



PROJECT MUSE®

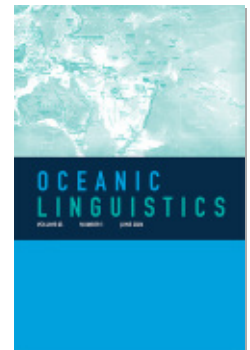
Sumatran

Blaine Billings, Bradley McDonnell

Oceanic Linguistics, Volume 63, Number 1, June 2024, pp. 112-174 (Article)

Published by University of Hawai'i Press

DOI: <https://doi.org/10.1353/ol.2024.a928205>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/928205>

Sumatran

Blaine Billings and Bradley McDonnell

UNIVERSITY OF HAWAII AT MĀNOA

The island of Sumatra and the Barrier Islands lying off its west coast are home to a diverse array of Austronesian languages. For at least a century and a half, a close genetic relationship between many of the non-Malayo-Chamic languages of the region—namely the Batak languages, Simeulue, Nias, Mentawai, and Enggano—has been proposed. Evidence in support of such a group was first thoroughly detailed by Nothofer, who outlined sound correspondences and phonological innovations for a Barrier Island–Batak subgroup. Building upon Nothofer’s proposal and recent observations about the languages of Sumatra, this paper proposes a far-reaching language group we call Sumatran, comprising Nothofer’s Barrier Island–Batak group as well as Gayo, spoken in northern Sumatra, and Nasal, spoken in southwestern Sumatra. We also provide stronger evidence for the inclusion of Enggano, spoken on the southernmost Barrier Island, which Nothofer only tentatively included. To support this proposal, we outline shared innovations that establish the foundation of what constitutes the Sumatran language subgroup.

Keywords: Sumatran Languages; Historical Linguistics; Gayo; Nasal; Enggano

1. INTRODUCTION.¹ Sumatra and the Barrier Islands (BI) off its west coast are home to a diverse array of Austronesian languages. While many of these languages are firmly established as members of a wider Malayo-Chamic group (Blust 1994), nearly two dozen lie outside of this subgroup (McDonnell and Truong, 2024). It is these languages—Gayo, the Batak languages, Haloban, Leukon, Simeulue, Sigulai, Nias, Mentawai, Nasal, Enggano, Rejang, and the Lampungic languages—that are discussed here in consideration of a larger Sumatran subgroup. Such a proposal has been present since some of the earliest linguistic work in the region (e.g., Marsden 1811:479). However, although this proposal is not novel, it has yet to be fully established and detailed in terms of the phonological innovations that are

1. We would like to thank two anonymous reviewers for valuable feedback on this chapter. We would also like to thank Owen Edwards for allowing us to use his map. We would also like to acknowledge Johan Safri, Wawan Sahrozi, and Anton Supriyadi, all of whom assisted with the Nasal data. We are also grateful to our research counterpart in Indonesia, Yanti (Atma Jaya Catholic University of Indonesia), and to the National Research and Innovation Agency (BRIN) for allowing us to conduct research on Nasal. Discussion of the Nasal data is based upon work supported by the National Science Foundation under Grant BCS–1911641. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

characteristic of the subgroup. We begin with a discussion of several previous proposals for this group of languages.

1.1. PREVIOUS PROPOSALS. There have been many iterations and names for the group of languages argued for in this paper, each comprising overlapping but different languages: Lafeber’s (1922) “Batak–Nias”; Nothofer’s (1986) “BI–Batak”; Blust’s (2013) “BI–North Sumatra”; Smith’s (2017) “Sumatran”; and several others. For the present discussion, we follow Smith’s use of the term “Sumatran” since it best encapsulates the geographic distribution and history of the subgroup. Proto-Sumatran (PSUM) is proposed as the language of the first group of Austronesian-speaking peoples in Sumatra who then spread across the island and the BI to its west, only to be later surrounded and likely displaced by Malayic- and Acehnese-speaking peoples. Of the discussions of the Sumatran group, the most detailed have been those made by Nothofer (1986) and Smith (2017).

Nothofer’s (1986) proposal of a BI–Batak subgroup is by far the most thorough. Based on shared phonological innovations, phonological irregularities, lexical innovations, and semantic innovations from Proto-Malayo-Polynesian (PMP), Nothofer considered data from the Batak languages, Simeulue (Simalur), Sigulai (Sichule), Nias, Mentawai, Enggano, and, tentatively, Gayo. Of the listed shared phonological innovations, three of nineteen were demonstrated to be common to all six languages from which data were considered. A fourth, relevant to later discussion, was demonstrated to be shared among all but Enggano. The remaining fifteen innovations were either shared among a subset of the languages or did not have enough evidence from some languages to be considered shared among all six. The four shared innovations are as follows:

- $*-j-$ > $g/x/h/\zeta/\emptyset$: listed as a “consonant shift,” a proposed change from PMP $*-j-$ (theorized as a palatalized velar [qʲ] in Blust 2013:586) to a number of phonetically close reflexes, likely via a series of successive lenition changes (e.g., $g^j > g > x > \zeta > h > \emptyset$), as discussed in Smith (2017:455–59).
- $*-\tilde{n}- > n$ and $*\tilde{n}, *n > n$: listed as two separate changes—a “consonant shift” and a “consonant merger”—these can be combined as evidence for a single shared innovation, a merger of palatal and alveolar nasals.²
- $*-Z-, *-D- > r/l/d$: listed as a “consonant merger,” evidence demonstrates reflexes of PMP $*-Z-, *-D-$ (here, $*-z-, *-d-$) are the same in all languages except Enggano, where data are insufficient.

Of the phonological irregularities and semantic innovations provided by Nothofer (1986:103–06), none can be considered shared by all languages of the BI–Batak group. Finally, more than a hundred shared lexical innovations are listed

2. The $n \sim d$ variation in Enggano appears to have been taken as counterevidence for being a shared innovation. It is considered here, however, as evidence in favor of a shared innovation (see section 3.1.9).

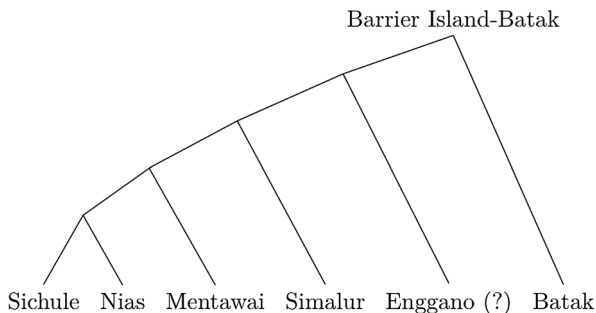
to bolster the proposal. Of these, many must be discounted for a variety of reasons: (i) they are PMP inheritances (e.g., Sigulai *uhu* ‘to string’ and Nias *susu* ‘to string’ from PMP **susun* ‘to stack up’); (ii) they consider only closely related languages (e.g., Sigulai *ufe* ‘to look’ and Nias *uve* ‘eye’); (iii) they do not follow regular sound correspondences (e.g., Mentawai *laje* ‘hungry’ and Karo Batak *lāhe* ‘hungry’ but either Mentawai ***lei* or Karo Batak ***layə*(h/s) is expected); or (iv) they are attributable to loans whether from another Sumatran language (e.g., Simeulue *tenəŋ* ‘torch, match’ borrowed as Sigulai *tini* ‘torch, match’ where Sigulai ***təno* or ***(s)hino* would be expected) or a Malayo-Chamic language (e.g., Simeulue, Gayo *gəməto* ‘wasp’ and Acehese *gömöto* ‘wasp’).

Based on the total evidence presented, an overarching BI–Batak subgroup was proposed with two primary branches—Batak and BI—and further subdivision as shown here in figure 1, reproduced from Nothofer (1986:107).

On the basis of these three rather restricted phonological innovations (one total merger, one intervocalic merger, and one intervocalic shift) shared by all members of the BI–Batak subgroup, along with a number of lexical innovations, the proposal has been widely accepted (see Blust 2013:744).

Nothofer’s proposal was revisited by Smith (2017) in a broader discussion of the Western Malayo-Polynesian languages. Three points from that paper are relevant to the present discussion. First, reflexes of PMP **j* were reinvestigated, and it was proposed that “Sumatran Barrier Islands languages, Batak, and Nasal probably merged **j* with **g*, with diverse changes thereafter,” thus strengthening the change earlier proposed by Nothofer (Smith 2017:457). Second, Enggano and, preliminarily, Nasal were subgrouped with the Batak and other BI languages, primarily on the basis of this PMP **j*, **g* > **g* merger. Third, it was remarked that “the entire [Sumatran] subgrouping proposal rests on a single sound change,” demonstrating the need to reevaluate the subgroup’s foundation (Smith 2017:459). While Nothofer (1986) did additionally describe a merger of PMP **ñ*, **n* > **n*, this certainly is neither a unique change nor one that holds strong weight in subgrouping on its own given how common it is crosslinguistically.

FIGURE 1. BI–BATAK SUBGROUP, ACCORDING TO NOTHOFFER (1986:107).

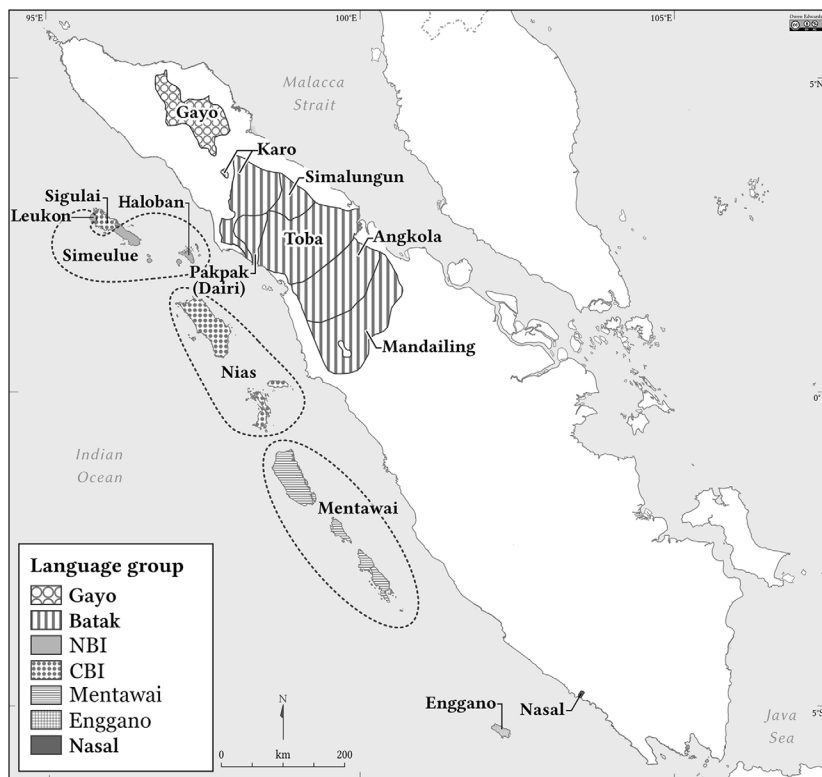


Both of these important works, however, have left a number of unanswered questions. First, as is clear from Smith's (2017) discussion, "[u]ltimately, more evidence for such a [Sumatran] subgroup needs to be located" (p. 459). Second, the place of both Nasal and Enggano in such a subgroup remains tentative. For Nasal, no thorough historical analysis has yet been carried out. Enggano, on the other hand, has been the subject of much debate. For example, Edwards (2015) argued against Enggano's membership, stating that "there are no sound changes [...] shared between Enggano and all other Barrier Islands-Batak members" (p. 89). Third, Gayo has been left almost entirely unconsidered in the subgrouping discussion. As Blust (2013) put it, "[Nothofer] includes some data from Gayō in his proposed lexical innovations without explicitly indicating that this language belongs in the proposed group" (p. 744). In Smith (2017), too, Gayo was left unmentioned. Finally, there remains much unknown regarding the phonological histories of the subgroup and its daughter languages. Though the focus of Nothofer's (1986:98) proposal was not to "give a full account of the phonological history of each of the languages," many of these are still lacking in detail.³

1.2. PROPOSAL. This paper aims to respond to the questions raised above in the following ways. First, in section 2, phonological inventories are reconstructed for seven pre- and proto-languages—Pre-Gayo (Pre-G), Proto-Batak (PB), Proto-Northern BI (PNBI), Proto-Central BI (PCBI), Pre-Mentawai (Pre-M), Pre-Enggano (Pre-E), and Pre-Nasal (Pre-N). In this section, the close relationship between certain language groups is assumed, namely the Batak languages (Dairi/Pakpak, Karo, Simalungun, Toba, Angkola, and Mandailing), the NBI languages (Haloban, Simeulue, and Leukon), and the CBI languages (Sigulai and Nias). Because the close relationship of these languages in their respective subgroups has already been established (Adelaar 1981, for Batak; Amery and Aziz 2019, for NBI; Lafeber 1922, for CBI), justification for a Batak, NBI, or CBI group is not explicitly stated, though the phonological innovations justifying these subgroups become apparent in the reconstruction of PSUM. The four remaining languages—Gayo, Mentawai, Nasal, and Enggano—are all considered in their own subsections, with only minor discussion relevant to introducing their phonologies and providing reconstructions before dialect splits, where relevant. From these pre- and proto-languages, a proposed PSUM phonology is reconstructed in section 3. Subsequently, in section 4, it is argued that six clear phonological innovations form the foundation of the Sumatran language subgroup and justify the PSUM reconstruction made in the preceding section. On the basis of these shared innovations, it is proposed that Gayo, Nasal, and Enggano, along with the more solidified Batak, NBI, CBI, and Mentawai languages, are all members of the Sumatran subgroup (figure 2). Finally, in section 5, a number of issues in reconstruction are briefly addressed, as is the place of the two remaining non-Malayo-Chamic language groups of Sumatra, Rejang and Lampungic.

3. The Batak subgroup (Adelaar 1981), Nias (Kähler 1937), and, most recently, Enggano (Edwards 2015) are exceptions to this.

FIGURE 2. LANGUAGES OF THE PROPOSED SUMATRAN SUBGROUP.



At the end of this paper, appendix A outlines the phonological inventory of PMP (as theorized in Blust 2013), divided into four tables for consonants, vowels, diphthongs, and consonant clusters. This inventory serves as the basis for the division of the analysis of historical phonology in section 3 and the comparison between PMP and PSUM in section 4.

2. RECONSTRUCTING LOWER-LEVEL PHONOLOGIES. In this section, phonological inventories are reconstructed for the various lower-level languages/subgroups here proposed to be members of the larger Sumatran subgroup. As such, this section is split into seven subsections: Pre-G (section 2.1), PB (section 2.2), PNBI (section 2.3), PCBI (section 2.4), Pre-M (section 2.5), Pre-E (section 2.6), and Pre-N (section 2.7). The three larger groups—Batak, NBI, and CBI—have already been established; the phonological innovations defining each of these subgroups are outlined in section 3. Gayo, Mentawai, Enggano, and Nasal, on the other hand, have historically been treated on their own as they do not have any immediately apparent close affinities to any of the other languages of the region (further proposals of subgrouping are discussed in section 5). The subsections regarding these four languages deal primarily with

lower-level changes that have happened either since linguistic documentation has begun or as a result of dialect divergence since a hypothetical Pre-G, Pre-M, Pre-E, and Pre-N.

Each of these seven subsections opens with a brief discussion of the language(s) before providing reconstructions with corresponding justifications. These reconstructions and discussions for the four pre- and three proto-languages serve as the foundation for the PSUM reconstructions in section 3. For PB, PNBI, and PCBI, a small number of example lexical items are cited as justification for reconstruction from each correspondence set. The data for these can be found in appendices B, C, and D, respectively. For the other four subsections, reference is made to the sources from which justifications are drawn. In section 3, where reconstructions are made to PSUM, lexical evidence is cited from all seven groups and can be found in appendix E. In sections 2–4, a number of correspondences and sound changes are discussed. The correspondences for phonemes between PMP, PSUM, and each of the seven groups are summarized in appendix F, with a broad overview summary of sound changes from PMP to all fifteen daughter languages provided in appendix G.

2.1. PRE-G. Gayo is a language with at least five distinct varieties—Bukit, Cik, Dëret, Lues, and Serbejadi—spoken in the central and southeast regions of the Indonesian province of Aceh. Two primary references serve as the basis of the analysis presented in this section—Hazeu’s (1907) Gayo–Dutch dictionary and Eades’ (2005) descriptive grammar of the Bukit variety of Gayo. Beyond these, the present analysis also draws on Shorto’s (1976) discussion of Gayo consonant correspondences.

The Pre-G reconstructed phonological inventory is given in table 1. In addition to what is shown in table 1, three nasal-stop sequences are reconstructible to Pre-G—*mp, *nt, and *ŋk. It must be noted that *b, *d, and *g are not reconstructed here for Pre-G despite there being many examples in Hazeu’s dictionary, likely the result of borrowing. This issue will be reconsidered in section 5.

TABLE 1. PRE-G RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Glott.		Front	Central	Back
Stop	vd.	^m b	ⁿ d		^ŋ g		High	i		u
	vl.	p	t		k		Mid	ɛ	ə	ɔ
Fricative			s			h	Low	a		
Nasal		m	n		ŋ		Diphthongs			
Trill			r				High			
Lateral			l				Mid			
Approximant		w		y			Low	ay, aw		

There are two primary correspondence sets identifiable among the varieties of Gayo, the first regarding diphthongs and the second “funny nasals.” The diphthongs /ai/ and /aɔ/ appear primarily in the Lues variety of Gayo and vary with the mid-high vowels /e/ and /o/, respectively, in the Bukit and Dëret varieties. In following Eades (2005:27), these are reconstructed to Pre-G *ay and *aw.

The second correspondence set regards the so-called funny nasals, phonetically post-ploded nasals that are also encountered in neighboring Acehese (Durie 1985) as well as elsewhere in Sumatra (see, e.g., Court 1970; Coady and McGinn 1982; see also Cohn and Riehl 2016, for a broader discussion of post-ploded nasals in Austronesian). According to Eades (2005), a contrast between voiced stops and funny nasals is primarily maintained by the older generations, whereas the younger generations have since merged them as plain voiced stops. These correspondences are reconstructed to Pre-G $*mb$, $*nd$, and $*ng$.

2.2. PB. The Batak languages are a group of at least seven closely related languages spoken in Northern Sumatra. The most significant early work with the Batak languages was carried out by Van der Tuuk, who wrote a dictionary (1861) and grammar (1971 [1864–1867]) of Toba Batak and identified many regular sound correspondences between the Batak languages and other Austronesian languages. Since then, the Batak languages have been the subject of numerous comparative analyses. The most thorough of these comparative works was done by Adelaar (1981), who provided a reconstruction of PB phonology and some lexical items as well as a subgrouping proposal for the Batak languages into two distinct branches, namely Northern Batak (NB)—Dairi/Pakpak and Karo—and Southern Batak (SB)—Simalungun, Toba, Angkola, and Mandailing. This detailed analysis and reconstruction remains widely accepted and is the most thorough analysis of Batak historical phonology. As such, this section presents a simplified overview of PB historical phonology following primarily from Adelaar (1981).

