-Kapuas–Mahakam have exerted influence on BB, but both subgroups have innovations other than *alud, and could not have been the source of apparent reflexes in BB. The regularity of reflexes and the unlikelihood of a borrowing explanation force the reconstruction of *alud beyond PGNB.

4.1.1.7. *kəraʔ 'long tailed macaque'. As Blust (2010) points out, this word is similar in form to an onomatopoetic word referring to the chatter or cries of monkeys, which he lists as *kəraq in Blust (2010) and as *akərahaq on the *ACD*. Evidence for this word comes from Tagalog *aklahaʔ* 'cry of monkeys', Maranao *kəraʔ* 'sound made by monkeys', Ngadha *kəra* 'shrill cry of monkeys'. Its presence outside of Borneo weakens its status as an innovation. In addition, Malayic and Sabahan terms disagree about the presence or absence of a final glottal (which would become *h* in Malay, assuming that this is

from *-q). There are other inconsistencies as well. Words of similar shape and meaning are found throughout Borneo, some likely the result of parallel innovation as onomatopoetic words. A summary of words with similar shapes referring to various monkeys in Borneo is given below:

Long Tailed Macaque
Tawoyan *kodeʔ* 'long tailed macaque', Benuaq *kodeʔ* 'long tailed macaque', Tunjung *kodeʔ* 'long tailed macaque', Paser *kode* 'long talied macaque'.

Proboscis Monkey
Kadorih *bakaraʔ* 'proboscis monkey', Ngaju *bakaraʔ* 'proboscis monkey', Tawoyan *bəŋkaraʔ* 'proboscis monkey', Basap *bəkaraʔ*, Busang *bəkroʔ*, Kelai *bəklaʔ*, Punan Tuvu *bəkaro*, Buket *bəkəroʔ*, Seputan *mokəroʔ*, S. Bisaya *bakaru*, Bonggi *kəraʔ*.

Red Leaf Monkey
Kapuas *bəkaraʔ*, Benuaq *bəkaraʔ*, Tunjung *bəkəraʔ*, Burusu *bəkaro*

Because similar onomatopoetic words are found both outside of Borneo and in BB, and also because these words are applied to all types of monkeys and often have irregular or unexpected sound correspondences, even within the proposed *kəraq innovations, this word is probably best explained as a result of onomatopoetic similarity, in some cases possibly inherited from PMP *akərahaq but in others perhaps arising through parallel innovation.

4.1.1.8. *lunək 'soft; mushy'. Words of similar meaning are found in Barito languages. Although Blust (2010) dismisses these words as likely unrelated, it is not clear at all that they should be excluded. This includes Ngaju Dayak *lunək* 'edible flesh of a fruit that clings to the seed (of jackfruit, rambutan, etc.)', Malagasy *lunaka* 'rich, good. Applied to soil'. Ngaju Dayak 'edible flesh of a fruit' and Iban 'pulpy, fleshy (of fruits)' could easily be related, and the Ngaju Dayak reflex of schwa is diagnostic, indicating that it was not borrowed from Malay. The semantic differences are minimal, and this word is therefore not likely a GNB innovation.

4.1.2. Weak innovations. The following list contains proposed innovations that are considered weak. The main distinction between words in this list and words in the previous list is that these words remain somewhat ambiguous as to their status. Often, this is because the counterevidence is restricted to fewer languages or because straightforward phonological evidence for exclusion is not available.

4.1.2.1. *tukul 'hammer'. There are at least two reflexes of *tukul in Barito languages that are regular: Maanyan and Dusun Witu *tukun*, both of which show the regular *-l > -n sound change. However, there are also clear loans in other Barito languages, such as Kadorih *tukul* (expected *tukun*) and

Tawoyan *tukul* (expected *tukur*). The Maanyan and Dusun Witu words may therefore be borrowings that have been incorporated into their respective phonologies. However, other Malay-sourced words can be found in Maanyan and Dusun Witu that did not alter the final *l*, such as Maanyan *tumpul* and Dusun Witu *tompol*, both from Malay *tumpul* 'blunt; dull'. If *tukun* is indeed native than *tukul* should be rejected as GNB evidence.

4.1.2.2. *puʔan 'squirrel'. Smith (2017b) proposed *puʔan as additional GNB evidence. This reconstruction should probably be *puʔər, however, because of Daic evidence: Long Bawang *puər* and Kelabit *puur*. PMP *-R becomes *-r* in Daic, but since Daic *-r* from PMP *-R corresponds with Ø in Kenyah and Berawan–Lower Baram, both of which have *-n* in this word, 'squirrel' cannot be reconstructed as *puʔəR. To explain these reflexes, it is hypothesized that *-ər became *-ar in these languages, and *-r later merged with *–n. It is, admittingly, difficult to find evidence for this hypothesis, but Long Semadoh *miər* and Lun Bawang *miər*, both 'to see', have a similar correspondence with Miri *tiʔan*, and Kenyah *taʔan* (note that Daic languages often drop and reanalyze initial consonant, with *m-* as an active voice prefix attaching to a reanalyzed root *iər*).

This makes the comparison with Kajang *puʔan* a bit more troublesome. There is already an established borrowing relationship between Lowland Kenyah and Kajang so borrowing, rather than the parallel innovation of *-ər > *-an, could explain the matching reflexes between Kajang and Kenyah.

4.1.2.3. *sulap 'temporary shelter'. Blust admits that Iban *sulap* may be from a Southwest Sabah source. The Land Dayak word is also problematic. Although Blust gives Bidayuh *surap* 'a top cover, as in roofing' Smith (2017b) records Bekati' *sirap* 'wooden roofing', suggesting Proto-Land Ddayak *sirap*, not *surap. Interestingly, a similar word is also found in Kiput *sila:p* 'field hut'. The vowel here indicates that there may be some competition among reconstructions, *silap/*sulap, or perhaps that this word is a near-cognate, rather than a word inherited from a common ancestor.

4.1.2.4. *ajən 'fish'. The apparent native reflexes of this word are confined to Central Sarawak and the Murik-Merap branch of Kayanic. It does not replace PMP *hikan 'fish', although reflexes of *hikan are surprisingly sporadic in Borneo and there are numerous innovations for 'fish' found all over the island.

The first issue with this reconstruction is the form itself. Smith (2017) proposed *ajən, but the initial vowel is not well supported. Because of reflexes like Kanowit *jən*, Sekapan *jən*, and Aoheng *ocen*, a better supported reconstruction is *əjən. The second issue is with its distribution. This innovation is rather well attested in Central Sarawak, although Melanau language retain *hikan. In Kayanic, *əjən only appears in Murik-Merap. Kadorih *ocin* is a borrowing from Central Sarawak (Upper-Kapuas–Mahakam). A similar word also appears in Basap, *ujən/ajən*, but the reflexes here are irregular since schwa did not change in penultimate position in Basap.

Because of the limited distribution, a possible explanation for this word is that it is an innovation that spread through contact specifically in central Borneo, away from the coast, which explains why it is absent in Melanau. There is not direct phonological evidence to throw it out from Central Sarawak and Murik-Merap.

4.1.3. Notes on the remaining lexical evidence. The remaining innovations do appear to be genuine. They have regular sound correspondences and are restricted to languages that Blust includes in GNB. In some cases, it is necessary to elaborate a bit on individual reconstructions in this class. Those elaborations are given in this section.

4.1.3.1. *tuzuq 'seven'. Blust (2010) describes this as a semantic shift replacement innovation. PMP *tuzuq 'to point' shifted to *tuzuq 'seven' in PGNB and replaced PMP *pitu. He further claimed that this shift took place in all GNB languages. However, two subgroups in Borneo, Land Dayak and Upper-Kapuas–Mahakam, retain *tuzuq 'to point' without the semantic shift. Words for 'seven' in these languages are formally similar yet importantly distinct from *tuzuq: Proto-Land Dayak *ijuʔ 'seven' and Proto-Upper-Kapuas–Mahakam *ticuʔ 'seven'. Although one may attempt to explain these as irregular reflexes of *tuzuq, since both Land Dayak and Upper-Kapuas–Mahakam have regular reflexes of *tuzuq 'to point', such attempts would be unfounded (Bakati' *nujuʔ* 'to point' vs *ijuʔ* 'seven' and Seputan *nucu* 'to point' vs. *ticu* 'seven'). It is possible that similar words to *tuzuq entered the vocabulary of Land Dayak and Upper-Kapuas–Mahakam from outside sources, just as they did in some Barito languages, but no matter the source, these languages did not shift their native reflexes of *tuzuq from 'to point' to 'seven'. To be clear, however, even though these languages do not reflect the semantic shift from 'to point' to 'seven' in reflexes of *tuzuq, they also do not reflect PMP *pitu 'seven'. Blust's observation that GNB languages replaced *pitu remains valid, but the replacement of *pitu does not appear to have been universally caused by a shift in the meaning of *tuzuq.