The PB reconstructed phonological inventory is given in table 2. Of these reconstructed phonemes, it must be noted that neither $*c$ nor $*j$ is reconstructible to word-final position (an inherited constraint from PMP; see Blust 2013:213) nor $*y$ to word-initial position.⁴ Six nasal-stop sequences— $*mb$, $*nd$, $*ng$, $*mp$, $*nt$, and $*nk$ —along with $*rC$ sequences are also reconstructed in intervocalic position based on trivial correspondence sets.⁵

TABLE 2. PB RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Glot.	Front Central Back		
Stop	vd.	b	d	j	g		High	i	u
	vl.	p	t	c	k		Mid	ə	
Fricative			s			h	Low	a	
Nasal		m	n		ŋ		Diphthongs		
Trill			r				High		uy
Lateral			l				Mid	ey	ow
Approximant		w		y			Low		

The nontrivial reconstructions and correspondence sets of PB that justify the above analysis are provided in table 3. As with all other tables throughout this

4. Note that only for PMP does $*j$ represent [g]. Elsewhere, $*j$ represents [d͡ʒ].

5. As mentioned in Adelaar (1981), Simalungun, Toba, and Angkola geminates correspond to nasal-stop sequences elsewhere. Following the conclusion made there, these are simply reconstructed as nasal-stop sequences.

paper that outline sound correspondences, the final column cites example lexical items evidencing the correspondences for each row. For correspondence sets where evidence can be found in more than five example lexical items in the associated appendix, only the first five are listed. The first correspondence, *N-S̥-S̥*, reflects word-final nasals in NB, voiced stops in Simalungun, and voiceless stops in the other SB languages. Given that trivial correspondences *N-N* and *S̥-S̥* already attest PB word-final nasals and voiceless stops, respectively, PB **S̥* is reconstructed here (i.e., *b, *d, and *g). For a similar reason, *c is reconstructed from *c-s*, a trivial *s-s* set already providing evidence for PB *s. PB *k and PB *h are reconstructed from the sets *k-h-∅* and *h-∅*, respectively, based on the common elision path *k > h > ∅*. PB *y is reconstructed from *y-∅* with an assumed reduction in Toba, and PB *l is reconstructed from *l-r* given the common occurrence of liquid assimilation. All three sets with *ə(-u)-o* are reconstructed to *ə as this reflects a canonical four vowel system (Gordon 2016:59), vowel assimilation in Karo, and the greater likelihood of a *ə > *o sound change. The *i-e* and *u-o* sets are reconstructed to *i and *u, respectively, with lowering conditioned by the following vowel in Dairi. Three final diphthongs are reconstructable to PB—*ey, *ow, and *uy—all attested by Simalungun.

TABLE 3. PB NONTRIVIAL RECONSTRUCTIONS AND CORRESPONDENCE SETS.

	PB	Env.	Dairi	Karo	Correspondence set				PB examples
C	* <i>S̥</i>	_#	N	N	S̥	S̥	S̥	S̥	(appendix B) 3, 6, 20, 37, 38, etc.
	*c		c	c	s	s	s	s	34, 52, 53, 66, 75
	*k	#_V_V	k	k	h	h	h	∅	15, 18, 19, 20, 21, etc.
	*h		h	h / # ∅ / else	h / # ∅ / else	∅	∅	∅	4, 7, 8, 29, 57, etc.
	*y		y	y	y	∅	y	y	5, 18, 46, 56, 74
	*l	...r	r	l	l	r	r	r	22, 23, 50, 55, 63
V	*ə	_h#	o	ə	o	o	o	o	4, 9, 39, 40, 68
	*ə	uC_	o	u	o	o	o	o	13, 27, 35, 36, 46, etc.
	*ə		ə	ə	o	o	o	o	3, 6, 11, 16, 17, etc.
	*i	_C[ə/u]	e	i	i	i	i	i	14, 15, 16, 17, 52, etc.
	*u	_C[ə/i]	o	u	u	u	u	u	8, 13, 20, 26, 27, etc.
VV	*ey	_#	e	e	ei	e	e	e	1, 19, 24, 33, 51
	*ow	_#	o	o	ou	o	o	o	5, 25, 31, 49, 58, etc.
	*uy	_#	i	i	ui	i	i	i	2, 12, 62, 65, 72

2.3. PNBI. It is clear from recent linguistic surveys in the region that there are at least three distinct NBI languages spoken on Pulau Simeulue and Pulau Banyak, namely Leukon, Simeulue (also known as Devayan), and Haloban. While Simeulue, the most widely spoken of the three, has received some linguistic attention in the form of a dictionary (Kähler 1961) and phonological comparison with Sigulai (Kähler 1955), there has been relatively little linguistic work in the region until recently, with a Haloban dictionary having been published in the past decade (Tim Kamus Balai Bahasa Banda Aceh 2013) and documentary work having recently been started with Leukon (Candrasari and Khalsiah 2018; Lubis and Williams 2019). These serve as the primary sources for the PNBI reconstructions below.

The PNBI reconstructed phonological inventory is given in table 4. The voiced stops, *b and *d, are not reconstructible to word-final position, and *r, *w, *y, and *ʔ are not reconstructible to word-initial position. The voiceless stop *p has a limited distribution in PNBI, only appearing word-finally or intervocally following *ə (see sections 3.1.1 and 3.1.4). Finally, five consonant sequences are reconstructed in intervocalic position—*mb, *nd, *mp, *nt, and *ŋk.

TABLE 4. PNBI RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Glott.	Front Central Back			
Stop	vd.	b	d				High	i		u
	vl.	p	t		k	ʔ	Mid	e	ə	o
Fricative			s			h	Low		a	
Nasal		m	n		ŋ		Diphthongs			
Trill			r				High			uy
Lateral			l				Mid	ew		
Approximant		w		y			Low		ay, aw	

The nontrivial correspondence sets and PNBI reconstructions are provided in table 5 below. The correspondences *w-b-b* and *w-f-f* are reconstructed to PNBI *b on the basis of the common lenition paths of *b > w and *b > f. Other common lenition paths support the reconstructions of PNBI *d and *k from the sets *r-d-d* and *x-x-k/x*, respectively. From the set *h/∅-h-∅*, PNBI *h must be reconstructed with an incipient or incomplete loss of *h in Haloban. PNBI *ʔ is reconstructed from the set *∅-∅-ʔ* with conditioned loss in Haloban and Leukon. The PNBI nasal-stop sequences *mb and *nd are reconstructed on the basis of Leukon with an assumed reduction in Haloban and Simeulue. Four word-final diphthongs are reconstructible for PNBI, all of which are reduced in Haloban and Leukon and retained with minor changes in Simeulue.

TABLE 5. PNBI NONTRIVIAL RECONSTRUCTIONS AND CORRESPONDENCE SETS.

	PNBI	Env.	Correspondence set			Examples (appendix C)
			Haloban	Leukon	Simeulue	
C	*b	#_	w	b	b	6, 7, 8, 9, 10, etc.
	*b	V_V	w	f	f	2, 3, 4, 5, 8, etc.
	*d	#_	r	d	d	1, 13, 14, 15, 16, etc.
	*k	#_, V_V	x	x	k/x	19, 21, 26, 30, 31, etc.
	*h		h/∅	h	h/∅	20, 22, 28, 32, 33, etc.
*ʔ		V ₊ Back_V V ₋ V ₊ Back	∅	∅	?	1, 10, 25, 38, 39
CC	*mb	V_V	b	mb	b	41, 44, 54, 56, 58
	*nd	V_V	d	nd	d	17, 49, 52, 57, 59
VV	*ay	_#	e	e	ay	3, 10, 26, 29, 31, etc.
	*oy	_#	e	e	oy	20, 24, 47, 48
	*aw	_#	o	o	aw	2, 12, 23, 25, 27, etc.
	*ew	_#	e	e	ew	4, 5, 45, 53, 55

2.4. PCBI. Nias and, to a lesser degree, Sigulai have both received more linguistic attention than the NBI languages. The most in-depth linguistic work with Nias started with Thomas and Taylor Weber’s (1887) Nias–Malay–Dutch dictionary. Since then, a number of other dictionaries and grammars, as well as some historical/comparative analyses (Lafeber 1922; Kähler 1937; Pätsch 1978; Brown 2001; and a recent presentation by Zobel 2021), have been published. As such, these previous analyses greatly inform the discussion below. Publications regarding Sigulai are found mainly in the work of Hans Kähler, who wrote both a dictionary (Kähler 1959) and a grammar (Kähler 1955), which includes notes on comparative phonology with Nias and Simeulue.

The PCBI reconstructed phonological inventory is given in table 6. No word-final consonants are reconstructed for PCBI, though syllable-final consonants are reconstructible in the form of the medial nasal-stop sequences *mb and *nd (but not *ŋg). In addition, the phonemes *ʔ, *x, *r, and *y are not reconstructible to word-initial position. It is worth noting that, unlike Pre-G, PB, and PNBI, word-final diphthongs are not reconstructible for PCBI.

TABLE 6. PCBI RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Glott.		Front	Central	Back
Stop	vd.	b	d		g		High Mid Low	i		u
	vl.		t			ʔ		e	ə	o
Fricative		f	s		x	h			a	
Nasal		m	n		ŋ					
Trill			r							
Lateral			l							
Approximant		w			y					

The nontrivial correspondence sets and PCBI reconstructions are provided in table 7. Common lenition paths justify the reconstruction of PCBI *b, *s, and *h from the correspondence sets *f-β*, *h/∅-s*, and *∅-h*, respectively. For the set *l-r*,

PCBI *r is reconstructed since PCBI *l is reconstructible from an existing *l-l* correspondence set, and no conditioning environment is readily found for a supposed PCBI *l to Nias *r* sound change. The historical development of Nias *b* and *d'*, compared with Sigulai *mb* and *nd*, as a derivation from nasal-stop sequences has been proposed by Catford (1988) and supported by Brown (2001) and is thus adopted here. The last correspondence set *ə-r* is reconstructed to PCBI *ə to reflect the greater likelihood of a reduced load on the mid-back vowel space.

TABLE 7. PCBI NONTRIVIAL RECONSTRUCTIONS AND CORRESPONDENCE SETS.

	PNBI	Env.	Correspondence set		Examples (appendix D)
			Sigulai	Nias	
C	*b	V_V	f	β	1, 3, 7, 16, 25, etc.
	*s		h-∅	s	6, 8, 20, 21, 28, etc.
	*h		∅	h	4, 5, 15, 16, 23, etc.
	*r	V_V	l	r	12, 13, 14, 26, 33
CC	*mb	V_V	mb	b	15, 36, 39, 40, 41
	*nd		nd	d'	34, 35, 37, 38, 42
V	*ə		ə	ɤ	1, 4, 6, 9, 10, etc.

2.5. PRE-M. Though the name “Mentawai” is often used on its own, in actuality it refers to a group of linguistic varieties that are highly diverse both in phonology and in lexicon spoken across the Mentawai Archipelago. Of the grammars and dictionaries of Mentawai, the majority focus on the variety spoken in the Pagai Islands, though this is known as the region with the least amount of linguistic diversity and is proposed to be the place last settled by the Mentawai people. Among those who have considered the diversity of the Mentawai varieties are Pampus (1989, 1994) and Lenggang (1992). It is primarily in these sources, along with Morris’s (1900) grammar and dictionary, that we find the best comparative evidence for a Pre-M, although much is still unknown in terms of Mentawai dialectology and certainly further work remains to be done in the region.

The Pre-M reconstructed phonological inventory is given in table 8. None of the voiced stops are found in word-final position, and the voiced palatal affricate *j is only reconstructed here to medial position. Neither *k nor *s is reconstructible to word-final position.

TABLE 8. PRE-M RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Front Central Back		
Stop	vd.	b	d	j	g	High	i	u
	vl.	p	t		k	Mid	e	o
Fricative			s		Low	a		
Nasal		m	n	ŋ	Diphthongs			
Lateral			l		High	iw		
					Mid	ey, ew		
					Low	ay, aw		

There are only four nontrivial correspondence sets for Mentawai. The first two concern word-final nasals and voiceless stops. As discussed in Blust (1997:158–60), while there are a number of Mentawai varieties that make a distinction between the two, some varieties have merged both as voiceless stops and others as nasals. It is based on the evidence from the varieties that maintain the distinction that Pre-M reconstructions with final nasals and final voiceless stops are made. This reconstruction is consistent with the conclusion made by Pampus (cited in Blust 1997). The third set pertains to correspondences between geminate stops and nasal-stop sequences. On the basis of the common process by which nasals undergo full assimilation to following stops, Pre-M nasal-stop sequences are reconstructed. Finally, while all Mentawai varieties have intervocalic /r/, some exhibit word-initial /d/ whereas others have word-initial /r/. In both cases, /d/ and /r/ are in apparent complementary distribution. As such, all instances of *d-r* are reconstructed to Pre-M *d for the same reasoning provided in section 2.3.

2.6. PRE-E. Enggano is a language spoken solely on the island of Enggano, the southernmost of the inhabited BI lying to the west of Sumatra. Though firmly established as an Austronesian language by Edwards (2015), its significant aberrancy compared with other languages of the region has posed issues for its subgrouping. While at one point multiple dialects were spoken on Enggano, this diversity was apparently eliminated in an island-wide dialect leveling resulting directly from contact with Europeans and subsequent colonization (Edwards 2015:56–58). Given that Enggano has changed rapidly in the past seventy years (see Yoder 2014, for these developments), the discussion here only treats the data leading up to and including Kähler’s (1987) dictionary. Much of the current analysis follows from the correspondences laid out by Edwards (2015).

The reconstructed Pre-E phonological inventory is given in table 9. Neither consonants are reconstructed to coda position nor diphthongs to word-final position.

TABLE 9. PRE-E RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Glott.	Front Central Back			
Stop	vd.	b	d				High	i		u
	vl.	p			k	ʔ	Mid	e	ə	o
Fricative						h	Low		a	
Lateral			l							
Approximant				y						

The nontrivial correspondence sets and respective Pre-E reconstructions are provided in table 10. The correspondence *f-p* is one characteristic of the dialect divide and is reconstructed to Pre-E *p on the basis of the common lenition path. Both correspondences with *d/r-d/r* and *d/r-t* are reconstructed to

Pre-E *d. For the correspondences *d/r-d/r*, *d is reconstructed on the basis that *r* is an apparent free-variant of *d* (Kähler 2020 [1940]:84). Pre-E *d is also reconstructed for the *d/r-t* correspondences that were identified by Kähler (1987) as a dialectal variant of Western Enggano that was supposedly caused by sporadic devoicing. This stands in contrast to Edwards' (2015) reconstruction to *t, which was made on the basis of the prevalence of *t* in words not marked as dialect variants. A reconstruction to *d is made here primarily because many instances not marked as dialectical are clear Malay adaptations (e.g., *e-tiri* < Malay *sirih* 'betel') and many of the forms containing *t* can be traced back to PMP *d (e.g., *doho-i-toho-i* < PMP *dəŋəR 'to hear'), supporting Kähler's hypothesis of sporadic devoicing. In agreement with Edwards (2015), the correspondence $\widehat{d}_3\text{-}\widehat{t}$ is reconstructed to Pre-E *d on the grounds that the conditioning environment (only occurring before /i/) provides reasonable evidence to assume an earlier, nonpalatal source. This conditioning points to earlier *t/d* and *d being reconstructed for the same reasons given above. For the correspondence *d/r-l*, where *d/r* indicates the already discussed free variation, Pre-E *l is reconstructed on the basis that there is no apparent conditioning environment for a supposed *d > l change and given that *d/r-d/r* correspondences already attest to a *d reconstruction.

TABLE 10. PRE-E NONTRIVIAL RECONSTRUCTIONS AND CORRESPONDENCE SETS.

Pre-E	Env.	South	West
*p		f	p
*d		d~r	d~r
*d		d~r	t
*d	_i	\widehat{d}_3	\widehat{t}
*l		d~r	l

2.7. PRE-N. Nasal is a language spoken in three villages in the southernmost part of Bengkulu province, close to the border with Lampung. It was spoken in three villages at the mouth of the Nasal River until a flood in the 1980s forced one of the villages to relocate further from the other two. There do not appear to be any dialectal differences in Nasal spoken in these three villages.

Nasal was not known to linguists until 2007 (Anderbeck and Aprilani 2013), although a wordlist was collected by an unknown Dutch colonial officer in 1895 (Stokhof 1987), and a few Nasal words appear in Helfrich (1891). The analysis that is presented in this section primarily relies upon the documentation that is currently being carried out by a team of researchers that includes the authors (see McDonnell 2017a,b, 2019, forthcoming), but reference is also made to the previous sources.

The Pre-N reconstructed phonological inventory is provided in table 11. The voiced obstruents *b, *d, and *g do not occur in final position, nor does *y

occur in initial position. Pre-N *o (from PMP *a) is only reconstructed word-finally, as is *ʔ, where it contrasts with *k. Though all PMP voiced obstruents have been considerably reduced in Nasal in native vocabulary, they seem to have been well-retained adjacent to *ə (e.g., *rəbuŋ > ləbuŋ, *pəzəm > pədum, and *hapəjəs > pəgus) and in nasal-stop sequences. Elsewhere, PMP *b became *w, *d became *χ, and *g became *h, except word-finally, where all were devoiced. The majority of other instances of *b*, *d*, and *g* in modern Nasal can be attributed to borrowing.

TABLE 11. PRE-N RECONSTRUCTED PHONOLOGICAL INVENTORY.

		Bil	Alv	Pal	Vel	Uvu	Glo				
								Front	Central	Back	
Stop	vd.	b	d		g			High	i	u	
	vl.	p	t		k		ʔ		Mid	ə	o
Fricative		s				χ	h			Low	a
Nasal		m	n		ŋ			Diphthongs			
Lateral			l					High		uy	
Approximant		w		y					Mid		
										Low	ay, aw

3. RECONSTRUCTING A PSUM PHONOLOGY. In this section, a proposed PSUM phonology is reconstructed based on the sound correspondences of the seven pre- and proto-languages discussed in section 2. The proposed reconstructed inventory is shown in table 12. In addition to the consonants, vowels, and diphthongs, reconstructions are also made for heterosyllabic consonant clusters (of the form nasal-stop and *rC) and vowel sequences. The evidence justifying PSUM as a valid intermediary between PMP and those pre- and proto-languages is discussed in section 4.

TABLE 12. PSUM PHONOLOGICAL INVENTORY.

		Bil.	Alv.	Pal.	Vel.	Glott.					
								Front	Central	Back	
Stop	vd.	b	d		g		High	i		u	
	vl.	p	t		k			Mid	ə		
Fricative		s				h			Low	a	
Nasal		m	n		ŋ		Diphthongs				
Trill			r				High	iw		uy	
Lateral			l					Mid			
Approximant		w		y					Low	ay, aw	

A few remarks on organization and notation must be made. First, in listing the sound correspondences, the most restrictive conditioning environments are stated first, followed by those that include all the remaining reconstructions. For example, in section 3.1.1, PSUM *b is reconstructed from the set *w-b-p-b-b-p-b* in intervocalic position where the first vowel is a schwa. For the second correspondence set, *w-b-b-b-p-w*, word-initial and intervocalic positions are considered, but this intervocalic environment excludes what was already listed, that

is, is restricted to V_V where the first vowel is not schwa. Second, where clear cognates have not yet been identified, a box has been filled with gray to show that its reflex is uncertain; see, for example, the correspondence of PSUM *b word-finally in Pre-N in table 13. Finally, for Pre-E, where nasality is a word-level suprasegmental feature, if a set regularly corresponds with a nasal word in Pre-E, it is preceded by a tilde (~). This is particularly important for the correspondences regarding liquids and nasals.

3.1. CONSONANTS.

3.1.1. *b. PSUM *b is reconstructed from the three regular correspondence sets outlined in table 13. The first two are nearly equivalent, though PNBI *b was devoiced and Pre-N *b was retained following *ə. Otherwise, these two sets are rather straightforward, with Pre-G and Pre-N having lenited *b to *w and Pre-E having merged earlier *b and *p. The final set for *b finds its analogs in *d and *g, revealing a common word-final devoicing in nearly all languages that retained word-final consonants; the only retention of original voiced stops is found in PB.

TABLE 13. CORRESPONDENCE SETS FOR PSUM *b.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*b	ə_V	w	b	p	b	b	p	b	68, 80, 133, 158, 159
	#_ V_V	w	b	b	b	b	p	w	9, 15, 21, 23, 27, etc.
	_#	p	b	p	∅	p	∅		54, 194

3.1.2. *d. As seen in table 14, PSUM *d is reconstructed from a variety of correspondence sets. All of these reflexes can be accounted for by common paths of sound change (e.g., lenition, $d > r > \chi$). Of the changes included in such a reconstruction, the most striking is the lateral assimilation in PCBI and Pre-M whereby the initial *d preceding *l in the following syllable is assimilated to *l.

TABLE 14. CORRESPONDENCE SETS FOR PSUM *d.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*d	#_Vl	r	d	d	l	l		χ	35, 36, 39, 42, 181
	#_ r	d	d	d	d	d	d	χ	32, 37, 38, 43, 44, etc.
	V_V	r	d	r	r	d	d	χ	51, 95, 102, 109, 131, etc.
	_#	t	d	r	∅		∅	t	72, 114, 148, 168, 170, etc.

3.1.3. *g. PSUM *g is reconstructed for the correspondence sets in table 15, with the proposed changes of word-final devoicing (as with *b and *d) in many of the languages, shift to PNBI *k, lenition to Pre-E and Pre-N *h, and inter-vocalic fricativization to PCBI *x. From PSUM to Pre-M, final *g formed a diphthong with a preceding *a or *ə; however, cases of original *ig and *ug remain unattested. This change is also found with Pre-M final *k, again providing evidence for the Pre-M final stop devoicing.

TABLE 15. CORRESPONDENCE SETS FOR PSUM *g.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*g	#_	∅	g	k	g	g	h	g	47, 48, 186
	V_V	∅	g	k	x	g	h	h	57, 82, 96, 103, 111, etc.
	_#	k	g	k	∅	w	∅	k	36, 39, 55, 106, 127, etc.

3.1.4. *p. PSUM *p is reconstructed from the three correspondence sets in table 16 based on common paths of lenition to PNBI *h and PCBI *f. The first set finds its analog in section 3.1.1—whereas PSUM *p was normally lenited to PNBI *h intervocally, it was retained as PSUM *p if preceded by a PSUM *ə. As with PSUM *t, voiceless stops were sporadically debuccalized in Pre-N, although this may be a more recent change resulting from contact with Lampungic.

TABLE 16. CORRESPONDENCE SETS FOR PSUM *p.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*p	ə_V	p	p	p	f	p	p	p	41, 92
	#_ V_V	p	p	h	f	p	p	p	46, 61, 103, 104, 105
	_#	p	p	p	∅	p	∅	p~?	33, 109, 125, 126, 166, etc.

3.1.5. *t. Apart from those exhibited by Pre-E, the correspondences exhibited in table 17 are fairly typical, the only change being the crosslinguistically common change of spirantization of *t before the high front vowel *i, resulting in PCBI *s.

TABLE 17. CORRESPONDENCE SETS FOR PSUM *t.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*t	_i	t	t	t	s	t	k	t	4, 149, 163, 164, 165, etc.
	#_ V_V	t	t	t	t	t	k	t	69, 121, 152, 154, 161, etc.
	_#	t	t	t	∅	t	∅	t~?	46, 66, 70, 74, 121, etc.

3.1.6. *k. The majority of the correspondences in the four sets in table 18 involve /k/ or /ʔ/ in the daughter languages. Because debuccalization of *k is a well-known path of change, these are reconstructed to *k. The loss of initial *k for PNBI and PCBI is not unexpected, assuming the change *k > *ʔ happened first since the distinction between vowel-initial and glottal-stop-initial words is often leveled. PNBI exhibits another interesting sound change—the loss of *k (or possibly *ʔ after the change *k > *ʔ) preceding *i. As with *g, PSUM *k is reflected as Pre-M *w in the final position.

TABLE 18. CORRESPONDENCE SETS FOR PSUM *k.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*k	_i	k	k	∅	ʔ	k	ʔ	k	16, 34, 73, 138
	#_	k	k	∅	∅	k	ʔ	k	64, 65, 66, 70, 71
	V_V	k	k	ʔ	ʔ	k	ʔ	k	1, 24, 33, 58, 63, etc.
	_#	k	k	ʔ	∅	w	∅	ʔ	2, 16, 23, 85, 89, etc.

3.1.7. *s. As seen in table 19, the correspondence sets for *s are rather straightforward, the only nontrivial change being the loss of *s word-finally in Pre-M and its merger with *t as *k in Pre-E.

TABLE 19. CORRESPONDENCE SETS FOR PSUM *s.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*s	#, V_V	s	s	s	s	s	k	s	52, 106, 145, 146, 147, etc.
	_#	s	s	s	∅	∅	∅	s	3, 4, 132, 151, 183

3.1.8. *h. Reflexes of PSUM *h are relatively scarce, being best attested by Pre-N but also found word-finally in Pre-G, finally and medially between like vowels in PB, and medially and initially in PCBI. These regular correspondences are outlined in table 20.

TABLE 20. CORRESPONDENCE SETS FOR PSUM *h.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*h	#, V_V	∅	h	∅	h	∅	∅	h	29, 121, 123, 125, 170, etc.
	_#	h	h	∅	∅	∅	∅	h	9, 11, 12, 22, 37, etc.

3.1.9. *m, *n, and *ŋ. As can be seen in table 21, the three nasals reconstructible to PSUM—*m, *n, and *ŋ—are stable in all of the Sumatran languages except Pre-E. In Pre-E, *m and *n correspond regularly with *b and *d. As discussed in Edwards (2015) and Smith (2020), this is a result of Pre-E’s suprasegmentalization of nasality, in which nasality transitioned from a segment-level to a word-level feature. Such a claim is supported by the fact that PSUM coda nasals (except when preceded by schwa) led to word-level nasality in Pre-E. For Pre-E *h as a reflex of PSUM *ŋ, we find an analog in the correspondences that justified the reconstruction of PSUM *g. Just as PSUM *g is reflected as Pre-E *h, so is PSUM *ŋ, indicating that the two underwent a merger first as a result of the suprasegmentalization of nasality and were later reduced to *h via a common path of lenition (as was proposed by Smith 2017).

TABLE 21. CORRESPONDENCE SETS FOR PSUM NASALS *m, *n, AND *ŋ.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*m	#, V_V	m	m	m	m	m	b	m	64, 83, 85, 86, 87, etc.
	ə_#	m	m	m	∅	m	∅	m	35, 45, 102, 141, 164
	_#	m	m	m	∅	m	~∅	m	60,88
*n	#, V_V	n	n	n	n	n	d	n	2, 45, 89, 90, 91, etc.
	ə_#	n	n	n	∅	n	∅	n	61, 160
	_#	n	n	n	∅	n	~∅	n	32, 51, 112, 124, 131, etc.
*ŋ	#, V_V	ŋ	ŋ	ŋ	ŋ	ŋ	h	ŋ	40, 74, 75, 94, 95, etc.
	ə_#	ŋ	ŋ	ŋ	∅	ŋ	∅	ŋ	47
	_#	ŋ	ŋ	ŋ	∅	ŋ	~∅	ŋ	27, 52, 57, 97, 113, etc.

3.1.10. *r. The correspondence sets for *r in table 22 show a variety of reflexes—Pre-G *y and Ø, PB *r, PNBI and Pre-N *l, and PCBI, Pre-M, and Pre-E Ø. For PCBI, Pre-M, and Pre-E, this *r appears to have caused changes to adjacent vowels (discussed in section 3.3) but has left little trace for its reconstruction otherwise. Of those few seemingly irregular forms where *r did not disappear entirely in PCBI and Pre-M, we see such reflexes as PCBI *h (PSUM *tirəm > PCBI *siho expected **sio ‘k.o.crustacean’) and Pre-M *ŋ (PSUM *urat > Pre-M *uŋat expected **uat ‘tendon’; PSUM *hənur > Pre-M *enuŋ expected **enu ‘path’), possibly indicating a lenition path through some medial glottal/velar consonant. The reconstruction of these sets to *r is made on a similar basis as that described by Blust (2013:595) for the reconstruction of Proto-Austronesian (PAN) *R.

TABLE 22. CORRESPONDENCE SETS FOR PSUM *r.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*r	#_	Ø	r	l	Ø	Ø	Ø	l	51, 132, 133, 134, 135, etc.
	i_#	Ø	r	l	Ø	Ø	~Ø [†]	1	19, 21, 107, 122, 144, etc.
	V_V, _#	y	r	l	Ø	Ø	Ø	1	10, 11, 37, 40, 187

[†] While the conditioning environment appears strange, the only reflexes of PMP final *r have been found following *i, and all such reflexes triggered nasality. On the other hand, only one PMP final *R has been found to trigger nasality, the only identified reflex of PMP *iR#—PMP *qabata/iR ‘sago grub’ > Pre-E *pākī ‘wood worm’. Reflexes of PMP final *r following *a and *u must be identified in order to verify if this is truly a merger.

3.1.11. *l. As is evident from the correspondence sets in table 23, PSUM *l is the most stable phoneme of all those reconstructed, being retained in all daughter languages initially and medially. In final position, *l is lost, as expected, in PCBI and Pre-E. It is likely that this final *l first became a final nasal *n in Pre-E before disappearing since, as proposed by Smith (2020), it served as a trigger for nasality (except after *ə, as discussed in section 3.1.10).

TABLE 23. CORRESPONDENCE SETS FOR PSUM *l.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*l	#, V_V	l	l	l	l	l	l	l	74, 76, 83, 112, 128, etc.
	ə_#	l	l	l	Ø	l	Ø	l	68
	_#	l	l	l	Ø	l	~Ø	l	65, 67, 186

3.1.12. *w and *y. There are two approximants reconstructible to PSUM: *w, found initially and medially, and *y, only found medially. The correspondence sets for both are shown in table 24 and are rather straightforward, with the reduction of PSUM *wa to PNBI *o in initial position (see section 3.3.1), fortition of PSUM *w to Pre-M, Pre-E *b, and fortition of PSUM *y to Pre-M *j.

TABLE 24. CORRESPONDENCE SETS FOR PSUM APPROXIMANTS *w AND *y.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*w	#_a	w	w	∅	w	b	b	w	175, 176, 177, 178, 179, etc.
*w		w	w	w	w	b	b	w	67, 77, 113, 140, 145, etc.
*y		y	y	y	y	j		y	29, 62, 78, 79

3.2. CONSONANT CLUSTERS. Two types of heterosyllabic consonant clusters, both shown in table 25, are clearly reconstructible for PSUM—homorganic nasal-stop sequences and *rC sequences, though not many such terms are reconstructible to PSUM. For the nasal-stop sequences, all languages underwent common paths of change, with Nʃ developing into Pre-G prenasalized stops, Nʃ developing into PCBI voiced stops, and *ŋg becoming PNBI *ŋk (via regular *g > *k) and PCBI *g. These nasal-stop sequences also triggered nasality in Pre-E. The *rC sequences are retained only in PB and Pre-G (where they underwent the normal *r > *y change) and have lost all evidence of the original *r in PNBI, PCBI, and Pre-N. No tautosyllabic consonant clusters are reconstructible for PSUM.

TABLE 25. CORRESPONDENCE SETS FOR PSUM CONSONANT CLUSTERS.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*mb	V_V	mb	mb	mb	mb	mb	~p	mb	113, 197
*nd		nd	nd	nd	nd	nd		nd	163
*ŋg		ŋg	ŋg	ŋk	g	ŋg		ŋg	49, 81, 97
*mp		mp	mp	mp	b	mp	~p	mp	173
*nt		nt	nt	nt	d	nt	~k	nt	119
*ŋk		ŋk	ŋk	ŋk	g	ŋk	~ʔ	ŋk	16, 136, 189, 193
*rC		yC	rC	C	C			C	17, 116, 185

3.3. VOWELS. Four vowels are reconstructed for PSUM, showing clear continuity from the inherited PMP vowel system. Each of these reconstructions is discussed in turn below.

3.3.1. *a. Many of the changes regarding PSUM vowels were triggered by an adjacent *r in PCBI, Pre-M, and Pre-E. The changes seen in PCBI follow two primary rules. First, the vowels only underwent centralization from PSUM *a, *ə, and *u to PCBI *o and from PSUM *i to PCBI *e. Second, the vowel that changed was: (i) the vowel adjacent to *r in word-initial or word-final position; (ii) the leftmost vowel if *r was between identical vowels; or (iii) the vowel lowest on the hierarchy *i > *u > *a > *ə if *r was between two dissimilar vowels. The only exception to this rule is the PSUM sequence *ara, which became PCBI *o (rather than *oa). Pre-M and Pre-G have similarly reduced PSUM *ara to *o and *ɛ (< *aya), respectively. Pre-M appears to have undergone many, but not all, of the same changes as PCBI. In Pre-G and Pre-N, word-final *a became *ə and *o, respectively. All of these correspondences for PSUM *a are shown in table 26.

TABLE 26. CORRESPONDENCE SETS FOR PSUM *a.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*a	#w_	a	a	o	a	a	a	a	175, 176, 177, 178, 179, etc.
	_r#	e	a	a	o	a	a	a	76, 78, 143, 187
	#r_		a	a	o	o		a	132
	ra, ar	ε	a	a	o	o		a	11, 37, 66, 98
	_ri		a	a	o	o		a	100, 180
	_ru		a	a	o			a	14, 184, 194
	ir_		a	a	o	a	a	a	20, 117
	ur_	a	a	a	o	a	a	a	171, 174
	_#	ə	a	a	a	a	a	o	24, 29, 69, 83, 103, etc.
	else	a	a	a	a	a	a	a	1, 2, 3, 5, 6, etc.

3.3.2. *i. As seen in the correspondence sets in table 27, the PSUM vowel *i is remarkably stable in all languages, only shifting to *e preceding a word-final *r in PCBI and Pre-M.

TABLE 27. CORRESPONDENCE SETS FOR PSUM *i.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*i	_r#	i	i	i	e	e	i	i	19, 21, 107, 122, 144, etc.
	else	i	i	i	i	i	i	i	4, 16, 17, 19, 20

3.3.3. *u. PSUM *u is similarly stable. When preceding a word-final *r, *u became *o in PCBI and Pre-E and *i in Pre-G via monophthongization after the *r > *y change (*ur# > *uy# > *i#). In addition, PSUM *u became PCBI *o following initial *r and preceding *ri. These correspondences are outlined in table 28.

TABLE 28. CORRESPONDENCE SETS FOR PSUM *u.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*u	_r#	i	u	u	o	u	o	u	18, 58, 59, 80, 81, etc.
	#r_ri	u	u	u	o	u	u	u	135, 136, 137
	else	u	u	u	u	u	u	u	1, 6, 8, 22, 23, etc.

3.3.4. *ə. PSUM *ə is reconstructed from a variety of correspondence sets where reflexes are found as *ə, *o, *u, and *e as shown in table 29. Most of these, as already discussed in the previous sections, were triggered by an adjacent *r and, as such, will not be further discussed. In the seventh correspondence set, a PSUM *ə in words ending in *əCur# became *u, though it is interesting to note that other words ending in *əCuC# (e.g., PSUM *ləsuj ‘mortar’ > PCBI *ləsu ‘mortar’) did not see a similar change in *ə. PNBI and Pre-N both lost initial *ə, *ə in closed final syllables merged with *u in Pre-N, and for the second-to-last set, Pre-G underwent a change similar to that seen for Dairi, whereby *ə became *o if preceding a final *h. The regular reflexes of *ə are Pre-M *e, Pre-E *o, and *ə in all the others.

TABLE 29. CORRESPONDENCE SETS FOR PSUM *ə.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*ə	#r_	ə	ə	ə	o	o	o	ə	51, 133
	_r#	ə	ə	ə	o	o	o	u	40, 162, 163, 196, 197
	_ru	u	ə	ə	o			ə	10, 195
	_Cu	u	ə	ə	ə	e	o	ə	92, 101, 121, 159, 161, etc.
	ar_ir_		ə	ə	o		o	ə	12, 38, 164, 179
	_[k/ŋ]		ə	ə	o	e	o	ə	40, 49, 110, 130, 162
	_Cur#		ə	ə	u	e		ə	18, 80, 81, 119, 120
	_h#	o	ə	ə	ə	e	o	u	12, 38, 49, 56, 134, etc.
	_C#	ə	ə	ə	ə	e	o	u	12, 33, 35, 38, 39, etc.
	#_	o	ə	∅	ə	e	o	∅	45, 46, 53
	else	ə	ə	ə	ə	e	o	ə	39, 41, 68, 102, 160, etc.

3.4. VOWEL SEQUENCES AND DIPHTHONGS. Based on the correspondences given in table 30, two medial heterosyllabic vowel sequences and four tautosyllabic word-final diphthongs⁶ are reconstructible to PSUM. The first two correspondences are reconstructed to vowel sequences rather than simple vowels because the reduction to PNBI *o for these sequences follows from an expected process of monophthongization. Although PSUM *ia and *ai have not yet been identified root internally, they, too, were monophthongized across morpheme boundaries to PNBI *e (see notes in appendix C; this process did not apply to the third-person singular pronoun *ia). A number of vowel sequences containing *ə (i.e., *uə, *əu) may also be reconstructible but are only marginally attested.