4.1.3.2. *kitan 'binturong'. This word also appears in Basap with an unknown infix, *kawitan*. It is possible that this word reflects *kawit-an, with the root *kawit 'hook' to describe the binturong's long curved tail, but not enough evidence is available to know for sure, so this is kept as a potential GNB innovation.

4.1.3.3. *gaduŋ 'grue' (green/blue). In Malay, this word is only found in the Upper Kapuas Iban variety and could easily have come from the nearby Upper-Kapuas–Mahakam or Punan languages. Blust and Trussel (2020) note that similar words appear in the Philippines through contact with Sama–Bajaw. This may be evidence that the word is also found in BB, but at this point not enough is known to rule out borrowing into Sama–Bajaw from a GNB source.

4.1.3.4. *lamin 'room of a house'. Kayan languages in Smith (2017b) have *amin*, with no initial consonant, although Blust (2010) cites Kayan *hamin* without giving an indication of which dialect it is from or the source material. An

initial consonant that yields Kayan Ø but likely went through an earlier stage where it was *h- is PMP *R. Kayan evidence therefore points to *Ramin, not *lamin.

Highland Kenyah languages also have *amin*, while Lowland Kenyah has *lamin*. The same issue is again found in Berawan–Lower Baram; Berawan reflects *lamin, but Lower-Baram reflects *amin. Both Highland Kenyah and Lower-Baram witnesses could reflect *Ramin. This suggests a doublet, *lamin/Ramin, or possibly a near-cognate.

4.1.3.5. *tikus 'rat'. Blust (2010) lists Lara Bekati *tikus* as evidence for this proposed GNB innovation in Land Dayak. Although there are no diagnostic phonemes with which to test the status of this word, larger sources on Land Dayak suggests that this is not native. Both Rensch et al. (2012), and Smith (2017b) list reflexes of *belabaw in dozens of Land Dayak languages including Lara Bekati, but no reflexes of *tikus. It is not clear where Lara *tikus* was sourced from in Blust (2010), but it is probably best considered nonnative vocabulary, again from a Malay source.

The *(ba)labaw/*tikus innovation is said to be a homonymic innovation, with *(ba)labaw being retained from PMP and *tikus newly innovated. It is now clear that only three subgroups reflect *tikus, Malayic, Southwest Sabah, and Northeast Sabah. These subgroups have totally replaced *(ba)labaw with *tikus. Other subgroups in Borneo show no sign of *tikus, except in the case of borrowing. This makes one wonder, were *(ba)labaw and *tikus truly GNB homonyms or was *tikus a later replacement? The only reason to reconstruct both is that GNB is assumed to have been a single language which gave rise to discrete subgroups in the familiar manner. According to conventional logic, since *tikus is found in three branches of GNB it must be reconstructed, but this still does not explain why no other subgroups show any evidence of *tikus, despite its apparent presence in PGNB nor does it explain why *tikus appears to have replaced *(ba)labaw in Malayic, Southwest Sabah, and Northeast Sabah. Is it appropriate to propose a homonymic innovation when no languages show evidence of homonymy?

4.1.3.6. *guaŋ 'heart; desire'. Blust gives Bonggi *guaŋ* 'center of emotions', Idaan *guaŋ* 'chest cavity', Tagol *guaŋ* 'heart (emotional)', and Iban *guaŋ* 'sweetheart'. To this we may add Burusu *guaŋ* 'to want'. A second possible addition is Modang (Woq Helaq) *guəŋ saŋ* 'sweetheart', but the complex historical phonology of this language makes it difficult to determine if this indeed continues PGNB *guaŋ. The word is still only witnessed by a few examples, and its limited attestation remains an issue.

4.1.3.7. *laŋkaw 'large temp. shelter'. Apparent reflexes of *laŋkaw* are found throughout Ibanic, including in Keninjal *laŋkaw* where it means 'granary', Seberuang *laŋkaw* where it means 'house' but may be modified as *laŋkaw uma* 'field hut', and Mualang *laŋkaw*, where the primary meaning is

'field hut'. The word was borrowed from Tamanic, where Taman has *laŋko* 'house'. The monophthongization of the final *-aw suggests that borrowing was quite early. This makes the presence of a possible cognate in Rara Bekati', *laŋko* 'granary', difficult to analyze, since although it has a mono-thphized diphthong there are no other Land Dayak languages with a potential reflex. It may have been borrowed early in Rara as well and therefore should not be considered evidence linking Land Dayak to GNB.

4.1.3.8. *sakay. Blust (2010) lists *qaRta 'outsider' > *sakay 'outsider' as a replacement innovation. This, however, is not an accurate representation of the innovation. PMP *qaRta was replaced before any putative PGNB language by *qulun, a term which is widespread in WIN. From a conventional subgrouping perspective, *sakay did not replace *qaRta, which had already been replaced, nor did it replace *qulun, which is retained in numerous GNB languages. It is therefore not a lexical replacement innovation.

4.1.3.9. *saʔay. Blust (2010) lists PMP *bakbak (?) > *saʔay 'kind of frog with a loud croak' as a lexical replacement innovation. Reflexes of *bakbak are found only in the Philippines, perhaps explaining the question mark that Blust places next to the reconstruction that he assigns to PMP. Words that are restricted to the Philippines are not typically considered viable PMP recon-structions. The gloss for *bakbak in the *ACD* is simply 'frog'. Considering both the restriction of *bakbak to the Philippines and its different meaning, *saʔay is not a true replacement innovation.

4.1.3.10. *sapaw. Another issue in Borneo is the presence of a well-attested lexical replacement innovation that directly contradicts the GNB-BB split, the *sapaw 'field hut' > *sapaw 'roof' innovation. In all languages of Borneo except Malayic and Southwest Sabah, the word *sapaw replaced PMP *qatəp as 'roof", and *qatəp was lost. Malayic and Southwest Sabah both retain PMP *qatəp 'roof'. This is a replacement innovation, which makes it potentially strong for subgrouping, but it joins Barito and GNB to the exclusion of Malayic and Southwest Sabah. Does the existence of *sapaw 'roof' force one to propose a new subgroup that crosscuts GNB and is thus incompatible with it? No, it does not. Rather, its existence shows how lexical replacement innovations appear to have spread throughout Borneo without much regard to subgrouping boundaries. Even though lexical replacement innovations have the potential to form strong subgrouping evidence, they can give contradictory results. This calls into question the very premise that lexical innovations alone are enough to unite a large subgroup like GNB.

4.1.4. What remains of the GNB lexical innovations. A list of remaining, removed, and doubtful GNB lexical innovations is given below, along with the subgroups in which they are found. There are eighteen remaining lexical innovations which have few or no issues. There are four weak innovations which are not included in the list of eighteen remaining innovations. Finally,

there are eight innovations that are clearly not GNB innovations and are removed outright.[7]

Remaining (18)

1	*damək 'blowpipe dart'	CS, MAL, SwS
2	*laŋkaw 'large temp. shelter'	MAL, SWS
3	*guaŋ 'heart; desire'	MAL, NES, SWS
4	*tikus 'rat'	MAL, NES, SWS
5	*təgap 'firm; sturdy'	KAY, LD, MAL, NS, SWS
6	*lamin 'room of a house'	KAY, LD, MAL, NS SWS
7	*kuju 'heron'	KAY, MAL, NS, SWS
8	*labi 'soft shelled turtle'	CS, KAY, MAL, NS, SWS
9	*sakay 'stranger'	CS, KAY, MAL, NS
10	*tujuʔ 'seven'	CS, KAY, MAL, NES, NS, SWS
11	*gaduŋ 'grue'	CS, MAL, NES, NS, SWS
12	*saʔay 'frog'	CS, KAY, LD, NES, NS, SWS
13	*kuyad 'long tailed macaque'	CS, KAY, NS
14	*təməduR 'rhinoceros'	CS, KAY, NS
15	*kəlit 'small bat'	CS, NS
16	*ñaRu 'eagle'	CS, KAY, Bulungan
17	*saʔup 'parang handle'	CS, KAY, NS
18	*kitan 'binturong'	CS, KAY, NS

Weak (4)

*tukul 'hammer'	CS, KAY, MAL, NS, SWS (Plus Barito)
*sulap 'temporary shelter'	LD, MAL, SWS
*puʔər 'squirrel'	CS, NS
*ajən 'fish'	CS, KAY

Removed (8)

*kəraʔ 'long tailed macaque'	MAL, SWS
*ambay 'sweetheart'	LD, MAL, SWS
*alud 'canoe'	LD, NES, NS, SWS
*kadis 'grasshopper'	CS, NS, LD
*lipəs 'cockroach'	CS, KAY, MAL, NES, NS, SWS
*kuini 'mango species'	CS, KAY, LD, MAL, NS
*cəRaʔuŋ 'sun hat'	CS, KAY, LD, NS
*lunək 'soft; mushy'	LD, MAL, NS, SWS

4.2. DISTRIBUTION OF LEXICAL INNOVATIONS IN GNB AND BB.

The next task in evaluating the lexical evidence for GNB is to examine the distribution of lexical innovations in GNB subgroups. A summary of the distribution of the eighteen remaining innovations is presented in table 3 below. A "1" indicates the presence of the innovation, and a "0" indicates the absence of the innovation. Cells with a "1" are further shaded to make the distributions clear.