TABLE 30. CORRESPONDENCE SETS FOR PSUM VOWEL SEQUENCES AND DIPHTHONGS.

PSUM	Env.	Pre-G	PB	PNBI	PCBI	Pre-M	Pre-E	Pre-N	Examples (appendix G)
*au	C_C	au	au	o	əu		au	au	6, 32, 72, 182
*ua	C_C	ua	ua	o	ua	ua	ua	ua	8, 22, 43, 54, 146, etc.
*ay	_#	ay	ey	ay	e	ay	e	ay	48, 53, 82, 87, 93, etc.
*uy	_#	i	uy	oy	i	uy		uy	5, 50, 75
*aw	_#	aw	ow	aw	ə	ew		aw	30, 99, 111, 169, 191, etc.
*iw	_#		i	ew	i	iw			7, 13

3.5. SUMMARY. Based on the discussion above, the PSUM phonological inventory is reconstructed as shown in table 12 at the beginning of this section. Most consonants were allowed in initial, intervocalic, and final positions, with *y not being permitted word-initially. PSUM *a, *i, and *u were allowed in all positions, with *ə being allowed in all but open word-final syllables. Certain vowel sequences were permitted, as were word-medial consonant clusters, which were restricted to those of the form nasal-stop and *rC.

6. In continuing from Blust (1998), among others, for PAN and PMP, these are denoted “diphthongs” but may represent vowel-consonant sequences (see Clynes 1997).

4. JUSTIFYING PSUM. Now that we have reconstructed a tentative PSUM phonological system, in this section, we argue that the differences in this and the PMP reconstructions present enough evidence to propose that PSUM is a separate subgroup of PMP. By comparing the reconstructed PSUM phonological inventory with that of PMP, six shared phonological innovations—five mergers and one shift—are identified, providing strong evidence for the validity of PSUM. These changes are outlined in table 31 and discussed further below.

TABLE 31. PHONOLOGICAL INNOVATIONS FROM PMP TO PSUM.

Type of Change	PMP	Evidence	PSUM
Merger	*j	36, 39, 49, 55, 57, etc.	*g
	*g	47, 48, 81, 97	
	*d	32, 33, 34, 109, 125, etc.	*d
	*z	51, 102, 131, 163, 172, etc.	
	*R	11, 12, 18, 135, 136, etc.	*r
	*r	21, 81, 133, 144, 165	
	*n	2, 8, 92, 93, 160, etc.	*n
	*ñ	88, 89, 90, 101, 114	
	*h, Ø	6, 23, 25, 32, 43, etc.	Ø
Shift	*q	6, 9, 10, 108, 109, etc.	*h

4.1. PMP *j, *g > PSUM *g. As seen in the examples in table 32, none of the Sumatran daughter languages present evidence for a **j*/**g* distinction, both having merged as PSUM **g*. This reconstructed **g* has a variety of reflexes due to common phonological processes as lenition to Pre-G Ø, PCBI **x*, and Pre-E **h*, as well as final devoicing to **k* in Pre-G, PNBI, Pre-M (> **w*), and Pre-N. In all of these cases, however, only PSUM **g* is reconstructible. The phoneme **g* is vanishingly rare in PMP basic vocabulary, but in languages that demonstrate apparent reflexes of PMP **g*, there is a clear merger with **j*.

TABLE 32. SUMATRAN REFLEXES OF PMP *g, *j > PSUM *g.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Pre-M	Pre-E	Nasal	Gloss
<i>*pusəj</i>	<i>*pusəg</i>	pusok	pusəg		fusə	pusow	puko	pusuk	‘navel’
<i>*pajay</i>	<i>*paɣay</i>		pagey	akay	faxe			pəhay	‘rice’
<i>*sijəm</i>	<i>*sigəm</i>		sigəm	sikəm		sigəm	kiho	sihum	‘ant’
<i>*ijun</i>	<i>*iguŋ</i>	iuŋ	iguŋ	ikuŋ	ixu				‘nose’
<i>*gələŋ</i>	<i>*gələŋ</i>		gələŋ	lakələŋ		geleŋ	e-hodo	gələŋ	worm

4.2. PMP *z, *d > PSUM *d. Reflexes of PMP **z* and **d* point to a merger in all daughter languages, as can be seen in table 33. As will be discussed in section 5.1, there are, however, a large number of cases where Gayo, Batak languages, NBI languages, or Nasal appear to have a voiced palatal affricate without corroborating evidence from other Sumatran languages. It is clear that these are Malay loans. The only apparent possible reflexes of **z* in Enggano are *kuru* ‘cape’ < PMP **tuzuq* ‘to point’ and *~dada* ‘breakers, surf’ < PMP **zalan* ‘road’, although these semantic connections are shaky at best.

TABLE 33. SUMATRAN REFLEXES OF PMP *z, *d > PSUM *d.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Pre-M	Pre-E	Nasal	Gloss
*haRəzan	*rədan		rədan		ora	oran			'ladder'
*pəzəm	*pədəm		pədəm		mərə	merem		pədum	'sleep'
*tuzuq	*tuduh	tə-turuh	tuduh	ituturu	turu	turu	(kuru ?)		'point'
*zalan	*dalan	ralan	dalan	dalan	lala	lalan	(~dada ?)		'road'
*tazuk	*taduk			taru	taru			taxu?	'plant'
*daRaQ	*darah	rayoh	darəh	rala	do			çalah	'blood'
*duRi	*duri	rui			doi	rui	e-du-dui	çuli	'thorn'
*ɲuda	*ɲuda		ɲuda	ɲura			e-huda	ɲuxo	'young'
*qudip	*hudip	urip			uri	uri		huçi?	'live'
*duha	*dua	roa	dua	do	dua	dua	dua	çuo	'two'

4.3. PMP *R, *r > PSUM *r. Though the presence of two rhotics *R and *r in PAN is a matter of debate (Adelaar 2005:360), the present analysis assumes their existence. Only one rhotic is reconstructible for PSUM, *r, which has derived from both PMP *R and *r. Though identifying reflexes of PMP *r in the daughter languages indeed proves difficult (as it is all but absent from basic vocabulary), the above sections have shown that the few clear cases indicate a merger of the two earlier PMP phonemes. Examples of this merger are shown in table 34.

TABLE 34. SUMATRAN REFLEXES OF PMP *R, *r > PSUM *r.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Pre-M	Pre-E	Nasal	Gloss
*birbir	*birbir	iwih	bibir		bebe	bebe			'edge'
*ləŋgur	*ləŋgur		ləŋgur	ləŋkul		le-ləŋgu			'thunder'
*rəbuŋ	*rəbuŋ		rəbuŋ	ləpuŋ	ombu	ombuŋ	(~opu ?)	ləbuŋ	'bamboo shoot'
*sisir	*sisir		sisir		sese		~kiki		'comb'
*tirtir	*tirtir		tirtir				~kiki		'shiver'
*wahiR	*wair	wəih		oil	we		be	wail	'water'
*iluR	*ilur	ilih		ilul	ilo	elu	edo		'saliva'
*Rumaq	*rumah	umah	rumah	luma	omo	uma	uba		'house'
*paRi	*pari		pari	hali	foi	poi			'rayfish'
*qasiRa	*asira		sira	asila	asio			silu	'salt'

4.4. PMP *ñ, *n > PSUM *n. Reflexes of PMP *ñ, like *r, are difficult to identify in the Sumatran languages due to the phoneme's relative rarity. However, as shown in table 35, all clear cases reveal a merger of PMP *ñ and *n as PSUM *n.

No clear evidence for or against this merger is found in Nasal. The one putative case of a PMP *ñ inheritance that is not attributable to a loan is PMP *baqəñan 'to sneeze' > Nasal *wahun*. Though this word is found all across Sumatran both with and without the final syllable (e.g., Gayo *waun~waudnən*, Simeulue *fanan*, Nias *bahŋ~bahŋna*), it is problematic for two reasons. First, this reconstruction has a number of doublets, namely PMP *bañan and *bəñan alongside Proto-Philippines *bahén (Blust and Trussel, ongoing), the last of which could very well be shared with Sumatran (assuming an earlier

TABLE 35. SUMATRAN REFLEXES OF PMP *ñ, *n > PSUM *n.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Pre-M	Pre-E	Nasal	Gloss
*ñam-ñam	*nanam		nanam	nanam		nanam			'taste'
*ñamuk	*namuk	namuk	namuk	namuk					'fly'
*ñamuR	*namur	nami	namur		namo				'dew'
*pəñu	*pənu		pənu	əno	fənu	penuŋ	~e-punu-oʔoi		'turtle'
*qañud	*[h]anud	anut	anud						'wash away'
*nanaq	*nanah		nanah	nana	nana			nana	'pus'
*duniq	*dunih					luni	dudi	χə-χunih	'rainbow'
*ənəm	*ənəm	onom	ənəm	nəm	ənə	enem		num	'six'
*anak	*anak	anak	anak	anaʔ	ono		ada	anaʔ	'child'
*manuk	*manuk	manuk	manuk	manoʔ	manu	manu		manuʔ	'bird'

PMP *baqən) with a suffixed *-ən or *-an to account for the trisyllables in the daughter languages. Second, should *baqəñan be the best reconstruction, it would assume deletion of the final *-an in Nasal. This would leave the palatal nasal *ñ in word-final position, which itself is a common environment for a merger with *n, rather than necessarily indicating an earlier *ñ, *n merger. Given that Nasal provides neither strong evidence for nor against this merger, it is only assumed to have been shared by the whole Sumatran subgroup. With regards to Enggano, though the single reflex of PMP *ñ would otherwise seem dubious due to there being no apparent trigger for nasality, Mentawai shares a similar irregularity (an appended -ŋ), which would have triggered the word-level nasality in Enggano.

4.5. PMP *h, Ø > PSUM Ø. As seen in the examples in table 36, the PMP phoneme *h has been unambiguously lost in all Sumatran languages. In many cases, this loss later resulted in monophthongization (e.g., PMP *ahu > PSUM *au > PNBI *o) or erosion of word-initial *ə (e.g., PMP *#hə > PSUM *#ə > PNBI Ø). A PSUM phoneme *h is reconstructible, though all cited cases of such clearly derive from earlier PMP *q, as discussed below.

TABLE 36. SUMATRAN REFLEXES OF PMP *h, Ø > PSUM Ø.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Pre-M	Pre-E	Nasal	Gloss
*hulij	*ulij	ulik	ulij						'hug'
*hasaŋ	*asaŋ			asaŋ		asaŋ	~aka		'cheek'
*hapuy	*apuy		apuy	ahuy				apuy	'fire'
*duha	*dua	roa	dua	do	dua	dua	dua	χuo	'two'
*haRəzan	*rədan		rədan		ora	oran			'ladder'

4.6. PMP *q > PSUM *h. The PSUM *h has merged with Ø in the majority of the languages here, but on the basis of /h/ reflexes in Gayo, the Batak languages, Nias, and best preserved in Nasal, it is assumed that PMP *q shifted to *h in PSUM and then later merged with Ø in the respective daughter languages. Evidence of this shift can be seen in the examples in table 37.

TABLE 37. SUMATRAN REFLEXES OF PMP *q > PSUM *h.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Pre-M	Pre-E	Nasal	Gloss
*qulu	*hulu	ulu	ulu	ulu	ulu	ulu	ulu	ulu	'head'
*quləj	*huləg		uləg	ulag	ulə	ulow		huluk	'worm'
*Rumaq	*rumah	umah	rumah	luma	omo	uma	uba		'house'
*sulūq	*suluh	suluh	suluh	sulu	sulu	sulu	kudu	suluh	'torch'
*baqəRu	*bahəru	ayu	baru	balu	bohəu	bau			'new'
*taqi	*tahi	tai	tai	tai	tai		kai	tahi	'feces'
*tuqəd	*tuhəd				tuhe		kuo	tuhut	'knee'

4.7. STRENGTH OF EVIDENCE. In total, there are six phonological innovations that support the Sumatran subgroup. Of the languages discussed here, all clearly exhibit these six changes, apart from the absence of evidence for/against the nasal merger in Nasal and for the three questionably attested sound changes in Enggano.

The strongest evidence for Sumatran comes from the merger of PMP *j and *g to PSUM *g, which, outside of the Cordilleran languages of the Northern Philippines (Reid 1974), is exhibited by few other Austronesian languages (see Smith 2017, for a similar comment). As is the case there, the fact that these languages all exhibit velar (or velar-derived, e.g., *g > Enggano /h/) reflexes of *j in an area where a *j-*d merger dominates (as occurred in Proto-Malayo-Chamic) attests to the close relationship of these languages.

Although similarly unencountered in Proto-Malayo-Chamic, the merger of the PMP palatals *z and *ñ with their alveolar counterparts *d and *n, respectively, is widely encountered elsewhere, as with both in many languages of the Philippines (Blust 1991:88) and Proto-Tanimbar-Kei (Mills 1991) and the merger of the palatal and alveolar nasals in Proto-Bungku-Tolaki (Mead 1998).

The last three changes are even more common, and many other subgroups have similarly innovated them, as with all three in Proto-Malayo-Chamic (Adelaar 2005), the merger of *R and *r in the Seko languages (Laskowske 2006), the shift of *q > *h in Proto-Kimbe (Goodenough 1997), and the loss of *h in Proto-Kayanic (Smith 2017).

In light of all of this, the strength of the argument for PSUM does not lie in individual sound changes (apart from, perhaps, PMP *j, *g > PSUM *g) but rather from the combination of all six shared by Sumatran languages.

5. FURTHER ISSUES IN THE SUMATRAN SUBGROUP. In this section, we address three issues of significance: section 5.1 responds to the issues that were raised from the reconstructed phonological inventories of section 2, section 5.2 discusses further proposals for internal subgrouping, and section 5.3 briefly discusses the place of the other non-Malayo-Chamic languages of Sumatra, namely Rejang and Lampungic.

5.1. ISSUES DURING RECONSTRUCTION.

5.1.1. Inheritances and loans. The identification of clear inheritance versus loan vocabulary is a difficult matter in many Sumatran languages. This is especially true because the source of the loans is other Malayo-Polynesian languages. This issue is particularly pronounced in the languages of mainland Sumatra (Gayo, Batak, and Nasal) as well as NBI. It may come as no surprise that the source of much of the borrowing is from Malayic languages. Gayo, Batak, and Nasal have likely been in contact since the spread of Malays in Sumatra over the past two millennia, and NBI shows evidence of contact with Malayic speakers, primarily Minangkabau, who have settled on Pulau Simeulue and Pulau Banyak. There are also loans from non-Malayic languages. In northern Sumatra, Gayo and NBI have a number of apparent loans from Acehnese, and in southern Sumatra, Nasal has a large number of apparent loans from Lampungic. In the cases of Gayo and Nasal, the borrowing is extensive. Shorto (1976) writes that “[i]n this core vocabulary, the proportion of inherited items [is] 56 per cent [and] Gayo may well prove to be an extreme case of linguistic acculturation” (pp. 206–07). Anderbeck and Aprilani (2013) also estimate a high number of loans from both Lampungic and Malay in Nasal.

In order to identify what is inherited and what is loan vocabulary, we had to make a decision between two subsets of the vocabulary—the larger set with phonological derivations from PMP that match the histories of surrounding languages (here, Malayic, Acehnese, and Lampungic) and the smaller set with phonological derivations from PMP that are unique to the region. Though the latter is often smaller and less well attested, we take it as the core inherited vocabulary since it cannot be accounted for by contact, whereas the prior can. In doing so for Gayo and Nasal, for example, we overlook the apparent /r/ reflexes of PMP *j since these can be attributed to Acehnese and Lampungic, respectively. Instead, we take *g, *k, and Ø as the true reflexes for Gayo, and *g, *k, and *h for Nasal. In Gayo, the same can be said for reflexes of *b, *d, and *g, where all of these appear to have historically shifted elsewhere (*w/p, *r/t, and *g/k/Ø, respectively). It is for this reason that the voiced stops *b, *d, and *g are not reconstructed here for Pre-G. One significant borrowing in all four of these languages/language groups is the voiced palatal affricate /dʒ/, often found in loans with a reflex of PMP *z. Despite this fact, section 4.2 demonstrates a merger of PMP *z and *d in all of these languages.

In the same vein, it is worth mentioning that the long-standing debate surrounding PMP *c is not considered here. As opposed to the other phonemes considered, there are no clear cases of inherited PMP *c attested by multiple Sumatran languages and not attributable to loans. By far the majority of claimed PMP *c inheritances to be found in the Sumatran group are encountered in the Batak languages, although this is perhaps unsurprising given that the Batak languages have been used as one of the foundations for reconstructing PMP *c (Blust 2013:570). What is important, however, is that those lexical

items that contain *c have no apparent cognates in the other Sumatran languages. Those that have been cited in other sources can be easily attributed to loans due to irregular sound correspondences, as with those in table 38.

TABLE 38. APPARENT *c REFLEXES IN SUMATRAN LANGUAGES OUTSIDE BATAK.

PMP	Loan language		Sumatran attestation	Irregularity
*calapa? ‘betel case’	Malay	celapa	Toba salapa Simeulue salapa Nias salafa Mentawai salipa Nasal sələpo	*p > /p/ *a > /i/ *a > /ə/
*cəkək ‘house lizard’	Acehnese	cicak	Gayo cicak Karo cikcak	*ə > /i/, /a/ *ə > /i/, /a/
	Minangkabau	cacak	Haloban cacak Nias sasa	*ə > /a/ *ə > /a/
*cabaŋ ‘branch’	Malay	cabaŋ	Karo cabaŋ Simeulue sabaja	*b > /b/
*pacək ‘wooden nail’	Malay	pacak	Angkola pasak Nias fasa	*ə > /a/ *ə > /a/

5.1.2. Enggano’s aberrancy. Enggano certainly stands out as a language that is highly distinct phonologically and lexically from what are proposed here to be its closest relatives within the Austronesian language family. However, all of the consonant sound changes apart from two that it has undergone since PSUM are rather common crosslinguistically. Those discussed here have all already been laid out by Edwards (2015), with some elaboration by Smith (2017, 2020), though they are repeated here to further justify Enggano’s membership in Sumatran. Enggano’s consonant sound changes since PSUM are visually summarized in table 39 and discussed more thoroughly below.

TABLE 39. CONSONANT SOUND CHANGES FROM PSUM TO ENGGANO.