There are clear zones where these innovations are concentrated in certain subgroups. Southwest Sabah and Malay share numerous innovations (with

---

7. CS = Central Sarawak. MAL = Malayic. SwS = Southwest Sabah. NES = Northeast Sabah. KAY = Kayanic. LD = Land Dayak. NS = North Sarawak. KEN = Kenyah. UKM = Upper-Kapuas–Mahakam. PUN = Punan. MEL = Melanau. KAJ = Kajang. BLB = Berawan–Lower Baram. DAI = Daic (Lun Dayeh, Kelabit). BIN = Bintulu. BAS = Basap. SwB = Southwest Barito. NEB = Northeast Barito. TUN = Tunjung. NwB = Northwest Barito. SEB = Southeast Barito.

the possible inclusion of NES), suggesting a particularly strong Malay influence, for example. Kayan and Kenyah have nearly identical innovation sets. Berawan–Lower Baram and Daic also share a certain closeness, as do Upper-Kapuas–Mahakam and Punan. These shared innovations do not imply a unique genetic relationship, since they are rarely exclusively shared, but rather show a closeness in the larger network of GNB languages and the diffusion of lexical innovations within this network. Note, however, that in some cases these shared lexemes do overlap with already established subgrouping boundaries, as is the case between Upper-Kapuas–Mahakam and Punan, and indeed Central Sarawak as a whole. Overall, the distribution of lexical innovations more closely resembles the linkage-like distribution of innovations discussed earlier.

4.3. WHERE DOES LAND DAYAK FIT? Perhaps the most interesting result of this reevaluation of lexical evidence is the placement of Land Dayak in GNB. Most GNB languages fit somewhere along the distribution of lexical innovations shown in table 3. Land Dayak, however, has only three GNB innovations, and these innovations do not fit cleanly in the distribution in table 3. Previous Land Dayak evidence has been mostly thrown out as it was shown to contain phonological irregularities that were probably the product of borrowing from Malay. What remains is not a convincing number of innovations for what should be a GNB language like any other. Note that Bintulu also

TABLE 3. DISTRIBUTION OF WELL-SUPPORTED GNB LEXICAL INNOVATIONS.

	SwS	Mal	Ken	Kay	UKM	Pun	Mel	Kaj	BLB	Dai	NeS	Bin	Bas	SwB	NeB	Tun	LD	NwB	SeB
damək	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
laŋkaw	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
guaŋ	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
tikus	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
təgap	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
lamin	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
kuju	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
labi	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
sakay	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
tujuʔ	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
gaduŋ	1	1	0	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0
saʔay	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0
kuyad	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
təməduR	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0
kəlit	0	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
ñaRu	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
saʔup	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
kitan	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0

shares a few innovations with GNB, but this is due to an overall lack of Bintulu data. Land Dayak, on the other hand, is well documented with several large lexical resources yet shares only three innovations with GNB languages, the same number as Basap.

Where should a subgroup like Land Dayak be placed in a larger subgrouping proposal for Borneo? With no clear indication that Land Dayak belongs to any specific group, it is perhaps best left unclassified, that is, a subgroup with no special genetic affiliation with any other subgroup.

4.4. CONCLUSION: GNB IS A LEXICAL DIFFUSION AREA. There is only one GNB lexical replacement innovation, PMP *qantipa > *labi 'soft shelled turtle'.[8] Innovations that were formerly thought to have been replacement innovations have been either ruled out of consideration altogether or shown to be homonymic innovations. Reflexes of *labi are geographically restricted; they are found in Southwest Sabah, Malayic, Kenyah, Kayan, Upper-Kapuas–Mahakam, Punan, and Melanau, but not in Kajang, Berawan–Lower Baram, Daic, Northeast Sabah, or Land Dayak. The distribution of the non-replacement innovations is also greatly restricted, with each showing clear zones of influence. These facts force the rejection of GNB as a traditional linguistic subgroup. There was no PGNB language, and GNB itself is an innovation-defined zone of lexical diffusion that is the product of contact and borrowing. GNB languages therefore do not share an exclusive common ancestor.

5.   WIN. Like GNB, WIN must be reevaluated due to its overreliance on lexical innovations. There are also issues with the inclusion of non-Bornean languages in the subgroup, addressed below. It is concluded that WIN must be reanalyzed as an innovation-defined lexical group restricted in scope to only the languages of Borneo plus only those non-Bornean languages that originate from Borneo, excluding Tamanic.

5.1. SUMATRA, JAVA, MADURA, BALI, AND LOMBOK. As originally conceived by Blust (2010), WIN was far-reaching and included all Bornean languages plus all AN languages to the west and south of Borneo, excluding both Sulawesi to Borneo's east (including the Tamanic languages) as well as Central-Eastern Malayo-Polynesian. Smith (2017a) refined WIN and argued for the exclusion of Moken, the Sumatran languages (Batak, Sumatran Barrier Island languages, Nasal), and Lampung. Smith's classification of WIN kept the

---

8. Blust (2010:73) notes that Tombonuwo *ontipo* 'type of large tortoise' suggests that *qaCipa was retained, perhaps with a semantic shift to 'tortoise'. So even this word is not without its issues. If it shifted from 'soft shelled turtle' to 'tortoise', with *labi filling in the new semantic gap, then it may still be a replacement along the same lines as *tuzuq is a replacement, at least as it was originally conceived. The more problematic issue is the existence of Tawoyan *ləwawiʔ*. Tawoyan is a Barito language, so this word could invalidate the status of *labiʔ as a GNB innovation. However, the reflex is irregular, with an extra syllable, and restricted to Tawoyan. An apparent cognate is also found in Toba Batak *labi* 'turtle' (Stap 1912), although the phonology does not allow one to evaluate its status as native or borrowing (ancient or recent).

languages of Java, Bali, and Lombok, Madura, and Rejang within WIN, but noted, while discussing the lexical evidence for including these languages, that "one must wonder if this limited evidence is enough to justify their inclusion" (Smith 2017a:454). Adelaar (2023) takes up this issue in far more detail, casting serious doubt on the inclusion of these languages in WIN.

A closer examination of the lexical evidence from the new perspective argued for in this paper reveals little justification for keeping these languages within WIN. Balinese, for example, reflects a single lexical innovation, *kəjut 'surprised; startled', but this word, as discussed more below, is ultimately a retention of PAN *kəzuC. Sasak was also included in WIN with only a single innovation, *butbut 'coucal', but a single innovation in the context of AN ISEA is hardly convincing. There is also the issue of the possible parallel innovation of this word, since *butbut is an imitative, mimicking the main song of the coucal. Javanese reflects four innovations, but considering the degree of contact between Malay and Javanese, these four shared lexical innovations still do not form convincing evidence. Adelaar (2022) demonstrates that some dialects of Rejang do not in fact merge *j with *d, causing immediate issues for its inclusion in WIN. The simplest conclusion to draw from these observations is that these languages should not be included in "Western Indonesian"; therefore, the term will only refer to the languages of Borneo and not to those languages which are spoken outside of Borneo for the remainder of this paper.

5.2. THE EVIDENCE FOR WIN. Like GNB, WIN is largely defined by lexical innovations. A list of those innovations is shown in table 4.

Smith (2017a) includes some phonological evidence for the subgroup as well, but the strength of this evidence is limited and cannot hold its own without the accompanying lexical evidence. That evidence is described in (1).

(1) Merger of PMP *d and *j as PWIN *d
Assimilation of heterorganic nasal+obstruent clusters into homorganic clusters

TABLE 4. WIN LEXICAL INNOVATIONS.