	PSUM		Stage 1		Stage 2		Pre-E
#1	*h	→	∅				
	*r	→	∅				
	*p	↘	*p	→	*p	→	*p
	*b	↗					
	*k	→	*ʔ	→	*ʔ	→	*ʔ
#1/2	*t	↘	*t	→	*k	→	*k
	*s	↗					
#2	*m	→	*m	↘	*b	→	*b
	*w	→	*w	↗			
	*n	→	*n	↘	*d	→	*d
	*d	→	*d	↗			
#2/3	*ŋ	→	*ŋ	↘	*g	→	*h
	*g	→	*g	↗			
	*l	→	*l	→	*l	→	*l
?	*y	?					

Given the large number of changes Enggano has undergone, three distinct stages are needed to account for some of the large chains. While this does not necessarily reflect the order of the changes (e.g., the loss of *h need not have happened before the shift of *t to *k), it is useful for understanding the larger picture.

5.1.2.1. Stage 1. The first stage of Enggano sound changes consists of the following: (i) merger of *b and *p as *p, (ii) merger of *s and *t (possibly also *c?) as *t, (iii) loss of *h, (iv) changes involving *r, and (v) shift of *k to *ʔ.

*Merger of *b and *p as *p.* This change is well attested crosslinguistically, though it is rather rare in Austronesian, having happened most notably in the Oceanic languages (Blust 2013:737) as well as in Makuva (van Engelenhoven 2009).

*Merger of *s and *t as *t.* This is the first of two rather unique sound changes. As was proposed would be the case by Edwards (2015:65), the merger of *s and *t as *t is evidenced as happening before a shift of *t to *k by the earliest sources we have of Enggano (e.g., PMP *sa-ŋa-puluq ‘ten’ > Pre-E ta-ha-pulu from Rosenberg 1855a).

*Loss of *h.* A very common sound change that warrants no further discussion.

*Changes with *r.* The phoneme *r underwent a number of changes in the first stage that participated in or wholly resulted in the apparent disappearance of the phoneme from the phonological inventory. First, it appears that final *r⁷ (along with *l) in certain lexical items later triggered word-level nasality (Stage 2), likely via an intermediate merger with the alveolar nasal. That this merger has occurred should not be considered strange, since, as Smith (2020) says, “[t]he nasalization of nonnasal sonorants in coda position is common throughout the Austronesian family” (p. 354). Second, as with the CBI languages, final *r furthermore had an effect on preceding *u, that is, PSUM *ur# > Enggano *o#.

*Shift of *k to *ʔ.* This change, too, is very widely attested and is also exhibited by PNBI and PCBI.

5.1.2.2. Stage 2. The changes that took place in the second stage are as follows: (i) shift of *w to *b, (ii) shift of *t to *k, and (iii) suprasegmentalization of nasality.

*Shift of *w to *b.* Although Edwards (2015:66) considers this fortition “unusual,” it is rather well attested crosslinguistically and, as he points out, is found elsewhere in island Southeast Asia, such as with Mentawai. Whether this should properly be considered a shift or a merger is unclear, as it seems impossible to ascertain whether this or the suprasegmentalization of nasality change happened first. Were it the former, this is a shift, the latter, a merger, though whether it was a shift or merger need not matter much, as the change, regardless, is rather common crosslinguistically.

7. Whether this change acted on final *r or coda *r more generally is unclear.

*Shift of *t to *k.* While this would otherwise be considered a relatively strange sound change, the conclusion of Blust's (2004) extensive article singling out this sound change in the Austronesian languages should be heeded: "Despite initial appearances, then, the foregoing observations suggest that *t > k is a natural sound change in consonant inventories that have first lost *k" (p. 394). Given that Enggano necessarily first lost *k via the shift *k > *ʔ (otherwise we would expect *s and *t to also be reflected as *ʔ), it should then not be considered too far-fetched that the shift *t > *k would occur in Enggano's history.

Suprasegmentalization of Nasality. This change is unique to Enggano and is not attested anywhere else in the Austronesian family. There can be no argument that diminishes the rarity of such a sound change. As a result of this change, *m > *b, *n > *d, and *ŋ > *g. While the change *ŋ > *g is not directly evidenced, it makes the most sense given that the suprasegmentalization change happened to Enggano system-wide and since the reflexes of *ŋ and *g are the same—*h*. See Smith (2017) for an eloquent explanation of this. As previously discussed, this change resulted in nasality becoming a word-level feature in words that originally contained a coda nasal⁸ not preceded by *ə.

5.1.2.3. Stage 3. The final stage of Enggano historical phonology discussed here comprises two changes: shift of *g to *h and the loss of coda consonants.

*Shift of *g to *h.* This change necessarily happened in the last stage given the merger of *ŋ and *g, as explained above, and follows from a well-attested path of lenition.

Loss of Coda Consonants. Whether this was a single change or multiple changes is unclear, though the loss of coda nasals necessarily happened after the suprasegmentalization change since coda and final nasals triggered nasality. What happened to PSUM coda *r in medial clusters is unclear, given that no unambiguous reflexes have yet been identified in Enggano. The loss of coda consonants is a well-attested change crosslinguistically.

5.1.2.4. Stage 4. While not considered here for reconstruction purposes, a fourth stage, including the most recent changes from Old Enggano to the present day, can be described. The changes in Enggano since Kähler's work have been thoroughly detailed by Yoder (2014) and Edwards (2015), who notes a further erosion of final vowels as well as a split of Old Enggano /o/ into Modern Enggano /o/ and /ə/. The reader is directed to these sources for a more in-depth treatment of these changes.

5.1.3. Enggano demystified. Following from the above discussion, only two changes that Enggano has undergone should rightly be considered aberrant—the merger of *s and *t as *t and the suprasegmentalization of nasality. All other changes, though many in total, are well attested and rather natural. Given this,

8. Again, this includes the original PMP coda *l and *ir# that had already merged with *n.

and since Enggano appears to have undergone the six changes characteristic of Sumatran, it is reasonable to consider Enggano to be a Sumatran language.

5.2. FURTHER INTERNAL SUBGROUPING. It should now be clear that there are, at the very least, three lower-level groupings within the Sumatran subgroup, namely the Batak, NBI, and CBI languages. Below, a few possible higher-level subgroup proposals will be considered: Gayo–Batak (section 5.2.1), BI (section 5.2.2), NBI–CBI (section 5.2.3), CBI–Enggano (section 5.2.4), and Mentawai–Enggano (section 5.2.5), all of which are refuted without more evidence.

5.2.1. Gayo–Batak subgroup. Although there has been some discussion about the possible close relationship between Gayo and the Batak languages, there does not seem to be much evidence supporting this idea. Whereas Gayo and the Batak languages appear to have had some sort of mutual effect on each other, primarily in vocabulary, cognates for the innovative lexicon reconstructible for PB are mostly not found in Gayo. However, there are four phonological irregularities shared between PB and Gayo. The first two, PMP *daRaq ‘blood’ > PSUM *darah > irregular *darəh > Pre-G *rayoh, PB *darəh, and PMP *huaji ‘younger.sibling’ > PSUM *(w)agi > irregular *aŋgi > Pre-G *aŋgi, PB *aŋgi, are clearly not shared with other PSUM daughter languages as evidenced by PNBI *dala, PCBI *do, Nasal *χalah* and PNBI *aki, PCBI *axi, Pre-M *bagi (although Pre-E *āhĩ with unexpected nasalization), respectively. The remaining two irregularities, PMP *walu ‘eight’ > PSUM *walu > irregular *waluh > Pre-G, PB *waluh, and PMP *siwa ‘nine’ > PSUM *siwa > irregular *siwah > Pre-G, PB *siwah, could be argued to be unique to PB and Pre-G since Nasal, the only other Sumatran language which retained final *h, reflects *walu* and *suwai*, though this alone can hardly be taken as strong evidence for a Gayo–Batak connection. In addition, there are no clear phonological innovations shared between Pre-G and PB, further arguing against a Gayo–Batak subgroup.

5.2.2. BI subgroup. It may be tempting to try to subgroup all seven of the BI languages into one larger BI subgroup. While there do appear to be a number of shared lexical innovations between these languages, the shared historical sound changes between them do not constitute significant evidence for a BI subgroup. In fact, only two sound changes appear to be shared between these languages: PSUM *d to *r intervocalically and word-final obstruent devoicing, though both are very common crosslinguistically and present issues.

The first of these initially appears to be a shared innovation, but on further inspection, it is likely a parallel development. First, the shift of *d to *r (also an alveolar trill beside PSUM *r) intervocalically poses an issue as *r from PSUM *d remains as /r/ in PNBI, PCBI, Pre-M, and Pre-E, but *r from PSUM *r becomes PNBI *l and PCBI, Pre-M, and Pre-E Ø, thus providing evidence

of an unconditioned split. Such an issue could be resolved, however, if we assume that $*r_1 < \text{PSUM } *r$ and $*r_2 < \text{PSUM } *d$ were rhotics of two different qualities, one a tap and one a trill, though this contradicts the idea that both seem to be best reconstructible to an alveolar trill.

The second possibly shared sound change, word-final obstruent devoicing, similarly poses issues. The first is that this change is not evidenced in any way by the CBI languages or Enggano due to their complete loss of coda consonants. Thus, evidence of this change can only be found in the NBI languages and Mentawai. In addition, it is not possible that this change happened before the shift of $\text{PSUM } *k$ to $\text{PNBI } *ʔ$ since the NBI languages distinguish /k/ and /ʔ/ word-finally. Had word-final obstruent devoicing happened first, final $*k$ would necessarily have also shifted to $*ʔ$ (which clearly did not happen, as evidenced by Mentawai's merger of $\text{PSUM } *g\#, *k\#$). Again, on the basis of these complications and the fact that word-final obstruent devoicing is a very common sound change crosslinguistically, it is assumed that these are parallel developments. It follows that there is no strong phonological evidence for a BI subgroup within Sumatran.

5.2.3. Northern–Central BI subgroup. The evidence for a Northern–Central BI subgroup is not much stronger than that for a larger BI subgroup. Possible shared innovations from PSUM are as follows: shift of $*d$ to $*r$ intervocally, shift of $*p$ to $*f$ ($> \text{PNBI } *h$), and shift of $*k$ to $*ʔ$. The argument against the first of these has already been considered in section 5.2.2.

The sound change from $*p$ to $*f$ is shaky at best—there is no evidence for us to assume that $*p$ passed through an intermediate $*f$ before becoming $\text{PNBI } *h$. In addition, there are some PNBI words which contain $*p$ (namely in the environment ə_-V) that clearly derive from PSUM words with $*p$ but which are reflected as $*f$ in PCBI . If an earlier $*p$ to $*f$ shift is to be assumed, it would have overlooked such forms only for these to later shift in PCBI . These complications and the commonality of $*p$ to $*f$ and $*p$ to $*h$ changes suggest that this was not a single shared development.

As opposed to the first two sound changes, the shift of $*k$ to $*ʔ$ could realistically have been shared. PSUM forms ending in $*-ək$ versus $*-əg$ being inherited as PCBI final $*-o$ and $*-ə$, respectively, indicate that word-final devoicing had not occurred in PCBI , at least not before the $*k$ to $*ʔ$ shift as in PNBI (see section 5.2.2). Thus, the only shared phonological sound change in support of such a subgroup is the $*k$ to $*ʔ$ shift. Given that this is a very common sound change crosslinguistically, the supposed Northern–Central BI is not considered here to be a valid lower-level subgrouping.

5.2.4. Central BI–Enggano subgroup. The CBI languages and Enggano have undergone a considerable number of the same phonological innovations— $*k > *ʔ$, $*C\# > \emptyset$, $*R > \emptyset$ —though all of them must be attributed to parallel development.

While both Pre-E and PCBI *ʔ are traceable to PSUM *k, their development cannot have been shared since voiceless stops were unambiguously integrated with preceding nasals in PCBI, PSUM *ŋk sequences became PCBI *g but Pre-E *~ʔ. Should this innovation have been shared, one would instead expect PCBI *ŋʔ or *ʔ. This is clearly not the case, as exhibited by the PSUM word *ruŋkuŋ ‘throat’ > PCBI *ogu and Pre-E *ũʔü.

The loss of final consonants, too, must have been a parallel development. Had this been shared, one would expect PCBI to exhibit a system of nasality similar to that of Pre-E since it was precisely the final nasals which led to such a system’s development. The discrepancy is also found in the effects of PSUM final *r on preceding vowels. Though both PCBI and Pre-E share *ur#, *ər# > *o# changes, the vowels in original *ar#, *ir# are maintained in Pre-E but centralized in PCBI. Had the *r been lost at the same time, one would expect the vowels to share the same quality unless a later rule caused them to shift, though no such rule can be identified in PCBI since the original word-final *a and *i are maintained.

Finally, as was mentioned in the previous paragraph, the loss of *r must have been independent. Should it have been shared, the vowel changes should be evident in both daughter languages unless later developments caused them to change, though no such developments are evident. It is clear then that none of these three changes is shared between Enggano and PCBI.

5.2.5. Mentawai–Enggano subgroup. Mentawai is the next closest BI language to Enggano. However, there is little evidence that the two are closely related. While their relationship could be obscured due to the high lexical divergence of both languages, the low number of shared phonological innovations refutes such a claim. Those shared innovations are *h > Ø, *r > Ø, and *s > *t (Simalegi variety of Mentawai).

The first sound change, *h > Ø, is so common crosslinguistically that it should be considered only very weak evidence for a relationship between the two.

The loss of *r poses similar issues here as with the CBI–Enggano proposal. Whereas PSUM *u# is retained in both Pre-M and Pre-E, PSUM original *ur# sequences became Pre-M *u# and Pre-E *o#. Had *r been lost simultaneously in the two languages, the vowel correspondences would also be the same.

Counter to Edwards’s (2015) claim that “[n]o other language of western ISEA [Island Southeast Asia] displays such a change,” Pre-E’s remarkably strange development of *s, *t > *t is also found in one variety of Mentawai–Simalegi (Pampus 1994). Should this have been a shared innovation, however, Pre-E would necessarily have undergone all those changes characteristic of Pre-M, namely PSUM *ə, *y, *r > Pre-M *e, *j, Ø. However, all three of these pose issues: (i) there is no reason to believe that Pre-E *o derives from earlier Pre-M *e < PSUM *ə; (ii) there are no clear reflexes of PSUM *y in Pre-E to verify that *y > *j; and (iii) the issues with *r that were mentioned above.

Furthermore, there are seemingly no lexical innovations shared between Simalegi (or Mentawai as a whole, for that matter) and Enggano that might suggest a shared history, whether via a shared proto-language below PSUM, contact distribution, or network relation. Though it is tempting to suggest a closer connection on the basis of the shared **s*, **t* > **t* merger, the current evidence indicates that, albeit unique to these two speech varieties in Western Indonesia, the **s*, **t* > **t* merger should be considered coincidental parallel evolution, and thus there is no current demonstrable Enggano–Mentawai connection.

5.3. THE POSITION OF OTHER NON-MALAYO-CHAMIC LANGUAGES OF SUMATRA. We have proposed that all but two of the non-Malayo-Chamic language groups of Sumatra, Rejang and Lampungic, comprise a single subgroup. This raises questions about the position of Rejang and Lampungic within Malayo-Polynesian and whether there is evidence that either of them fall within Sumatran. Here, we present some of the challenges for including Rejang and/or Lampungic in Sumatran.

5.3.1. The position of Rejang. The history of Rejang, a language of southwest Sumatra, has been discussed at length. Blust (1984) was the first to outline Rejang historical phonology. Richard McGinn, in a series of articles, has further detailed these sound changes (McGinn 1997, 1999, 2000, 2003, 2005, 2009). McGinn (2005) outlines changes from PMP to Proto-Rejang (PR) based on data from five different varieties. If we consider the five mergers and one shift in PSUM, we see that four of the five mergers are attested in PR. However, there are significant complications in each case.

The mergers of PMP **z*, **d*, and PMP **ñ*, **n* are found in a subset of words, but not others. McGinn (2005) proposes that these are conditioned mergers where the merger of PMP **z*, **d* occurred word-initially in all but one word and the merger of PMP **ñ*, **n* only occurred intervocally (see McGinn 2005:21–22).

The merger of PMP **j* and **g* presents a different complication. All five Rejang varieties provide evidence of the merger word-medially. However, in four Rejang varieties, the merger occurs word-finally, but in the Rawas variety, PMP **j* > Rawas *-t*, which leads McGinn to posit that the merger did not occur until after the breakup of PR.

The merger of PMP **h*, \emptyset and the shift PMP **q* > PR **h* appear to have occurred as well. McGinn proposes that the distinction between PMP **R* and **r* is retained. He maintains that there was a later partial merger of PMP **R* and **r* in word-initial position in all varieties and word-final position in all but Rawas. There is evidence of the retention of PMP **R* and **r* found word-medially where PMP **R* surfaces as *-l-*, while PMP **r* surfaces as *h* in one variety and ʔ in the other four.

The complications involved in including Rejang in Sumatran are not trivial, especially the apparent retention of PMP **R* and **r* as well as the partial merger

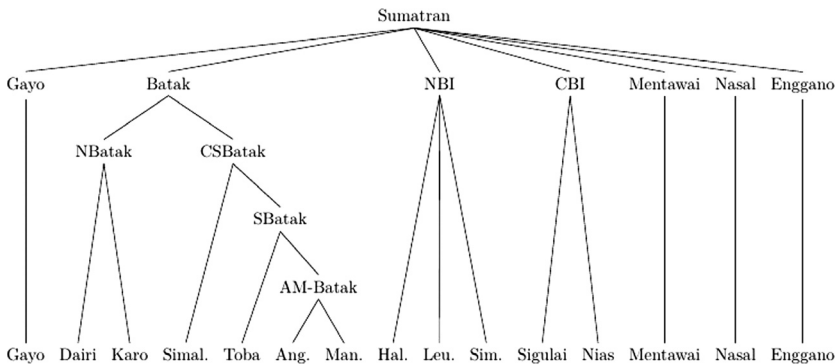
of PMP *j and *g. One potential avenue of research is an investigation of early loan words from Malay. Because such apparent loanwords have undergone significant sound changes, it is difficult to tell whether a given lexical item has undergone a series of adaptations or whether the item represents native vocabulary.

5.3.2. The position of Lampung. The history of the Lampungic languages of southern Sumatra has been detailed by Anderbeck (2007), in which he reconstructs parts of the phonology, morphology, and lexicon of Proto-Lampungic (PL). He also details a constellation of sound changes from PMP to PL. However, only one of the mergers found in PSUM occurs in PL, namely, PMP *h, Ø > PL Ø, in addition to the shift PMP *q > PL *h. Importantly, none of the PMP *j-*g, *z-*d, *ñ-*n, or *R-*r mergers occurred. Thus, Lampungic languages more clearly do not subgroup with Sumatran.

6. CONCLUSION. Based on the data considered above, it is clear that PSUM stands as its own branch of PMP, as defined by five distinct mergers and one shift. Of the fifteen languages considered here, all of them fulfill these six criteria and are established as daughter languages within Sumatran. Furthermore, it has been shown that Gayo, Nasal, and Enggano, which have until now remained unclassified, are members of the Sumatran subgroup.