*əluŋ	'river mouth; estuary'	*kubuŋ	'flying lemur; flying fox'	*kəlabət	'gibbon'
*jaʔa	'chin; jaw'	*bə[d/R]uk	'pig-tailed macaque'	*dəŋən	'river otter'
*suŋay	'river'	*Rimbaʔ	'primary forest'	*pəŋanən	'python'
*tupay	'tree shrew; squirrel'	*qulun	'outsider'	*pəlanuk	'mousedeer'
*pinaŋ	'betel nut'	*biRuaŋ	'sun bear'	*kəjut	'surprised; started'
*duRian	'durian'	*kəlasi	'red leaf monkey'	*puRaʔ	'crab'
*butbut	'coucal'	*buRis	'silver-leaf monkey'	*kəniw	'eagle'
*[t/k]iuŋ	'myna bird'	*təlaʔus	'barking deer'	*ukəd	'western tarsier'
*kəRiw	'orangutan'	*kuliR	'clouded leopard'	*ma-tuRun	'binturong'
*giRam	'river rapids'	*iban	'reciprocal affine'	*[t/s]iliŋ	'to fly'
*gətəm	'harvest'	*li(ŋ)kaw	'brow'	*madam	'rotten'
*bə-təRiʔ	'pregnant'				

WIN languages are therefore those languages which merge *d and *j as *d, assimilate heterorganic nasal+obstruent clusters, and also reflect at least some of the WIN lexical innovations shown above. A language that has the phonological mergers but not the lexical evidence, or a language that has the lexical evidence but keeps *d and *j distinct, cannot fit into WIN as it is defined by Smith (2017a).

The number and quality of the lexical innovations and the inclusion of at least some phonological evidence for the subgroup give WIN a more legitimate footing than GNB for serious consideration. Most of the lexical innovations remain valid even after closer inspection, and only four may need to be removed. Those four, *tupay, *pinaŋ, *iban, and *kəzut, are discussed below.

5.2.1. *tupay 'squirrel'. Blust (2010) lists *tupay with evidence from Malayic *tupay*, Ngaju Dayak *tupay*, and Bonggi *tufi*. There are also some Land Dayak words that Blust did not include, such as Ribun and Sanggau *tupay*, but the development of the final diphthong marks Land Dayak witnesses as borrowings. In fact, the Barito evidence originally provided by Blust also shows phonological irregularities indicative of borrowing. Ngaju Dayak reflexes of *-ay undergo regular raising to –*ey*, which is absent in Ngaju Dayak *tupay*. Similar issues are found in other Barito evidence. The only words that look native are the Malayic words plus Bonggi *tufi*. However, one must question the validity of an apparent subgroup-defining innovation that is found almost exclusively in a single subgroup with only one external witness. Since Sundanese retains PMP *buhət 'squirrel', the *tupay innovation is most likely a Malayic innovation and one which spread to some non-Malayic languages via borrowing.

5.2.2. *pinaŋ 'betel nut'. Blust proposed *pinaŋ as an innovation that replaced *buaq as a word for 'betel nut'. This was not a proper replacement, however, since *buaq had a wider meaning of 'fruit'. Rather, it was a was a semantic narrowing. The main issue with *pinaŋ is that there are competing reconstructions for 'betel nut', including *gəRat (Ngaju *gehat*, Basap *kərat*, Ngorek *gahat*, Kayan *gahat*, Kelai *gəhæt*, Lahanan *gaat*, Beketan *gehet*, Bulungan *gərat*) and *paʔan (Tunjung *paatn*, Busang Kayan *paʔan*, Punan Tubu *paʔan*, Aoheng *paʔan*). This situation, where there are numerous competing reconstruction spread throughout Borneo for a single item is a situation which one expects to form from multiple instances of contact, possibly with non-AN populations, after initial settling populations had already spread out. This casts doubt on the status of *pinaŋ as a subgroup-defining innovation.

5.2.3. *kəzut 'startle'. Blust also argues that *kəjut 'startle' is a PWIN innovation, but this word is reconstructable to PAN *kəzuC 'to jerk suddenly, jump when startled' with Paiwan *mi-kəduts*. Reflexes in Borneo are therefore retentions.

5.2.4. *iban 'reciprocal affine'. This word was proposed in Smith (2017b) as an innovation meaning 'parent-in-law; child-in-law'. It is now clear that this word follows directly from a suffixed form of PMP *ibah 'companion, close

relative, other one', which gave *ibah-an and eventually *iban* through reduction to two syllables after h-deletion and vowel coalescence.

5.3. THE DISTRIBUTION OF THE LEXICAL EVIDENCE. The distribution of WIɴ innovations in Bornean languages contrasts with the distribution of GNB evidence as shown here in table 5. As expected, there is much more representation in BB. Twenty-two innovations are found in Northeast and Northwest Barito, twenty-one in Southwest Barito, twenty in Basap, and Southeast Barito, and sixteen in Tunjung. Land Dayak is also much better represented, with twenty-one innovations. Overall, there is more robust attestation

TABLE 5. WIɴ LEXICAL INNOVATION DISTRIBUTION IN BORNEAN LANGUAGES.

	KAY	PUN	KAJ	NEB	NWB	UKM	SWB	Ken	LD	SEB	BAS	MEL	TUN	BLB	Day	SWS	MAL	NES	BIN
bə[d/R]uk	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
pəlanuk	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
suŋay	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
biRuaŋ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
qulun	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
kubuŋ	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0
dəŋən	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	0
[t/k]iuŋ	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0
durian	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0
kəlabət	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1	0	0	0
butbut	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0	1	1	0	0
kuliR	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	0	0	0	0
peŋanen	1	1	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
əluŋ	1	1	1	1	1	1	0	1	1	0	0	0	0	1	1	0	0	1	0
jaʔa	1	1	1	0	1	1	0	1	0	0	0	1	0	1	1	1	0	1	0
təlaʔus	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0
bə-tərRiʔ	0	1	1	0	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0
kəlasi	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	0	1	0	0
li(ŋ)kaw	1	1	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0
madam	1	1	1	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0
giRam	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0
rimbaʔ	0	0	0	0	1	1	1	1	1	0	1	0	0	1	0	0	1	0	0
kəniw	0	1	0	1	0	1	0	0	0	0	0	1	0	1	1	1	0	1	0
gətəm	1	0	0	1	0	1	1	0	1	0	1	0	1	1	0	0	0	0	0
buRis	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
ukəd	0	1	1	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0
kəRiw	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0
ma-tuRun	0	0	1	1	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0
[t/s]iliŋ	1	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0
puRaʔ	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0

all around and the distribution is much more random. Malay and Sabahan languages are less represented than they were in the GNB chart. Bintulu is an outlier, with only three innovations, but this is due to a lack of data, not a genuine lack of relevant innovations.

When compared to the distribution of GNB lexical innovations, the WIN innovations are more evenly distributed, resulting in an inability to parse centers of diffusion from the table. BB subgroups, as well as Land Dayak, are not easily separated from the other languages. If these innovations are the result of lexeme spread throughout a network, then the spread appears to be more even and not associated with any center of dispersal.

5.4. QUALITY OF THE WIN LEXICAL EVIDENCE. Regarding the quality of the lexical evidence for WIN, most innovations are novel concept innovations which were apparently innovated (or possibly borrowed) to describe new species of plants and animals that were encountered after the original AN settlers moved past the Wallace line into Borneo. Three lexical replacement innovations can be identified, including PMP *qazay 'chin; jaw' > PWIN *jaʔa 'chin; jaw', PMP *paniki 'flying fox' > PWIN *kubuŋ 'flying fox', and PMP *qaRta 'outsider' > PWIN *qulun 'outsider'. Both *qulun and *kubuŋ are widely attested, but *jaʔa is somewhat less widely attested. Some words, like *təlaʔus 'barking deer' and *kuliR 'clouded leopard' may be interpreted as replacements (replacing PAN *sakəC and *lukəNaw respectively), but a lack of Philippine witnesses suggests that these words were not present in PMP, which lost them after speakers moved to the Philippines where these animals are absent. The concepts would have been reintroduced to AN speakers as they moved into Borneo, making these novel concept innovations.

The interpretation of these data is complicated. On the one hand, there is more evidence of a more even distribution among WIN languages, suggesting more cohesion in the group than when compared to GNB. On the other hand, the evidence mostly consists of novel concept innovations, with few lexical replacement innovations to speak of. WIN is a more well-established group, and its influence has spread throughout all languages of Borneo, perhaps from the time of initial settlement. This differs from GNB, which arose in a lexical diffusion zone at a later date.