Within Sumatran, there appear to be at least seven distinct branches—Gayo, Batak, NBI, CBI, Mentawai, Enggano, and Nasal. As explored in section 5.2, no strong arguments based on shared phonological innovations can yet be made to establish further higher-level subgroups within Sumatran. It is clear that more work needs to be done in the realms of morphological and syntactic reconstructions as well as the identification of shared lexical innovations in order to further explore such proposals. The proposed Sumatran internal subgrouping tree is shown in figure 3.

FIGURE 3. PROPOSED SUMATRAN INTERNAL SUBGROUPING TREE.



APPENDIX B. PB RECONSTRUCTION OF SELECTED LEXICAL ITEMS

The lexical items cited here can be found in: Manik (2002) for Dairi; Neumann (1951) for Karo; Syarfina et al. (2015) for Simalungun; Warneck (1977) for Toba; and Eggink (1936) for Angkola. Since the best existing source for Mandailing lexical data is, primarily, an Angkola dictionary (Eggink 1936), and since Mandailing lexical forms are frequently near equivalent to their corresponding Angkola forms,⁹ Mandailing data will not be cited here. Geminate consonants are represented with their typical orthographic form (written as nasal-stop sequences) for Simalungun, Toba, and Angkola.

PMP	PB	Dairi	Karo	Simalungun	Toba	Angkola	Gloss
1. *ambay	*ambey	ambe	ambe	ambey	ambe	ambe	to swing
2. *babuy	*babuy	babi	babi	babuy	babi	babi	pig
3. *baləj	*baləg	baləŋ	baləŋ	balog	balok	balok	border
4. *baRəq	*barəh	baroh	barəh	bay/joh ¹⁰	baro	baro	ulcer
5. *bayaw	*bayow	bayo		bajo ¹¹	bayo	bayo	in-law
6. *bələj	*bəlag	bələŋ		bolag	bolak	bolak	broad
7. *biRaŋ	*birah	birah	birah	birah	bira	bira	k.o. taro
8. *buRiq	*burih	borih	burih	burih	buri		to wash
9. *daRaŋ	*darəh ¹²	daroh	darəh	daroh	daro	daro	blood
10. *dilaq	*dilah	dilah	dilah	dilah	dila	dila	tongue
11. *ənəm	*ənəm	ənəm	ənəm	onom	onom	onom	six
12. *hapuy	*apuy	api	api	apuy	api	api	fire
13. *hutək	*utək	otok	utuk	utok	utok	utok	brain
14. *ijunŋ	*igunŋ	egunŋ	igunŋ	igunŋ	igunŋ	igunŋ	nose
15. *ikuR	*ikur	ekur	ikur	ihur	ihur	ihur	tail
16. *ipən	*ipən	epən	ipən	ipon	ipon	ipon	tooth
17. *ipəs	*ipəs	epəs	ipəs	ipos	ipos	ipos	cockroach
18. *kahiŋ	*kayu	kayu	kayu	hayu	hau	hayu	wood
19. *kitay	*kitey	kite	kite	hitey	hite	hite	bridge

9. Of the major phonological differences, the clearest are the loss of /h/ and the retention of earlier nasal-stop sequences (rather than realization as geminates) in Mandailing.

10. Simalungun has sporadically changed intervocalic /t/ to /y/ and, in some forms, /y/ (whether original or from *r) to <j> /d͡ʒ/. See also #5, 56, and 71.

11. See the footnote for #4. PB *ow > Simalungun /o/ is irregular in this case.

12. PMP *a > PB *ə is an irregular development, one that would have made PB *darəh 'blood' homonymous with PMP *daRəq > PB **darəh 'clay'. The latter, however, is possibly instead reflected as PB *dah, as in Karo *dah*.

PMP	PB	Dairi	Karo	Simalungun	Toba	Angkola	Gloss
20. *kunj	*kunj	konij		hunj ¹³	hunik	hunik	turmeric
21. *laki	*laki	laki	laki	la-lahi	lahi	lahi	male
22. *lapaR	*lapar	rapar	lapar	lapar	rapar	rapar	hungry
23. *ləŋgur	*ləŋgur	rəŋgur	ləŋgur	ləŋgur	roŋgur	roŋgur	thunder
24. *m-atay	*matey	mate	mate	matey	mate	mate	to die
25. *panaw	*panow	pano	pano	panou	pano-pano	pano	k.o. skin disease
26. *pa-udəhi	*pudi	podī	pudi	pudi	pudi	pudi	back
27. *pusəj	*pusəg	posoj	pusuj	pusog	pusok		navel
28. *qabaŋ ¹⁴	*kabaŋ	kabaŋ	kabaŋ	habaŋ	habaŋ	habaŋ	to fly
29. *qadəp	*hadəp	hadəp	adəp	adop	adop	adop	front
30. *qajəŋ	*agəŋ	agəŋ	agəŋ	agoŋ	agoŋ	agoŋ	charcoal
31. *qaləjaw	*ləgow	ləgo	ləgo	logou	logo	logo	dry
32. *qapəjux	*pəgu	pəgu	pəgu	pogu	pogu	pogu	bile
33. *qatay	*atey	ate	ate	ate-atey	ate-ate	ate-ate	liver
34. *qə(n)səm	*acəm	acəm	acəm		asom	asom	sour
35. *quləj	*uləg	oləŋ		ulog	ulok	ulok	snake
36. *qunəj	*unəg	onəŋ	unəŋ	unok ¹⁵	unok	unok	pith
37. *saŋkəb	*saŋkəb	saŋkəm		saŋkob	saŋkop	saŋkop	to close
38. *səRəd	*səRəd	səRən	səRən	sorod		sorot	to sting
39. *tabəq	*tabəh	taboh	tabəh	taboh	tabo	tabo	fat
40. *tanəq	*tanəh	tanoh	tanəh	tanoh	tano	tano	earth
41. *təbuh	*təbu	təbu	təbu	tobu	tobu	tobu	sugarcane
42. *tələn	*tələn	tələn	tələn	tolon	tolon	tolon	to swallow
43. *təRəb	*təRəb	təRəm	təRəm		torop	torop	many
44. *ukir	*ukir	okir		uhir	uhir	uhir	to carve
45. *utaq	*utah	utah	utah	utah	uta	uta	vomit
46. *uyəg	*muyəg	moyəŋ	muyəŋ	muyog	miok	muyok	shaky
47. *zakan	*dakan	dakan	dakan	dahan	dahan	dahan	rice
48.	*alud	alun	alun	alud	alut		to massage
49.	*badow	bado	bado	badou		bado-bado	k.o. fish

13. PB *g > Simalungun /ŋ/ is irregular in this case.

14. PMP *q > PB *k in this lexical item is an irregular development.

15. PB *g > Simalungun *k is irregular in this case.

PMP	PB	Dairi	Karo	Simalungun	Toba	Angkola	Gloss
50.	*bəlur	bərur	bəlur	bulur ¹⁶	borur	borur	to swallow
51.	*buey	bue	bue	buey	bue	bue	many
52.	*ciŋəm	ceŋəm	ciŋəm		siŋom	siŋom	tight
53.	*cirəm	cirəm	cirəm	sirom			to smile
54.	*dəlэг	dəlэг	dəlэг	dolog	dolok	dolok	mountain
55.	*galar	garar	galar	galar	garar	garar	to pay
56.	*gayaw		gayo	ganjow ¹⁷	gayo	gayo	k.o. crab
57.	*habab	habam	abam	abab	abap	abap	ash
58.	*idow	ido	ido	idou	ido	ido	debt
59.	*imbow	imbo	imbo	imbou	imbo	imbo	k.o. monkey
60.	*kalak	kalak	kalak	halak	halak	halak	person
61.	*kaləd	kalən	kalən		halot	halot	swollen spleen
62.	*kəlukuy ¹⁸	kəliki	kuliki	haluhuy		halihi	k.o. bird
63.	*labar	rabar	labar	labar	rabar	rabar	k.o. food
64.	*məsэг	məsэг	məsэг	mosog	mosok	mosok	to burn
65.	*pəduy ¹⁹	pədi	pədi	ampoduy	ampodi	ampodi	bamboo mouse
66.	*picat	picat		pisat	pisat	pisat	to pinch
67.	*pituj	pətuj	pituj	pituj	pituj		blind
68.	*rəh	roh	rəh	roh	ro	ro	to come
69.	*sukat	sukat	sukat	suhat	suhat	suhat	k.o. taro
70.	*suki	soki	suki	suhi-suhi	suhi	suhi-suhi	corner
71.	*tarub	tarum	tarum	tayub ²⁰	tarup	tarup	roof
72.	*tənduy	təndi	təndi	tənduy	təndi	təndi	spirit
73.	*tukər	təkor	tukur	tuhor	tuhor		to buy
74.	*tuyuj	tuyuj		tuyuj	tiuj	tuyuj	k.o. eggplant
75.	*uncim	oncim	uncim		unsim		k.o. banana

16. PB *ə > Simalungun /u/ appears irregular in this case.

17. See the footnote for #4. The extracent nasal is irregular.

18. The developments of the penultimate and prepenultimate vowels for this form are irregular in many of the Batak languages. Until these irregular changes are better understood, PB *kəlukuy is tentatively reconstructed.

19. The presence of an initial *am-* in all of the SB languages is mysterious. This is tentatively reconstructed as PB *pəduy.

20. See the footnote for #4.

APPENDIX C. PNBI RECONSTRUCTION OF SELECTED LEXICAL ITEMS

The lexical items cited here can be found in: Tim Kamus Balai Bahasa Banda Aceh (2013) for Haloban; Candrasari and Khalsiah (2018) for Leukon; and Kähler (1961) for Simeulue and Leukon. There are few sources for NBI languages' lexicons, but even fewer for in-depth phonological descriptions. Given this, there are many vowel realizations/developments that are not well understood for the NBI languages. Most of these include simple *u~o* or *i~e* interchanges, but some are more complex. Given the lack of further lexical items to make comparison and the inaccuracy of certain transcription systems, these finer vowel changes from PNBI to the NBI languages will not be tackled.

	PMP	PNBI	Haloban	Leukon	Simeulue	Gloss
1.	*aku	*d-eʔu ²¹	reo	deu	deʔu	1SG
2.	*babaw	*a-babaw		a-fafo	a-fofaw ²²	shallow
3.	*bahi	*silabay	silawe	silafe	silafay	woman
4.	*baliw	*balew		fale ²³	faleu	to return
5.	*baRiw	*a-balew	a-wale		a-faleu	rotten
6.	*batu	*batu	watu	batu	batu	stone
7.	*bəRŋi	*bəŋi	wəŋi	bəŋi	bəŋi	night
8.	*bibiR	*bibil	wewel	bifil	bifil	lips
9.	*biRaŋ	*bila	wila		bila	k.o. taro
10.	*buka-i	*buʔay	wuwe	fue	buʔay	to open
11.	*buluŋ	*buluŋ	woluŋ	buluŋ	buluŋ	leaf
12.	*buRaw	*a-bulaw	a-wulo	a-fulo	a-fulaw	to go far
13.	*dahun	*don	ron		don	medicine
14.	*daləŋ	*dalək	ralək		dalək	k.o. fly
15.	*daŋan	*dan	ran		dan	long
16.	*daRaŋ	*dala	rala	dala	dala	blood
17.	*dəmdəm	*ma-rəndəm		mə-rəndəm	ma-rədəm	to remember
18.	*dilaŋ	*dila	rila	dila	dila	tongue
19.	*gumis	*kumis		xumi	kumi	beard
20.	*hapuy	*ahoy	ahe	ahe	ahoy	fire

21. From apparent earlier, *aku > *di-aku > *d-eku > *d-eʔu. Such vowel changes commonly occurred in PNBI, even across morpheme boundaries, as in PMP *ma-qitəm > *ma-itəm > PNBI *metəm.

22. PNBI *a > Simeulue /o/ in this case is taken to be a sporadic change under influence from the following *aw.

23. For PNBI *bobuy, *boluy, and *balew, Leukon and Simeulue initial /f/ may simply be a back-formation from intervocalic forms (as in Simeulue *a-falew*), where /f/ is expected.

	PMP	PNBI	Haloban	Leukon	Simeulue	Gloss
21.	*ijuj	*ikuŋ	ixuŋ	ixuŋ	ixuŋ	nose
22.	*ipən	*ihən	iyən	ihən	ehen	tooth
23.	*kasaw	*asaw		aso	asaw	rafter
24.	*l<um> aŋuy	*l<um> aŋoy	ma-laŋe	l<um> aŋe	l<um> aŋoy	to swim
25.	*lakaw	*laʔaw	ma-lao	l<um>alao	laʔo ²⁴	to walk
26.	*lijay ²⁵	*likay	lixé	lixé	lixay	palm rib
27.	*maN- quhaw	*maŋ-aw	maŋo- maŋo	maŋo	maŋaw	thirsty
28.	*ma-panas	*ma-hanas	ma-hanat ²⁶	ma-hanas	ma-anas	hot
29.	*m-atay	*matay	mate	mate	matay	dead
30.	*ŋajan	*kakan ²⁷		xaxan	kaxan	name
31.	*pajay	*akay	axe	axe	axay	rice
32.	*paRih	*hali	ali	hali	ali	stingray
33.	*puluq	*hulu		hulu	ulu	ten
34.	*punay	*unay		une	unay	k.o. dove
35.	*qabu	*abu	awu	afu	afu	ash
36.	*qalu-hipan	*alihan	alian	alihan	alian	centipede
37.	*quay	*uay	uwe	ue	uay	rattan
38.	*siku	*siʔu		siu	siʔu	elbow
39.	*suka	*suʔa		sua	suʔa	k.o. tree
40.	*taba	*t<in> aba ²⁸	t<in> awa ²⁹		t<in>afa	garden
41.	*timba	*timba		timba	tibaʔ ³⁰	to get water
42.	*tubah	*tuba	tuwa		tufa	k.o. plant
43.	*zauq	*ma-raw	a-ro	a-ro	a-raw	far

24. The PNBI diphthong appears to have been irregularly reduced in Simeulue. The fact that the diphthong is reconstructible to PNBI is apparent from the final vowel correspondences.

25. For this reconstruction, see Erik Zobel's PMP addenda at <https://ezlinguistics.blogspot.com/p/pmp-addenda.html> (accessed March 5, 2024).

26. The status of final PNBI *s is unclear. Given disagreement among sources and possible influence from Minangkabau final *s changes, these changes are not discussed here. The development of Haloban final /t/, however, appears to be irregular.

27. PMP *ŋ > PNBI *k is an irregular development. This change likely occurred via *ŋ > *g > *k under the influence of the PMP *j > *g > *k in the adjacent syllable.

28. PMP *taba 'cut with a knife' > PNBI *t<in>aba 'garden' (i.e., that which has been cut away).

29. See von Rosenberg (1855b:467), for this cited form.

30. The development of a final /ʔ/ in Simeulue is irregular.

PMP	PNBI	Haloban	Leukon	Simeulue	Gloss
44.	*alambuk		alambu?	alabu?	yard
45.	*a-lelew		a-lele	a-leleu	to follow
46.	*bəŋay	wəŋe	bəŋe	bəŋay	k.o. worm
47.	*boboy		fofe	bofoy	to visit
48.	*boloy		fole	boloy	to wash
49.	*əndəŋ	ədəŋ	əndəŋ	ədəŋ	pig
50.	*iban	iwan		ifan	betel
51.	*kaol	xaol	xaol	kaol	banana
52.	*ondil		ondil	odil	turmeric
53.	*sesew	sese	sese	seseu	to tell
54.	*simbi	sibi	simbix ³¹	sibix	chin
55.	*siŋkew		siŋke	siŋkeu	k.o. snail
56.	*təm̄ba-təm̄ba		təm̄ba-təm̄ba	təba-təba	palate
57.	*tənday	tədəh ³²	tənde	təday	middle
58.	*timbəl	tibəl	timbəl	tebəl	smoke
59.	*tindaw	tido	tindo	tidaw	to request

APPENDIX D. PCBI RECONSTRUCTION OF SELECTED LEXICAL ITEMS

The lexical items cited here can be found in: Kähler (1959) for Sigulai and Sundermann (1905) for Nias.

	PMP	PCBI	Sigulai	Nias	Gloss
1.	*babaw	*a-babə	a-fafə	a-βaβɣ	shallow
2.	*banua	*banua	banua	banua	world
3.	*baqbaq	*baba	bafa	baβa	mouth
4.	*baqəñan	*bahəna	baəna	bahɣna	to sneeze
5.	*baqəRaŋ	*boha	boa	boha	molar
6.	*basəq	*a-basə	a-baə	a-basɣ	wet
7.	*bibiR	*bebe	befe	beβe	lips
8.	*bitiəs	*bisi	bihi	bisi	shin
9.	*dəpah	*dəfa	dəfa	dɣfa	fathom
10.	*əpat	*əfa	əfa	ɣfa	four
11.	*hajək	*aŋo ³³	aŋo	ago	to sniff

31. The development of a final /x/ in Leukon and Simeulue appears to be irregular.

32. This final /h/ is an irregular development in Haloban.

33. This PCBI *g implies an excrescent nasal (i.e., PCBI *g < *ŋg, *ŋk), an innovation also found in PB *aŋgəh and PNBI *aŋkə. The PMP *ə > PCBI *o here is unexplained.

	PMP	PCBI	Sigulai	Nias	Gloss
12.	*haRəzan	*ora	ola	ora	ladder
13.	*pa-udəhi	*furi	fuli	furi	back
14.	*pudul	*a-furu	a-fulu	a-furu	blunt
15.	*qabaŋ	*homb ³⁴	ombo	hobo	to fly
16.	*qabu	*habu	afu	haβu	ash
17.	*qajəŋ	*axo	axo	axo	charcoal
18.	*qalu-hipan	*alifa	alifa	alifa	centipede
19.	*qaRta	*ata	ata	ata	person
20.	*qasin	*asi	ahi	asi	sea
21.	*qasiRa	*asio	aiyo	asio	salt
22.	*qatəluR	*adulo ³⁵	adulo	adulo	egg
23.	*qətut	*hətu	ətu	hɣtu	fart
24.	*qitəm	*a-itə	a-itə	a-itɣ	black
25.	*quban	*huba	ufa	huβa	gray hair
26.	*qudip	*a-uri	a-uli	a-uri	alive
27.	*qunəj	*hunə	unə	hunɣ	pith
28.	*siji	*sixi	ixi	sixi	to winnow
29.	*siku	*siʔu	iʔu	siʔu	elbow
30.	*tabəq	*tabə	tafə	taβɣ	fat
31.	*təbuh	*təbu	təfu	tɣβu	sugarcane
32.	*tuba	*tuba	tufa	tuβa	k.o. plant
33.	*tuzuq	*turu	tu-tulu	turu	to point
34.		*bandi	bandi	badʔi	k.o. fish
35.		*fa-əndu-ʔə	f-əndu-ʔə	fa-ɣdʔu-ʔɣ	to call
36.		*ləmba	ləmba	leba	to adhere
37.		*m-ondi	m-ondi	m-odʔi	to bathe
38.		*s<um>indo	um-indo	s<um>idʔo	to stand
39.		*simbi	imbi	sibi	chin
40.		*simbo	imbo	sibo	smoke
41.		*talimbo	talimbo	talibo	k.o. mushroom
42.		*tendo	tendo	tedʔo-tedʔo	support pole

34. This PCBI *mb implies an excremental nasal, an innovation also found in PNBI *hambaŋ.

35. This PCBI *d implies an excremental nasal (i.e., PCBI *d < *nt), an innovation also found in Pre-M *antelu.

APPENDIX E. PSUM RECONSTRUCTION OF SELECTED LEXICAL ITEMS

For the lexical data cited here for PB, PNBI, and PCBI, please see the references given in their respective appendices; reconstructions that are not found in appendices B–D are made based on data from the sources cited there. For the remaining four, unless otherwise specified, the lexical items cited here can be found in: Hazeu (1907) for Gayo; Morris (1900) for Mentawai; and Kähler (1987) for Old Enggano.³⁶ For Nasal data, the reader is referred to the ongoing documentation work (McDonnell 2019).

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
1.	*aku	*aku	aku	*aku	*aʔu		aku			1SG
2.	*anak	*anak	anak	*anak	*anaʔ	*ono		ada	anaʔ	child
3.	*atas	*atas		*atas	*atas		ata		atas	above
4.	*b[ə]ʔitəs	*b[ə]ʔitəs	tis	*bitis	*bitis	*bisi	biti			shin
5.	*babuy	*babuy		*babuy	*baboy	*babi	babuy ³⁷		babuy	pig
6.	*bahuq	*bau		*bau	*bo	*bəu		pau		to smell
7.	*baliw	*baliw			*balew	*bali	baliw			to return
8.	*banua	*banua		*banua	*bano	*banua	manua ³⁸	pado ³⁹		world
9.	*baqbaq	*babah	awah ⁴⁰	*babah	*baʔba	*baba		papa ⁴¹		mouth

36. As mentioned previously, Enggano phonology has changed drastically since the data for Kähler's (1987) dictionary were collected. See Yoder (2014) and Edwards (2015), for discussions of some of these changes.

37. See Adriani (1928:27), for this attested form.

38. PSUM *banua 'world' > Mentawai *manua* 'heaven'. Irregular *b > /m/.

39. PSUM *banua 'world' > Enggano *pado* 'placenta'.

40. In this and other forms, a zero-reflex of PSUM *b is found in Gayo rather than the expected *w. One proposal to explain such forms was made by Shorto (1976:205–06), who posited a set of semivowel rules in which there is a preference against semivowels in adjacent syllables and for /w/ before /i/. This explains many, but not all, of the irregularities of this sort seen here.

41. PSUM *babah 'mouth' > Enggano *papa* 'cheek'.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
10.	*baqəRu	*bahəru	ayu	*baru	*balu	*bohəu	bau			new
11.	*baRaq	*barah		*baru	*bala	*bo	bo			lung
12.	*baRəq	*barəh		*barəh	*balə	*bao		pao ⁴²		ulcer
13.	*baRiw	*bariw		*bari	*baləw					rotten
14.	*baRu	*baru		*baru	*balu	*bou				k.o. tree
15.	*batu	*batu	watu	*batu	*batu	*batu	batu	~paku ⁴³	watu	stone
16.	*bəŋkik	*bəŋkik		*bəŋkik	*bəŋi?	*bəgi				bat
17.	*bəRŋi	*bəŋi	inji	*bəŋi	*bəŋi	*boŋi			dəŋi	night
18.	*bəsuR	*bəsur		*bəsur	*bəsul	*buso				satiated
19.	*bibiR	*bibir		*bibir	*bibil	*bebe	bibo ⁴⁴			lip
20.	*biRaq	*birah		*birah	*bila	*bio	bio	pia ⁴⁵		taro
21.	*birbir	*birbir	iwi ⁴⁶	*bibir		*bebe	bebe			bank of a river
22.	*buaq	*buah		*buah	*bo	*bua	bua	pua	wah	fruit
23.	*buhək	*buək	uk	*buk	*bu?	*bu			uwu?	hair
24.	*buka	*buka	ukə	*buka	*bu?ay		buka		buko	to open
25.	*bukuh	*buku		*buku	*bu?u	*bu?u				joint

42. PSUM *barəh 'ulcer' > Enggano *pao* 'wound'.43. PSUM *batu 'stone' > Enggano *~paku* 'anchor'. Unexplained source of nasality.

44. Irregular *i > /o/.

45. PSUM *birah 'taro' > Enggano *pia* 'garden'.

46. Although not entirely regular, this and other forms listed here show PSUM *r > Gayo /h/ / i_# rather than Ø from vowel coalescence.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
26.	*bulu	*bulu		*bulu			bulu	pudu		body hair
27.	*buluŋ	*buluŋ	uluŋ	*buluŋ	*buluŋ	*bulu	buluŋ ⁴⁷	~pudu ⁴⁸		leaf
28.	*bunuq	*bunuh		*bunuh	*bunu	*bunu		pudu		to kill
29.	*buqaya	*buhaya		*buaya	*buaya	*buaya			bəhayo	crocodile
30.	*buRaw	*buraw	biow	*burow						to chase away
31.	*buRiq	*burih		*burih			bui			to wash
32.	*dahun	*daun		*daun	*don	*də-rəu				leaf
33.	*dakəp	*dakəp		*dakəp					χaku?	to hug
34.	*daki	*daki			*dai	*daʔi				filth
35.	*daləm	*daləm	rələm	*daləm			lalem ⁴⁹		dələm	inside
36.	*daliŋ	*daliŋ	ralik	*daliŋ	*daliŋ	*lali		~dadi		root
37.	*daRaŋ	*darah	rayoh ⁵⁰	*darəh	*dala	*do			χalah	blood
38.	*daRəŋ	*darəh			*dao				χaluh	clay
39.	*dələj ⁵¹	*dələŋ		*dələŋ	*dələŋ	*lələ	ləlew			mountain
40.	*dəŋəR	*dəŋəŋ			*rəŋə			dohor ⁵²	dəŋul	to hear
41.	*dəpəh	*dəpa		*dəpa	*dəpa					fathom

47. See Lenggang (1992:98), for the attested final /ŋ/.

48. Alternates with *pudu* (no nasality).

49. See Adriani (1928:6), for this attested form.

50. Irregular PSUM *a > Gayo /o/ (from earlier *ə > o / _h#). Innovation shared by PB.

51. For this reconstruction, see Erik Zobel's PMP addenda at <https://ezlinguistics.blogspot.com/p/pmp-addenda.html#delej> (accessed March 5, 2024).

52. Assumed fossilized PSUM *-i suffix.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
42. *dilaq	*dilah		*dilah	*dila	*lela	lila			tongue
43. *duha	*dua	rowa	*dua	*do	*dua	dua ⁵³	dua	χuo	two
44. *duRi	*duRi	rui			*doi	dui ⁵⁴	dudur ⁵⁵	χuli	thorn
45. *ənam	*ənam	onom	*ənam	*nam	*əna	enem		num	six
46. *əpat	*əpat	opat	*əpat	*hat	*əfa	epat	opa	paʔ	four
47. *gələŋ	*gələŋ		*gələŋ	*lakaləŋ		geleŋ ⁵⁶	hodo		worm
48. *gulay	*gulay		*guley	*kulay	*gule			gulay	a dish
49. *hajək	*angəh		*angəh	*angkə	*ago				to kiss
50. *hapuy	*apuy		*apuy	*ahoy				apuy	fire
51. *haRəzan	*rədan		*rədan	*aeran	*ora	orat ⁵⁷			ladder
52. *hasaŋ	*asaŋ			*asaŋ		asaŋ ⁵⁸	~aka		gills
53. *həmay	*əmay		*əmey					may	rice
54. *huab	*huab ⁵⁹	wap	*huab		*ua	oap			yawn
55. *hulij	*ulig	ulik	*ulig						to hug
56. *ihəq	*iəh	oyoh	*iəh		*iə			iuh	urine

53. See Lenggang (1992:111), for the attested initial /d/.

54. See Adhiani (1928:46), for the attested initial /d/.

55. Assumed fossilized reduplicant.

56. See Lenggang (1992:90), for the attested final /ŋ/.

57. None of the sources investigated here attest to the expected final /n/.

58. See Lenggang (1992:127), for the attested final /ŋ/.

59. Irregular PMP *h > PSUM *h.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
57.	*ijuj	*iguŋ	iuŋ	*iguŋ	*ikuj	*ixu		ihu ⁶⁰		nose
58.	*ikuR	*ikur	uki ⁶¹	*ikur		*iʔo		iʔo		tail
59.	*iluR	*ilur	ulih ⁶²		*ilul	*ilo				spit
60.	*inum	*inum	inum	*inum	*inum	*inu			inum	to drink
61.	*ipən	*ipən	ipən	*ipən	*ihən	*ifə				tooth
62.	*kahiw	*kayu	kayu	*kayu	*ayu	*eu			kayu	wood
63.	*kaka	*kaka	kaka			*ʔaʔa		ʔaʔa		elder sibling
64.	*kami	*kami	kami	*kami	*ami	*ami	kai ⁶³	ʔai	kam	1P.EXCL
65.	*kapal	*kapal		*kapal				~ʔapa		thick
66.	*kaRat	*karat	ket	*karat	*alat	*o				to bite
67.	*kawil	*kawil					kabili	~ʔabi	kawil	fishhook
68.	*kebəl	*kebəl		*kebəl	*əpəl					invulnerable
69.	*kita	*kita	kitə	*kita	*ita	*ita		ʔika	kito	1P.INCL
70.	*kultit	*kultit	kultit	*kultit	*kultit	*uli	kulit	ʔulit ⁶⁴		skin
71.	*kutu	*kutu	kutu	*kutu	*utu	*utu			kutu	louse
72.	*lahud	*laud	laut		*lor	*lau				sea

60. PSUM *iguŋ 'nose' > Enggano *ihu* 'bow of a boat'. Unexplained loss of nasality.

61. Irregular vowel metathesis.

62. Irregular vowel metathesis.

63. Irregular *m > Ø. Shared by Enggano.

64. See Modigliani (1894:275), for the attested //.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
73.	*laki	*laki		*laki	*lai	*laʔi				male
74.	*lanjit	*lanjit	lanjit	*lanjit		*lanji	lanjit		lanjit	sky
75.	*lanjuy	*lanjuy		*lanjuy	* <um> anjoy	*lanji	ranjoy ⁶⁵		lanjuy	to swim
76.	*lapaR	*lapar	lapey	*lapar		*lofo	lapa			hungry
77.	*lawaq- lawaq	*lawah- lawah	lə- lawah	*lawah- lawah	*awal- awal	*lawah- lawah			lə- lawah	spider
78.	*layaR	*layar			*layal	*loyo	lajo			to sail
79.	*layu	*layu	layu			*leu	laju			to wilt
80.	*ləbuR	*ləbur ⁶⁶			*ləpul		lebu ⁶⁷	lopo		mud
81.	*ləngur	*ləngur		*ləngur	*ləŋkul		lə-lenggu ⁶⁸			thunder
82.	*lijay ⁶⁹	*lijay		*liji	*likay	*lixé	ligéy ⁷⁰			palm leaf rib
83.	*lima	*lima	limə	*lima	*lima	*lima	lima	liba	limo	five
84.	*luhəq	*[ɔ̃]luh	luh	*luh			elu	elo ⁷¹		tears
85.	*manuk	*manuk	manuk	*manuk	*mano?	*manu	manu		manu?	bird

65. Irregular PSUM *l > Mentawai *r.

66. PMP *ləbuR 'mud' > PSUM *ləbur 'mud;land;earth'.

67. PSUM *ləbuR 'mud;land;earth' > Pre-M 'to wallow (as of pigs)'.

68. See Pampus (1989:87), for this attested form.

69. For this reconstruction, see Erik Zobel's PMP addenda at <https://ezlinguistics.blogspot.com/p/pmp-addenda.html#lijay> (accessed March 5, 2024).

70. See Pampus (1994:117), for this attested form.

71. Kähler (1987:51) cites *edo*. The original /l/ is attested by Helfrich and Pieters (1891:552) *belo*, possibly with the active prefix *b(u)-* (see Edwards 2015:73).

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
86.	*mata	*mata	mata	*mata	*mata	*mata	mata	baka	mato	eye
87.	*matay	*matay	matay	*matay	*matay	*mate	matay	ka-bake	matay	to die
88.	*ñam-ñam	*nanam		*nanam	*nanam		nanam			to taste
89.	*ñamuk	*namuk	namuk	*namuk	*namuk					fly
90.	*ñamuR	*namur	nami	*namur		*namo				dew
91.	*nanaq	*nanah		*nanah	*nana	*nana			nanah	pus
92.	*nəpuq	*nəpuh			*nəpu	*nəfu				k.o. fish
93.	*nipay	*hənipay	nipey	*nipey					hənipay	snake
94.	*-ŋa-	*-ŋa-			*-ŋa-	*-ŋa-	-ŋa-	-ha-		numeral ligature
95.	*ŋuda	*ŋuda		*ŋuda	*ŋura			huda ⁷²	ŋuɣo	young
96.	*pajay	*pagay		*pagey	*akay	*faxe			pəhay	rice
97.	*pangaj	*pangaj		*pangaj		*faga			pangaj	to roast
98.	*paRa	*para		*para	*ala	*fo			palo	attic
99.	*paRaw	*paraw	payow	*parow	*halaw	*fau				hoarse
100.	*paRih	*pari		*pari	*hali	*foi	poi			stingray
101.	*pəñu	*pəñu		*pənu	*əno	*fənu	penuŋ	~e-punu-oʔoi		turtle
102.	*pəzəm	*pədəm		*pədəm		*məɾə	merem		pədum	to sleep
103.	*pija	*piga	piən	*piga	*ika		piga	piha	piɣo	how many
104.	*pitu	*pitu	pitu	*pitu	*itu	*fitu	pitu		pitu	seven

72. PSUM *ŋuda 'young', > Enggano *huda* 'woman'.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
105.	*puluq	*puluh	puluh	*puluh	*hulu	*fulu	pulu	pulu ⁷³	puluh	ten
106.	*pusəj	*pusəg	pusok	*pusəg		*fusə	pusow	puko	pusuk	navel
107.	*qabat[ə]R	*bat[ə]r		*badir		*batəl		~paki		wood worm
108.	*qabu	*habu	wau	*habu	*abu	*habu	abu			ash
109.	*qədəp	*hadəp		*hadəp	*arəp					front
110.	*qəjəŋ	*agəŋ		*agəŋ	*akəŋ	*axo				charcoal
111.	*qələjəw	*ləgəw	low	*ləgəw	*ləkaw	*ləxə	legew			day
112.	*qalu-hipan	*halu-ipan	lipən	*lipan	*alihən	*alifa	alupa ⁷⁴		həlupan	centipede
113.	*qambəwəŋ	*hambəwəŋ			*ambəwəŋ	*hambəwə				k.o. mango
114.	*qañud	*hanud	anut	*anut						to wash away
115.	*qaqay	*hahay		*kahey		*ahe		ae		foot
116.	*qaRta	*arta		*ata		*ata				person
117.	*qasiRa	*asira		*sira	*asila	*asio			silo	salt
118.	*qatay	*hatay	atey	*atey	*atey	*ate	atey		hatay	liver
119.	*qatəluR	*hantəluR		*təluR		*adulo	antelu ⁷⁵		həntəluR	egg
120.	*qənuR	*hənuR				*huno	enu ⁷⁶			path

73. See Walland (1864:123), for the attested /l/.

74. None of the sources investigated here attest the expected final /n/.

75. See Lenggang (1992:229), for the attested /nt/ sequence.

76. Irregular *r > /ŋ/.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
121.	*qəʔut	*həʔut	utut	*[əʊ]ntut	*bəʔut	*həʔu	etut	~uku	həʔuʔ	fart
122.	*qilɪR	*hilɪr	kilɪh ⁷⁷		*elɪl	*ele				to flow downstream
123.	*quay	*huay	wey		*uay	*ue			huway	rattan
124.	*quban	*huban	uwən	*uban	*uban	*huban	uban ⁷⁸		huban	gray hair
125.	*qudɪp	*hudɪp	urɪp		*unap	*huna	(⁷⁹)		huyiʔ	to live
126.	*quhənəp	*huənəp							hənaʔ	fish scale
127.	*quləj	*huləg		*uləg	*ulak	*ulə	ulow		huluk	snake
128.	*qulu	*hulu	ulu	*ulu	*ulu	*ulu		ulu ⁸⁰	hulu	head
129.	*qunəj	*hunəg		*unəg	*unok	*hunə	unow			pith
130.	*qutək	*utək		*utək		*uto	utek			brain
131.	*quzan	*hudan	urən	*udan			uran ⁸¹			rain
132.	*Ratus	*ratus			*latus	*otu	otu			hundred
133.	*rəbuŋ	*rəbuŋ		*rəbuŋ	*ləpuŋ	*ombu	ombuk ⁸²	~opu ⁸³	ləbuŋ	bamboo shoot

77. Irregular *q > /k/.

78. None of the sources investigated here attest to the expected final /n/.

79. Possibly fossilized in *murimania* 'to live' as cited in Pampus (1989:83). Mentawai *murimania* could derive from *ma-qudɪp-ni-banua* 'ACT-live-GEN-world' > *m-urɪ-manua*.