5.5. REDEFINING "WESTERN INDONESIAN." With so much of Western Indonesia removed from the group, one may wonder what is left, and if "Western Indonesian" remains an appropriate title. It would be more accurate to use a term exclusive to Borneo. Although the term "Bornean language" has historically been used as a geographical title, it should now replace "Western Indonesian" in describing these languages. Bornean languages belong to a unique Bornean lexical group, that is, a group of languages that reflect the mostly novel concept innovations that were introduced to the language of the first PMP-speaking settlers to arrive on the island. Following the logic of the earlier thought experiment in section 3.2.1, the languages of

Borneo descend from this settling population of PMP-speakers, but because there was no unique Proto-Bornean language, there is no justification for grouping the languages of Borneo into an exclusive subgroup. The Bornean languages descended from PMP and reflect a regional-specific vocabulary that existed among the initial settling population, but they do not share an exclusive ancestor language.

6. TYING UP A LOOSE END: THE NORTH BORNEO SUBGROUP. Until now this paper has dealt with the issue of lexically defined subgroups and their reanalysis but there is one loose end to tie up before concluding; the issue of North Borneo and the terminal devoicing of PMP voiced plosives. North Borneo (NB, PNB), as conventionally conceived, contains Southwest Sabah (SwS, PSwS), Northeast Sabah (NES, PNES), and North Sarawak (NS, PNS). This section argues that terminal devoicing was incorrectly used to justify two subgroups in a parent-sibling relationship (North Borneo and North Sarawak), rendering the larger North Borneo subgroup invalid. This section begins with a discussion of North Sarawak, since the fate of North Borneo is closely intertwined with the fate of North Sarawak.

6.1. NORTH SARAWAK. North Sarawak has its genesis in early publications by Blust (1969, 1974), who first noted a split in reflexes of voiced plosives in intervocalic position in the languages of northern Sarawak that result in odd reflexes. Blust's original hypothesis on the origin of this split involved a series of vowel deletions and consonant cluster formations in a new set of trisyllabic PAN reconstructions. Although the trisyllable origin hypothesis has been abandoned and the voiced plosive split is now considered to be primarily motivated by the gemination of consonants after a penultimate schwa (Blust 2010), the split itself is still considered to have been an innovation that occurred in an ancestral language now named Proto-North Sarawak (PNS).

This split in voiced obstruents is the only piece of linguistic evidence put forward for a North Sarawak subgroup. The first ingredient in the PNS split is geminate consonants, which PNS inherited from two sources. First, consonants at the onset of the final syllable underwent gemination if the penultimate syllable was open and contained a schwa nucleus.

(2) PMP *dəpah 'a fathom' → pre-PNS *dəp:a
    PMP *təbuh 'sugarcane' → pre-PNS *təb:u

Second, heterorganic consonant clusters, mostly from reduplicated monosyllables, assimilated and formed geminates regardless of the penultimate vowel.

(3) PMP *butbut 'pluck, pull out' → pre-PNS *bub:ut

Pre-PNS therefore had singleton and geminate consonants in the medial position. These geminates were somewhat predictable, occurring mostly after a penultimate schwa, but not completely predictable since those formed from consonant clusters could appear after any vowel.

TABLE 6. TYPICAL REFLEXES OF VOICED PLOSIVES IN MEDIAL
POSITION IN NORTH SARAWAK LANGUAGES.

PMP	*b		*d/*j [gʲ]		*z [dʒ]		*g	
pre-PNS	*b	*bː	*d	*dː	*j [dʒ]	*jː [dʒː]	*g	*gː
PNS	*b	*bʰ	*d	*dʰ	*j	*jʰ [dʃ]	*g	*gʰ
Bario Kelabit	b	bʰ	d	dʰ	d	dʰ	g	gʰ
Long Semadoh	b	bʰ	d	dʰ [dʃ]	d	dʰ [dʃ]	g	gʰ
Lebo' Vo'	ɓ	ɓ	ɗ	ɗ	ʃ	ʃ	ɠ	ɠ
Lepo' Tau	b	p	d	t	j	c	g	k
Penan	v	b	r	d	j	j	g	g
Kiput	b	s	r	s	c	s	g	k
Miri	b	f	d	s	j	s	g	k
Bintulu	v	ɓ	ɣ	ɗ	j	ʃ	g	ɠ

    The next ingredient in the split is terminal devoicing. Blust (2010) argues
convincingly that the voiced geminate plosives underwent a process of terminal
devoicing in PNS, resulting in what he describes as "true voiced aspirates," that
is, stops that begin voiced, end voiceless, and have an extended voice onset
time. These consonants are transcribed as PNS *bʰ, *dʰ (from PMP *d and
*j, which had already merged), *jʰ (from PMP *z [dʒ]), and *gʰ. The voiced
aspirates persist in some languages, for example, in the Bario dialect of
Kelabit and the Long Semadoh dialect of Lun Dayeh. In other languages,
the voiced aspirates have unexpected and complex reflexes, sometimes devoic-
ing entirely, sometimes becoming implosives, and other times shifting to sibi-
lant articulations. Table 6 summarizes some of these reflexes. Several North
Sarawak languages representing all first-order subgroups are given with their
respective reflexes of the voiced plosives in medial position.
    Other than this split in voiced plosives, there has been no attempt to further
defend a North Sarawak subgroup. Blust (2010:51) considers this change
unique enough to warrant a subgrouping proposal and states that "the distinc-
tiveness of this change, and hence the low probability that it would be a product
of convergence, provides strong evidence for a North Sarawak subgroup." The
strength of this evidence, according to Blust, hinges on its uniqueness.
Attempts to identify lexical innovations that are exclusive to the North
Sarawak group have proven frustrating, so North Sarawak truly is a subgroup
totally defined by a single sound change.

6.2. NORTH BORNEO. Much like North Sarawak, the North Borneo hypoth-
esis has been around for some time, beginning with Blust (1974) but not being
truly fleshed out until Blust (1998) and especially Blust (2010). The main con-
tribution of the North Borneo hypothesis, especially early on, was the recognition
that languages of Sabah, despite sharing many typological traits with languages
of the Philippines, actually subgroup with the other languages of Borneo. Today,
Sabahan languages are rarely included in Philippine subgroupings.

**TABLE 7. REFLEXES OF VOICED PLOSIVES IN MEDIAL POSITION
IN BARIO KELABIT AND IDAAN BEGAK.**

PMP	*b		*d/*j/*z		*g		*R	
Bario Kelabit	b	bʰ	d	dʰ	g	gʰ	r	
Idaan Begak	b	bp	d	dt	g	gk	Ø	gk

North Borneo has three primary divisions, Southwest Sabah, Northeast
Sabah, and North Sarawak. Southwest Sabah includes the Murutic, Paitanic,
and Dusunic languages of Sabah and is by far the largest of the two
Sabahan groups. Northeast Sabah includes Bonggi, spoken on Banggi Island
at the northern tip of Borneo, and Idaan varieties spoken in eastern Sabah.
There is some disagreement about the validity of Northeast Sabah, but for
now it is assumed that Bonggi and Idaan form a subgroup. See Lobel
(2013, 2016) for an alternate view that groups Bonggi with Philippine lan-
guages, and Smith (2017a) for a defense of Northeast Sabah. Either way,
Idaan varieties are Sabahan (that is, not Philippine), and this is where the inter-
esting historical phonology is to be found.

The key piece of evidence for a North Borneo subgroup is a correspondence
between Idaan and North Sarawak languages. Both reflect the same split in
voiced plosives in the same environment. Some examples are shown below
with reflexes of PMP-voiced plosives in Bario Kelabit and Idaan Begak. An
important difference between these two languages is that Idaan Begak also
has voiced-voiceless clusters *gk* which originate from PMP *R. In Idaan
Begak *R deleted in medial position except after schwa, where it became
gk. In the final and initial positions, *R became *g*. The two are compared in
table 7.

Blust (2010:57) sees this as evidence for a larger subgroup, North Borneo:
"It seems likely, then, that correspondences such as PNS *bʰ : IB [Idaan Begak]
bp or PNS *dʰ : IB [Idaan Begak] *dt* reflect an innovation in a single proto-
language." That innovation, according to Blust, is the terminal devoicing of
voiced geminate plosives.

A potential problem for this hypothesis is the fact that Southwest Sabahan
languages, which Blust includes in NB, do not have the type of complex
reflexes of *bʰ and other voiced aspirated stops found in Northeast Sabah
and North Sarawak. Blust makes two observations which he claims supports
the hypothesis that SwS languages descend from an ancestor that had terminal
devoicing. First, Blust (2010:57–60) shows that after a schwa, *b and *d/*j are
reflected mostly unchanged in SwS but after other vowels they typically have
lenis reflexes such as v/β, w, r, or Ø. Blust (2010:60) further notes that native
vocabulary in Southwest Sabah tends to have implosive pronunciations of the
voiced stops *b* and *d* and that these implosives may have been generalized to
stops from a past voiced aspirate. A similar history is evident in some Lowland
Kenyah languages (see reflexes of voiced obstruents in Lebo' Vo'; table 6).
These observations are used to justify positing a series of PNB innovations

TABLE 8. STAGES IN THE DEVELOPMENT OF MEDIAL VOICED
PLOSIVES IN NORTH BORNEO.

PMP	*b		*d/*j		*z		*g	
PGNB	*b	*bː	*d	*dː	*j	*jː	*g	*gː
PNB	*b	*bʰ	*d	*dʰ	*j	*jʰ	*g	*gʰ
PSwS	*b	*ɓ	*d	*ɗ	*d	*ɗ	*g	*g(?)
PNᴇS	*b	*bʰ	*d	*dʰ	*d	*dʰ	*g	*gʰ
PNS	*b	*bʰ	*d	*dʰ	*j	*jʰ	*g	*gʰ

*bː > *bʰ, *dː > *dʰ, *jː > *jʰ, and *gː > *gʰ. My interpretation of Blust's arguments is presented in table 8.

In addition to the single piece of phonological evidence, Smith (2017b) attempts to assemble an accompanying list of lexical innovations in support of North Borneo. Unfortunately, there are few innovations to speak of. Only three were proposed, and they are not particularly strong.

The best of the three is an innovation that apparently filled a semantic gap. As Smith (2017b) points out, PMP had a word for 'pig (general term)', *babuy; a word for 'domesticated pig', *bəRək, but no word for 'wild pig'. Both Sabahan subgroups and all North Sarawakan subgroups (excluding Kenyah), reflect the term *bakas 'wild boar, wild pig' which fills this gap. The other two apparent innovations are *tulud 'to fly', with evidence found in Dusunic, Paitanic, Idaan, Bulungan, Daic, and Berawan–Lower Baram, and *ləbas 'naked', with evidence from Dusunic, Paitanic, Murutic, Northeast Sabah, and Daic. PNB *ləbas is especially suspect, since it is found mostly in languages of Sabah and appears only in the Daic group of North Sarawak, the group which also happens to be closest to Sabah and with which Sabahan languages have had closer contact. It could easily be a borrowing in Daic.

6.3. THE TROUBLE WITH NORTH BORNEO AND NORTH SARA-WAK.

The use of terminal devoicing as key evidence for both North Sarawak and North Borneo is a major problem, because the same sound change cannot serve double duty as critical evidence for subgroups if one subgroup is a daughter of the other. If the voiced aspirates were a PNB innovation, then they must be retentions in North Sarawak and retentions have no subgrouping value. If they were a PNS innovation, then they must have developed independently in Idaan Begak, not in PNB, which now lacks evidence.

These observations have a devastating impact on Blust's arguments, but not all is lost. In the following sections, I argue that North Sarawak should remain as a valid subgroup but that North Borneo should be rejected. The argument proceeds as follows: (i) Terminal devoicing is not a unique sound change, so it is not necessary to assume that it occurred only once in PNB nor is it necessary to assume that its presence in Idaan and North Sarawak forces the recognition of a North Borneo subgroup. (ii) Terminal devoicing has a clear phonetic motivation, which further implies that it may have arisen through

parallel innovation. (iii) The other member of North Borneo, Southwest Sabah, has no evidence for a past stage of terminal devoicing despite Blust's arguments. (iv) Although it is clear that terminal devoicing is not unique, it is still rare. The rarity of the sound change, even if it is not totally unique, means that it may still be important for subgrouping at lower levels.

6.3.1. Terminal devoicing is not unique. Both North Sarawak and North Borneo rely on the uniqueness of terminal devoicing as a sound change as the key evidence for their validity. No matter how one slices it, however, terminal devoicing of voiced geminate obstruents must have happened at least twice in Borneo. This is because of the change *R > g created *gk* clusters in Idaan.

If the North Borneo hypothesis is assumed correct, then terminal devoicing occurred in PNB, and was inherited in its daughter languages. In the Idaanic group, *R shifted to and merged with *g, which then fed gemination and later a second terminal devoicing event if *g from earlier *R followed a penultimate-syllable schwa. Blust (2010:57) states that the motivation for this second terminal devoicing event was that "IB [Idaan Begak] apparently extended this psychologically salient type of cluster to new geminates after the change of *R > g." The change of *R to *g is known to have taken place at a later stage, since Bonggi retains *R in cases where it metathesized with the final-syllable vowel: *waart* < *uRat and *baart* < *baRat. If one accepts Northeast Sabah as a subgroup, then Bonggi forces the reconstruction of *R without the shift to *g in intervocalic position, with the shift to *g occurring later, possibly under influence from Greater Central Philippines.

If North Borneo is not accepted as a valid subgroup, then terminal devoicing again happens at least twice; once in PNS and again in Idaanic. This time, the fact that *R also experiences terminal devoicing after the shift to *g does not add an additional terminal devoicing event, since it is no longer assumed that terminal devoicing in Idaanic was inherited. Devoicing may have occurred after the shift of *R to *g in this scenario.

It is therefore likely that terminal devoicing happened twice, no matter what position one takes on the interrelatedness of languages in North Borneo. With or without PNB, terminal devoicing is a rare, but not unique, sound change.

6.3.2. Terminal devoicing is phonetically motivated. Blust's main objection to the argument that terminal devoicing occurred in parallel innovations in North Sarawak and Northeast Sabah is that terminal devoicing is known nowhere else in the AN world and therefore of a high quality for subgrouping. It is true that terminal devoicing is uncommon, but despite its lack of attestation there is a clear phonetic motivation for terminal devoicing of voiced geminate stops. The presence of phonetic motivation increases the likelihood that the sound change would repeat itself, especially in a larger linguistic area like Borneo where languages and their respective phonologies are in close contact.

The phonetic motivation for terminal devoicing is the aerodynamic voicing constraint, which is based on the observation that voiced stops involve a buildup of pressure which must be overcome in order to maintain airflow over the vocal chords (Ohala 1983). An extended delay in the release of voiced stops increases the force necessary to overcome the buildup in air pressure. This motivates such common phonological realities as the preference for voiceless over voiced geminate stops, the preference for voiceless over voiced stops in word-final position, as well as sound changes which result in the removal of voiced stops from these environments (see Blust 2018, for more on the historical consequences of this constraint).

The fact that terminal devoicing is phonetically motivated increases the chance that it may arise through parallel innovation. In fact, similar uncommon yet phonetically motivated sound changes exist in Borneo which have interesting parallels to terminal devoicing. For example, in Land Dayak and certain surrounding Malayic languages, some Barito languages, Modang, and Bonggi (as well as Chamic and some Aslian (Mon-Khmer) languages outside of Borneo) final nasal stops have developed nasal preplosion (Phillips 2005). This is a phenomenon where historically nasal-final words develop a plosive element before the nasal stop, giving rise to sound changes like *-m > –pm, -bm, or -p. This sound change occurs in languages with onset-driven, rather than coda-driven nasality and may be analyzed as a means to prevent unwanted leftward spreading nasality (Blust 1997). Does the rarity of the change warrant its usage as subgrouping evidence? The consensus on this is no, it does not, because of the shared phonetic motivators that trigger the sound change. It is rather thought to be the product of parallel development in languages with similar phonetic pressures with additional sprachbund effects.[9]

A similar history may explain terminal devoicing in Idaan and NS: terminal devoicing arose as a product of parallel innovation due to the shared phonological pressures of the languages affected within a clearly defined northern Borneo linguistic area.

6.3.3. Southwest Sabah has no evidence for terminal devoicing. Although it is true that gemination does not lead to terminal devoicing elsewhere in the AN world, it does have an effect on voiced plosive development, often resulting in similar splits in the voiced plosive series. Blust (2010:57), in support of reconstructing voiced aspirates to PNB, sates that "The essential

---

9. Another similarly rare sound change that is attested in Borneo is the devoicing of plosive stops after nasals from historical *ND clusters. Smith (2015) identifies devoicing of such clusters, *mb > mp, *nd > nt, *nj > nc, *ŋg > ŋk, as evidence for a division within Highland Kenyah languages. This change is a bit different from preplosion in that the phonetic motivation for post-nasal devoicing remains controversial. Nevertheless, an identical parallel change is also found in Murik-Merap, a group that does not immediately subgroup with Kenyah (Smith 2017c). Type B Highland Kenyah languages and Murik-Merap have no evidence linking them together, but rather have strong evidence for their separation. Once again, it appears that a rare sound change, attested in only a handful of languages, occurred in parallel innovations in Borneo as the product of a sprachbund.

**TABLE 9. EXAMPLES OF GEMINATE-TRIGGERED SPLITS
IN MP LANGUAGES.**

PMP	Kambera	
*pija	pira	how many
*qapəju	ka-pidu	gall
*tuba	tuwa	derris root
*təbuh	tibu	sugarcane
PMP	**Tiruray**	
*pajay	farey	field rice
*qapəju	fədu	bile
*labuq	lawuʔ	fall
*təbuh	təbək	to prick, inject with something
*laki	lagey	man
*dəkət	dəkət	paste or glue
PMP	**Sekapan**	
*lubaŋ 'hole'	luveə̯	hole
*təbəŋ 'fell a tree'	təbəŋ	to fell a tree

correctness of this interpretation, [that voiced aspirates were a feature of PNB], is supported by other languages of Sabah that show a similar split of the PMP voiced obstruents, although neither reflex is phonetically complex." He therefore views a split in voiced stops in SwS as being derived from an earlier stage, PNB, where voiced aspirates were present. Smith (2023) lists multiple languages where a split in voiced stops is attributable to earlier gemination, none of which necessitate an intermediate period of voiced aspiration from terminal devoicing. Such cases are found in Tiruray of the Philippines (Schlegel 1971), Kambera of Central-MP (Klamer 1998), Kajang languages of Borneo (Smith 2017b), and others. Examples are shown in table 9, with Sekapan representing the Kajang subgroup.

These observations show that the split in voiced stops in SwS could have easily emerged from geminates, and do not add to the argument that PNB had voiced aspirates. A compelling hypothesis which explains the North Borneo situation is that pre-Proto-Southwest Sabah, pre-Proto-Northeast Sabah, and pre-PNS all had voiced geminate stops. In Proto-Southwest Sabah these stops resisted the intervocalic lenition that affected singleton voiced stops. In PNS these geminates underwent terminal devoicing, giving rise to the true voiced aspirates that Blust originally reconstructed for PNS. In Proto-Northeast Sabah, the geminates remained, and later in Idaanic languages developed into complex voiced-voiceless heterorganic clusters.

6.3.4. Terminal devoicing is rare and still useful. Do these observations mean that terminal devoicing is useless for subgrouping evidence? This question can be answered with a definitive no. Terminal devoicing remains a rare sound change which could be useful for subgrouping in certain scenarios. However, the observations above point out that terminal devoicing in Idaan

and NS may be explained as occurring through parallel innovation and that inheritance is not the only explanation for its existence in both subgroups. North Sarawak, for example, might still be valid because of terminal devoicing. All North Sarawak languages reflect terminal devoicing in identical environments. There is also no troublesome merger of *R and *g. Most languages do not shift *R to g, and those that do keep *R and *g separate. North Sarawak also does not rely on assuming terminal devoicing in Proto-Southwest Sabah, again simplifying the subgrouping of North Borneo.

6.4. SUMMARY OF NORTH BORNEO. This section has shown that the only piece of evidence for the North Borneo subgroup, terminal devoicing of voiced geminate stops, is not unique. In the original proposal from Blust (2010), terminal devoicing was used as key evidence for two subgroups in a parent-sibling relationship, rendering it invalid. There are two possible ways to reconcile this: (i) recognize a North Borneo subgroup and dismantle North Sarawak, or (ii) maintain North Sarawak but do away with North Borneo. Although the issue of terminal devoicing and its usefulness for subgrouping is far from resolved, it was demonstrated above that terminal devoicing happened at least twice in Borneo and that North Sarawak and North Borneo are incompatible. The position taken here is that Idaan developed terminal devoicing independently of North Sarawak due to shared phonetic pressures on the voicing of geminate plosives. In North Sarawak, due to the subgroup-side evidence for terminal devoicing, in the form of both retentions of terminally devoiced stops as well as odd reflexes of stops such as Kiput s < *bʰ, it is assumed that North Sarawak languages developed from an ancestor with terminally devoiced voiced aspirates and that North Sarawak is still a valid subgroup.

Another possible interpretation is that both North Sarawak and North Borneo are invalid and that terminal devoicing arose not only in Idaan and North Sarawak, but also in independent innovations within North Sarawak languages themselves. This scenario calls for multiple parallel innovations of terminal devoicing, rather than only two parallel innovations. North Sarawak certainly suffers from a lack of diverse evidence, but terminal devoicing, though not unique, still appears to be sufficiently rare for one to be skeptical of widespread parallel innovation as an explanation for the correspondences found in North Sarawak.[10] In addition to traditional evidence for North Sarawak, Smith and Rama (2022) performed a Bayesian phylogenetic analysis of Bornean languages and found support for North Sarawak but not for North Borneo. Multiple methods therefore support a North Sarawak subgroup.

---

10. Lobel and Riwarung (2009) and Lobel and Hall (2010) describe voiced-voiceless homorganic clusters in South Subanon and Danao languages which have certain similarities with terminally devoiced geminates in Borneo. However, those clusters arose through place assimilation of previously heterorganic voiced-voiceless clusters, not through a devoicing process that originates in singleton voiced geminates. So, they must be considered historically distinct. It is possible that contact could have created a linguistic situation that make South Subanen and Danao more amicable to such clusters, but this is not "parallel development of terminally devoiced clusters" as discussed here.

Yet another possible interpretation is that the North Sarawak subgroups (Daic, Kenyah, Berawan–Lower Baram, and Bintulu) plus Idaanic form a subgroup with one another, which excludes Southwest Sabah and with Bonggi in an ambiguous position. This may be interpreted as a compromise position which upholds the concept of a large subgroup defined by terminal devoicing while eliminating the contradiction of having North Borneo and North Sarawak both defined by the same sound change. The implication of this proposal is that Idaanic-speaking people would have originated from an area closer to Sarawak, where the majority of North Sarawak languages are located. This alternative is potentially interesting, but it still has terminal devoicing occurring twice due to the *R > *g sound change and is not any more economical than keeping North Sarawak and Northeast Sabah separate.

Rather than splitting North Sarawak up into its component parts of Daic, Kenyah, Berawan–Lower Baram, and Bintulu, it is for now kept intact, although no longer part of a larger North Borneo subgroup. The component subgroups of North Borneo, that is, Southwest Sabah, Northeast Sabah, and North Sarawak, are now designated as three separate subgroups descended not from North Borneo but from the language of the initial AN settlers. They are therefore in a sister relationship with other major Bornean subgroups.

7. CONCLUSION: A NEW MODEL FOR BORNEAN LINGUISTIC RELATIONS. There can be no doubt that the languages of Borneo share a unique common history, but that shared history does not necessitate a single, exclusive linguistic ancestor. The shared lexicon of Bornean languages is the result of vocabulary innovation within a connected community of PMP-speaking settlers, but those innovations did not result in divergence from PMP to a single shared ancestor language. Rather, the language of PMP-speaking settlers in Borneo diverged directly from PMP, giving rise to the seven subgroups and single linkage which dominate the island today. The shared Bornean lexicon unites the languages of Borneo, and some of that shared lexicon may even be traced back to the initial AN settlement. We must acknowledge, however, that regional lexicon is not in itself indicative of a traditional linguistic subgroup, which, according to established methodology, is a group of languages descended from a single common ancestor language itself defined by phonological, morphological, and, in some cases, high-quality lexical evidence. When lexical evidence is used, it is the quality of the lexical evidence which determines its usefulness as evidence for either traditional subgroups, linkages, or areas of lexical diffusion which arose through contact. In the case of Borneo, lexical evidence shows that Bornean languages descend from PMP without a single intervening proto-language that unites all languages of Borneo, but they do contain the Borneo-specific lexical innovations that entered the vocabulary of the first PMP-speaking settlers.

Within Borneo, the GNB "subgroup" is even less cohesive. The lexical innovations which define the group are sparsely attested and distributed in such a

way as to suggest centers of diffusion within a network of languages, not inheritance from a common ancestor and not as a shared lexicon attributable to some single group. These languages therefore form an innovation-defined area of lexical diffusion, that is, an area where lexical innovations spread horizontally between languages or dialects.

Finally, the former NB subgroups are reanalyzed as equidistant sister subgroups with no exclusive common ancestor. They are, like other subgroups in Borneo, descended from the initial PMP-speaking settlers of the island. They contain both Bornean and GNB lexicons but are not descended from a PNB language with terminal devoicing of voiced geminate stops.

This new model for Bornean linguistic relations presents some trouble for visual representation. Should the lexically defined contact areas be drawn in a hierarchical tree-like representation? This might overexaggerate their closeness. Should they all be drawn as independent primary branches of MP? This is probably the better solution, but it may also overstate the independence of languages that do share some level of historical co-development. The decision on modeling is not arbitrary since it has important implications for reconstruction. In an ideal situation, any lexeme that is present in at least two primary branches may be reconstructed into the proto-language, in this case PMP. Should an innovation which exists in lexically defined zones such as those in Borneo be reconstructed to PMP because it is found in two "primary branches"? Most scholars would agree that the answer to this question is "no." For example, one cannot reconstruct *tuzuq as 'seven' to PMP simply because it exists across multiple primary divisions in Borneo, even though those divisions are "primary" in the sense that no exclusive proto-language exists between Bornean languages and PMP. Only innovations which are not exclusive to lexically defined areas are valid candidates for reconstruction into PMP. Innovations which occur within a single lexical diffusion area, even if those innovations span multiple "primary branches," should not be used as sole evidence for a PMP reconstruction.

A mixed model that incorporates both tree representations of subgroups with strong evidence and horizontally defined zones that unite these subgroups is used to try and capture Bornean linguistic relations in figure 6. A thick dashed

FIGURE 6. RAKE-LIKE TREE REPRESENATTION OF BORNEAN
LINGUSITIC RELATIONS.
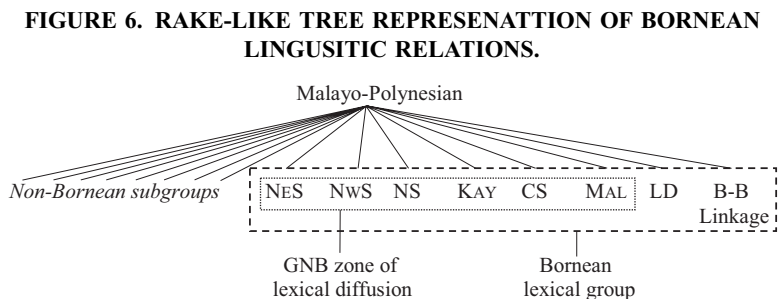
line unites all languages of Borneo in a Bornean lexical group, defined by lexical innovations that probably entered the vocabulary of the first PMP-speaking settlers of the island and were inherited into modern languages. A thin dotted line shows the approximate borders of a GNB zone of lexical diffusion, a loose collection of languages where some lexical innovations spread through various zones of contact and borrowing. Italics are used as shorthand for multiple primary-level subgroups (*Non-Bornean subgroups*) and do not represent any cohesive linguistic relation.

The arguments made in this paper for Borneo likely apply to other large subgroups of ISEA. The most obvious is the Philippine subgroup. It is likely that the innovations which supposedly define the Philippine subgroup more accurately define a zone of lexical diffusion. Hopefully, AN comparative studies may continue to move away from the traditional models and methods which have shaped western AN higher-order subgrouping to date. The result of such a shift will be a sometimes radically different, but historically more accurate, representation of linguistic relations in the area.

REFERENCES

Adelaar, K. Alexander. 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, Routledge Language Family Series 7, 1–43. London: Routledge.

———. 2023. Historical linguistics of the languages of Sumatra, Java, the Lesser Sunda Islands, and Moken Moklen. In *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*, ed. by K. Alexander Adelaar and Antoinette Schapper. Oxford: Oxford University Press.

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

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

———. 2007. *Prehistory of the Indo-Malaysian archipelago: Revised edition*. Canberra: Australian National University E Press.

Blust, Robert. 1969. Some new Proto-Austronesian trisyllables. *Oceanic Linguistics* 8:85–104.

———. 1974. The Proto-North Sarawak vowel deletion hypothesis. Ph.D. diss., Department of Linguistics, University of Hawai'i.

———. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *Working Papers in Linguistics* 9(2): 1–15.

———. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32(2): 241–93.

———. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36:149–79.

———. 1998. The position of the languages of Sabah. In *Pagtanáw: Essays on language in honor of Teodoro A. Lkamzon*, ed. by Ma Lourdes S. Bautista, 29–52. Manila: Linguistic Society of the Philippines.

———. 2001. Historical morphology and the spirit world: *qali/kali-prefixes in Austronesian languages. In *Issues in Austronesian morphology. A festschrift for Byron W. Bender*, ed. by Joel Bradshaw and Kenneth L. Rehg, 15–73. Canberra: Pacific Linguistics.

———. 2002a. Formalism or phoneyism? The history of Kayan final glottal stop. In *Between worlds: Linguistic papers in memory of David John Prentice*, ed. by K. Alexander Adelaar and Robert Blust, 29–37. Canberra: Pacific Linguistics.

———. 2002b. Kiput historical phonology. *Oceanic Linguistics* 41:384–438.

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

———. 2018. Two birds with one stone: The aerodynamic voicing constraint and the languages of Borneo. *Journal of the Southeast Asian Linguistics Society* 11(2): 1–18.

———. 2019. The resurrection of Proto-Philippines. *Oceanic Linguistics* 58(2): 153–256.

———. 2020. Response to comments on "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59:450–79.

Burkhardt, J. M. 2014. The reconstruction of the phonology of Proto-Berawan. Ph.D. diss., Institute for South-East Asian Studies, Faculty of Language and Cultural Sciences, Johan-Wolfgang-Goethe-University.

Camerini, Jane R. 1993. Evolution, biogeography, and maps: An early history of Wallace's line. *Isis* 84(4): 700–727.

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. London: Routledge.

Heggarty, Paul, Warren Maguire, and McMahon April. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3829–43.

Klamer, Marian. 1998. *A grammar of Kambera*. Berlin: De Gruyter Mouton.

Liao, Hsiu-chuan. 2020. A reply to Blust 2019 "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59:426–49.

Lobel, Jason William. 2013. Southwest Sabah revisited. *Oceanic Linguistics* 52:36–68.

———. 2016. *North Borneo sourcebook: Vocabularies and functors* (PALI Language Texts). Honolulu: University of Hawai'i Press.

Lobel, Jason William, and William C. Hall. 2010. Southern Subanen aspiration. *Oceanic Linguistics* 49:319–38.

Lobel, Jason William, and Labi hadji Sarip Riwarung. 2009. Maranao revisited: An overlooked consonant contrast and its implications for lexicography and grammar. *Oceanic Linguistics* 48:403–38.

Majid, Zuraina. 1982. *The west mouth, Niah, in the prehistory of Southeast Asia* (Special Monograph 3). Kuching: The Sarawak Museum.

Milroy, James, and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 12(2): 339–84.

Ohala, John. 1983. The origin of sound patterns in vocal tract constraints. In *The production of Speech*, ed. by P. F. MacNeilage, 189–216. New York: Springer.

Phillips, Timothy C. 2005. A survey of nasal preplosion in Aslian languages.

Reid, Lawrence A. 2020. Response to Blust "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59:374–93.

Rensch, Calvin R., Carolyn M. Rensch, Jonas Noeb, and Robert Sulis Ridu. 2012. *The Bidayuh language: Yesterday, today, and tomorrow (revised and expanded)* (SIL E-Books 33). Kuching: SIL International.

Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

———. 1997. Social networks and kinds of speech-community event. In *Archaeology and language 1: Theoretical and methodological orientations*, ed. by Roger M. Blench and Matthew Spriggs, 209–61. London: Routledge.

———. 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.

———. 2020. Comment on Blust "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59:366–73.

Schlegel, Stuart A. 1971. *Tiruray-English Lexicon*. Berkeley: University of California Press.

Smith, Alexander D. 2015. On the classification of Kenyah and Kayanic languages. *Oceanic Linguistics* 53(2): 333–57.

———. 2017a. The Western Malayo-Polynesian Problem. *Oceanic Linguistics* 56(2): 435–90.

———. 2017b. The languages of Borneo: a comprehensive classification. Ph.D. diss., Department of Linguistics, University of Hawai'i.

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

———. 2018. The Barito linkage hypothesis, with a note on the position of Basap. *Journal of the Southeast Asia Linguistics Society (JSEALS)* 11(1): 12–34. DOI: 10524/52418.

———. 2019. A second look at Proto-Land Dayak vowels. *Oceanic Linguistics* 58(1): 110–42. DOI: 10.1353/ol.2019.0005.

———. 2023. Reconstructing non-contrastive stress in Austronesian and the role of the mora in stress shift, gemination, and vowel shift. *Diachronica* 40(1): 111–52. DOI: 10.1075/dia.20032.smi.

Smith, Alexander D., and Taraka Rama. 2022. Environmental factors affect the evolution of linguistic subgroups in Borneo. *Diachronica* 39(2): 193–224.

Stap, Herm W. 1912. *Nederlandsch-Tobasche Woordenlijst*. Leiden: S.C. Van Doesburgh.

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.

Zorc, David. 2020. Reactions to Blust's "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59:394–425.

smith.alexander@gmail.com