80. See Helfrich and Pieters (1891:580), for the attested /l/.

81. See Lenggang (1992:88), for the attested final /n/.

82. None of the sources investigated here attest to the expected final /ŋ/. In opposition to Adriani (1928:36), we take the <g> written in <ogbuk> in Morris (1900:282) to signify a geminate, as in the Malay loan *kelambu* 'mosquito net' > Mentawai <kälabuŋ> (here *kelabuŋ*) beside <tilagbuŋ> (here *tilambuŋ*). This excrescent nasal would be an innovative development shared by PCBI *ombu.83. PSUM *rəbuŋ 'bamboo shoot' > Enggano ~*opu* 'kind of grass'.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
134. *Riaq	*riəh		*rih					liuh	imperata cylindrica
135. *Rumaq	*rumah	umah	*rumah	*luma	*omo	uma	uba		house
136. *Runkuŋ	*runkuŋ		*runkuŋ	*luŋuŋ	*ogu		~uʔu	luŋkuŋ	throat
137. *Rusuk	*rusuk		*rusuk	*lusuʔ	*osu	usu	uku		rib
138. *sakit	*sakit	sakit	*sakit					sakiʔ	sick
139. *sauŋ	*sauŋ		*sauŋ		*səu				umbrella
140. *sawa	*sawa		*sawa	*sawa	*sawa	saba			python
141. *sijəm	*sijəm		*sijəm	*sikəm		sigem ⁸⁴	kiho	sihum	ant
142. *siku	*siku	siku	*siku	*siʔu	*siʔu			siku	elbow
143. *sinaR	*sinar			*sinal	*sino				to shine
144. *sisir	*sisir		*sisir		*sese		~kiki		comb
145. *siwa	*siwa	siwah ⁸⁵	*siwah	*siwa	*siwa	siba			nine
146. *suan	*suan	suən	*suan		*sua				to plant
147. *suluq	*suluq	suluh	*suluh	*sulu	*sulu	sulu ⁸⁶	kudu	suluh	torch
148. *sumanəð	*sumanəð		*sumanəð	*sumanaŋ	*sumanaŋ				human spirit
149. *t<in>aqi	*t<in>ahi					tinay ⁸⁷	kidai		belly

84. See Pampus (1994:117), for the attested final /m/.

85. Irregular final /h/. Innovation shared by PB.

86. PSUM *suluh 'torch' > Mentawai *sulu* 'sun'.

87. See Pampus (1989:79), for this attested form.

	PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
150.	*tabəq	*tabəh		*tabəh	*tabə	*tabə	tabe			fat
151.	*taləs	*taləs			*taləʔ	*talə			talus	taro
152.	*taliḥ	*tali		*tali	*tali	*tali	tali	kadi	tali	thread
153.	*taliŋa	*taliŋa				*taliŋa	taliŋa	kaliha ⁸⁸		ear
154.	*tanəq	*tanəh	tanoh	*tanəh		*tanə			tanuh	dirt
155.	*tanən	*tanən		*tanən		*təŋa		~kaha		hand
156.	*taqi	*tahi	təi	*tai	*tai	*tai	kai	kai	tahi	feces
157.	*tawa	*tawa		*tawa		*tawa				laugh
158.	*təbiq	*təbiḥ			*təpi	*təfi	kopi			broken
159.	*təbuh	*təbu	tu	*təbu	*təpu	*təfu			təbu	sugarcane
160.	*tələn	*tələn		*tələn		*tələ		kodo		to swallow
161.	*təlu	*təlu	tulu	*təlu	*təlu	*təlu	təlu	kolu ⁸⁹	təlu	three
162.	*təjəR	*təjər			*təjəl	*tojə				mangrove
163.	*tinzəR ⁹⁰	*tindər		*tindər		*sindo				to stand
164.	*tiRəm	*tirəm			*tiləm	*siho				oyster
165.	*tirtir	*tirtir		*tirtir				~kiki		to shiver
166.	*tiup	*tiup					kiu ⁹¹		tiup	to blow

88. See Helfrich and Pieters (1891:555), for the attested /l/.

89. See Helfrich and Pieters (1891:593), for the attested /l/.

90. Changed from Blust and Truszel (ongoing) *ti(n)zəg because of ambiguity of Philippine evidence for final *g over final *R and Sumatran evidence for final *R.
 91. PSUM *tiup 'to blow' > Enggano *kiu* 'wind'. See the entry for "air" in the appendix to chapter five of part two in Logan (1855) for an earlier citation for Enggano *tiu*, confirming the *t* > *k* sound change.

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
167. *tubuq	*tubuh		*tubuh					tuwuh	to grow
168. *tuhud	*tuud		*tiud	*tor				tuwut	knee
169. *tupaw	*tupaw		*tupow	*tupaw	*tupə			tupaw	louse
170. *tuqəd	*tuhəd				*tuhe		kuo	tuhut	tree stump
171. *tuRa	*tura		*tura	*tula	*tuo				stinger
172. *tuzuq	*tuduh	turuh	*tuduh	*ituturu	*turu	turu			pointer
173. *umpu	*umpu				*ubu		~upu		ancestor
174. *uRat	*urat	uyət	*urat		*uo	uŋat ⁹²	ua ⁹³	ula?	tendon
175. *wahiR	*wair	wəih		*oil	*we		be	wail	water
176. *wakat	*wakat				*waʔa	bakat			root
177. *walu	*walu	waluh ⁹⁴	*waluh	*olu	*walu	balu		walu	eight
178. *wani	*wani		*wani	*oni	*wani				bee
179. *waRəd ⁹⁵	*warəd		*warəd	*olor	*wao-wao			walut	root
180. *waRi	*wari		*wari					wali	day
181. *zalan	*dalan	ralan	*dalan	*dalan	*lala	lalan ⁹⁶			road
182. *zauq	*dauh		*dauh	*daw	*dəw				far

92. Irregular *r > /ŋ/.

93. PSUM *urat 'tendon' > Enggano ua 'vine'.

94. Irregular final /h/. Innovation shared by PB.

95. Changed from Blust and Trusel (ongoing) *waRəj because of lack of evidence for final *d.

96. See Lenggang (1992:74), for this attested form. PSUM *dalan > Mentawai *lalan* 'swim'. For a similar semantic development, see PMP *panaw 'to go' > Sigulai *fana* 'to walk; to swim (of fish)' as cited in Amery (2019:2).

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Old-E	Nasal	Gloss
183.	*bagas	waas	*bagas	*baka?	*baxa	bag ⁹⁷			inside
184.	*baruŋ		*baruŋ	*baluŋ	*tum-bao				heron
185.	*bəŋŋa		*bəŋŋa		*boŋŋa				maggot
186.	*gahul	awal	*gahul	*kaul					banana
187.	*hasar	asey	*asar					hasal	nest
188.	*ijur			*ijol	*ijo			ijul	mucus
189.	*lanəkəp		*lanəkəp		*lagə				cover
190.	*sada	sara	*sada	*sara	*sara	sara			one
191.	*səraw					sow		səlaw	to cry
192.	*sua[ɾ]			*sol				sual	comb
193.	*taŋkaw		*taŋkaw		*tagə	taŋkow ⁹⁸			to steal
194.	*tarub		*tarub	*talup					roof
195.	*təruh	tuyuh	*təruh	*təlu	*tu				low
196.	*tigər		*tigər	*tikəl	*sixo				midday
197.	*timbər		*timbər	*timbəl	*simbo	timbo ⁹⁹	~ipo		smoke

97. See Pampus (1989:799), for this attested form.

98. See Modigliani (1898:291), for the attested medial /ŋk/ sequence.

99. See Pampus (1994:115), for the attested /mb/ sequence.

APPENDIX F. SIMPLIFIED OVERVIEW OF SOUND CORRESPONDENCES BETWEEN PMP, PSUM, AND PSUM SUBGROUPS

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Pre-E	Nasal
*b	*b	*p / # *w / else	*b	*p / ə *p / # *b / else	*b	*p / # *b / else	*p	*b / ə *w / else
*d	*d	*t / # *r / else	*d	*r / V_V *r / # *d / else	*r / V_V *l / #...l *d / else	*l / #...l *d / else	*d	*d / ə *t / # *ɣ / else
*j *g	*g	*k / # ∅ / else	*g	*k	*x / V_V *g / else	*w / # *g / else	*h	*g / ə *k / # *h / else
*p	*p	*p	*p	*p / ə *h / else	*f	*p	*p	*ʔ / # *p / else
*t	*t	*t	*t	*t	*s / i *t	*t	*k	*ʔ / #
*c								
*k	*k	*k	*k	∅ / # *ʔ / else	∅ / # *ʔ / else	*k	*ʔ	*ʔ / # *k / else
*q	*h	*h	*h	∅	*h	∅	∅	*h
*s	*s	*s	*s	*s	*s	*s	*k	*s
*h	∅							

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Pre-E	Nasal
*R *r	*r	*y	*r	*l	∅	∅	∅	*l
*l	*l	*l	*l	*l	*l	*l	*l	*l
*m	*m	*m	*m	*m	*m	*m	*b	*m
*n	*n	*n	*n	*n	*n	*n	*d	*n
*ñ								
*ŋ	*ŋ	*ŋ	*ŋ	*ŋ	*ŋ	*ŋ	*h	*ŋ
*mb	*mb	*mb	*mb	*mb	*mb	*mb	*p / ə̄ *~p / else	*mb
*mp	*mp	*mp	*mp	*mp	*b	*mp	*p / ə̄ *~p / else	*mp
*nd	*nd	*nd	*nd	*nd	*nd	*nd	*d / ə̄ *~d / else	*nd
*nt	*nt	*nt	*nt	*nt	*d	*nt	*k / ə̄ *~k / else	*nt
*ŋg	*ŋg	*ŋg	*ŋg	*ŋk	*g	*ŋg	*h / ə̄ *~h / else	*ŋg
*ŋk	*ŋk	*ŋk	*ŋk	*ŋk	*g	*ŋk	*ʔ / ə̄ *~ʔ / else	*ŋk
*w	*w	*w	*w	*w	*w	*b	*b	*w
*y	*y	*y	*y	*y	*y	*j		*y

PMP	PSUM	Gayo	PB	PNBI	PCBI	Mentawai	Pre-E	Nasal
*a	*a	*ə / _# *a / else	*a	*a	*a	*a	*a	*o / _# *a / else
*i	*i	*i	*i	*i	*i	*i	*i	*i
*u	*u	*u	*u	*u	*u	*u	*u	*u
*ə	*ə	*ə	*ə	*ə	*ə	*e	*o	*u / _C# *ə / else
*ay	*ay	*ay	*ey	*ay	*e	*ay	*e	*ay
*aw	*aw	*aw	*ow	*aw	*ə	*ew		*aw
*uy	*uy	*i	*uy	*oy	*i	*uy		*uy
*iw	*iw		*i	*ew	*i	*iw		

REFERENCES

- Adelaar, K. Alexander. 1981. Reconstruction of Proto-Batak phonology. In *Historical linguistics in Indonesia*, ed. by Robert A. Blust, 1–20. Jakarta: Atma Jaya University.
- . 2005. Malayo-Sumbawan. *Oceanic Linguistics* 44(2): 357–88.
- Adriani, Nicolaus. 1928. Spraakkunstige schets van de taal der Mentawai-eilanden. *Bijdragen tot de Taal-, Land- en Volkenkunde* 81(1): 1–117.
- Amery, Robert. 2019. Language is more than communication: Why we should maintain the mother tongue and promote linguistic diversity. In *Proceedings of the 2nd English Education International Conference (EEIC) in conjunction with the 9th Annual International Conference (AIC)*, ed. by Yunisrina Qismullah Yusuf, 1–5. Banda Aceh: Universitas Syah Kuala Press.
- Amery, Robert, and Zulfadli A. Aziz. 2019. Enumeration and classifiers in Pulau Simeulue / Pulau Banyak languages, Aceh. In *The proceedings of the 2019 GLOCAL conference in Asia, Asian linguistic anthropology*, ed. by Asmah H. Omar, 430–39. London: The GLOCAL.
- Anderbeck, Karl. 2007. An initial reconstruction of Proto-Lampungic: Phonology and basic vocabulary. *Studies in Philippines Languages and Cultures* 16:41–165.
- Anderbeck, Karl R., and Herdian Apriliani. 2013. *The improbable language: Survey report on the Nasal language of Bengkulu, Sumatra* (Electronic Survey Report 2013-012). Dallas, TX: SIL International.
- Blust, Robert. 1984. On the history of Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140(4): 422–50.
- . 1991. The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30(2): 73–129.
- . 1994. The Austronesian settlement of mainland Southeast Asia. In *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*, ed. by Karen L. Adams and Thomas J. Hudak, 25–83. Tempe, AZ: Program for Southeast Asian Studies, Arizona State University.
- . 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36(1): 149–79.
- . 1998. In defense of Dempwolff: Austronesian diphthongs once again. *Oceanic Linguistics* 37(2): 354–62.
- . 2004. *t to k: An Austronesian sound change revisited. *Oceanic Linguistics* 43(2): 365–410.
- . 2013. *The Austronesian languages*. Canberra: The Australian National University.
- Blust, Robert A., and Stephen Trussel. Ongoing. The Austronesian Comparative Dictionary. www.trussel2.com/ACD (accessed March 5, 2024).
- Brown, Lea. 2001. *A grammar of Nias Selatan*. Sydney: The University of Sydney.
- Candrasari, Ratri, and Khalsiah. 2018. *Kamus Indonesia – Leukon*. Lhokseumawe, Aceh: Sefa Bumi Persada.
- Catford, John Cunnison. 1988. Notes on the phonetics of Nias. In *Studies in Austronesian linguistics*, ed. by Richard McGrimes, 151–72. Athens, OH: Center for Southeast Asian Studies.
- Clynes, Adrian. 1997. On the Proto-Austronesian “diphthongs.” *Oceanic Linguistics* 36(2): 347–61.
- Coady, James, and Richard McGinn. 1982. On the so-called implosive nasals of Rejang. In *GAVA: Studies in Austronesian Languages and Cultures*, ed. by Rainer Carle, 437–49. Berlin: Dieter Reimer Verlag.
- Cohn, Abigail C., and Anastasia K. Riehl. 2016. Are there post-stopped nasals in Austronesian? *NUSA* 60:29–57.
- Court, Christopher. 1970. Nasal harmony and some Indonesian sound laws. In *Pacific linguistic studies in honour of Arthur Capell* (Pacific Linguistics C-13), ed. by S. A. Wurm and D. C. Laycock, 203–17. Canberra: The Australian National University.

- Durie, Mark. 1985. *A Grammar of Acehnese on the Basis of a Dialect of North Aceh*. Dordrecht: Foris Publications.
- Eades, Domenyk. 2005. *A grammar of Gayo: A language of Aceh, Sumatra* (Pacific Linguistics 567). Canberra: The Australian National University.
- Edwards, Owen. 2015. The position of Enggano within Austronesian. *Oceanic Linguistics* 54(1): 54–109.
- Eggink, H. J. 1936. *Angkola- en Mandailing-Bataksch – Nederlandsch woordenboek*. Bandung: A. C. NIX & Co.
- van Engelenhoven, Aone. 2009. The position of Makuva among the Austronesian languages of East Timor and Southwest Maluku. In *Austronesian historical linguistics and culture history: A festschrift for Bob Blust*, ed. by Alexander Adelaar and Andrew Pawley, 425–42. Canberra: The Australian National University.
- Goodenough, Ward H. 1997. Proto-Kimbe: A new analysis. *Oceanic Linguistics* 36(2): 247–311.
- Gordon, Matthew K. 2016. *Phonological typology*. Oxford: Oxford University Press.
- Hazeu, Godard Arend Johannes. 1907. *Gajösch-nederlandsch woordeboek*. Jakarta: Landsdrukkerij.
- Helfrich, Oscar Louis. 1891. Proeve van een Lampongsch-Hollandsche woordenlijst bepaaldelijk voor het dialect van Kroë. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen* 45.
- Helfrich, Oscar Louis, and J. A. J. C. Pieters. 1891. Proeve van eene maleisch-nederlandsch-enganeesch woordenlijst. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 34:539–604.
- Kähler, Hans. 1937. Untersuchungen über die Laut-, Wort- und Satzlehre des Nias. *Zeitschrift für Eingeborensprachen* 27:91–128, 212–22, 261–88.
- . 1955. *Die Sichule-Sprache auf der Insel Simalur an der Westküste von Sumatra* (Afrika und Übersee 27). Berlin: Dietrich Reimer.
- . 1959. *Vergleichendes Wörterverzeichnis der Sichule-Sprache auf der Insel Simalur an der Westküste von Sumatra*. Berlin: Dietrich Reimer.
- . 1961. *Simalur-Deutsches Wörterbuch mit Deutsch-Simaluresischem Wörterverzeichnis*. Berlin: Dietrich Reimer.
- . 1987. *Enggano-Deutsches Wörterbuch* (Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg 14). Berlin: Dietrich Reimer.
- . 2020 (1940). Grammatical summary of Enggano (trans. by Barnaby Burleigh and Mary Dalrymple). *Zeitschrift für Eingeborensprachen* 30:81–117, 182–210, 296–320.
- Lafeber, Abraham. 1922. *Vergelijkende klankleer van het Niasisch*. The Hague: Hadi Poestaka.
- Laskowske, Thomas V. 2006. *The Seko languages of South Sulawesi: A reconstruction*. Dallas, TX: SIL International.
- Lenggang, Zainuddin H. R. 1992. *Geografi dialek bahasa Mentawai*. Padang: Pusat Pembinaan dan Pengembangan Bahasa.
- Logan, J. R. 1855. *The journal of the Indian Archipelago and Eastern Asia*, vol. 9. Singapore: Jacob Baptist.
- Lubis, Tasnim, and Nicholas Williams. 2019. Preliminary documentation of Leukon language. Endangered Languages Archive. <http://hdl.handle.net/2196/00-0000-0000-0014-134E-C> (accessed March 5, 2024).
- Manik, Tindi Radja. 2002. *Kamus Pakpak-Indonesia*. Medan: Penerbit Bina Media.
- Marsden, William. 1811. *The history of Sumatra containing an account of the government, laws, customs, and manners of the native inhabitants with a description of the natural productions and a relation of the ancient political state of that island*. London: John M'Creery.

- McDonnell, Bradley. 2017a. Prospects of a community based language project in the Nasal speech community. In *Communities in control: Learning tools and strategies for multilingual endangered language communities*, ed. by Nicholas Ostler, Vera Ferreira, and Chris Moseley, 64–69. Hungerford, England: Foundation for Endangered Languages.
- . 2017b. Documentation of Nasal: An overlooked Malayo-Polynesian isolate of southwest Sumatra. Endangered Languages Archive. <http://hdl.handle.net/2196/00-0000-0000-0010-798B-E> (accessed January 15, 2024).
- (collector). 2019. Documentation of the multilingual linguistic practices of the Nasal speech community. Collection BJM02 at catalog.paradisec.org.au (accessed March 5, 2024). doi: [10.26278/5f46870d43f29](https://doi.org/10.26278/5f46870d43f29).
- . Forthcoming. Documenting multilingualism in Southwest Sumatra. In *Indonesian Languages and Linguistics: State of the Field*, ed. by Maya Ravindranath Abtahian, Abigail C. Cohn, Thomas J. Connors, Joseph J. Errington, and Yanti. Ithaca: SEAP Publications, Cornell University Press.
- McDonnell, Bradley, and Christina L. Truong. 2024. The non-Malayic languages of Sumatra and the Barrier Islands. In *The Oxford guide to the Malayo-Polynesian languages of South East Asia*, ed. by Alexander Adelaar and Antoinette Schapper. Oxford: Oxford University Press.
- McGinn, Richard. 1997. Some irregular reflexes of Proto-Malayo-Polynesian vowels in the Rejang language of Sumatra. *Diachronica* 14(1): 67–107.
- . 1999. The position of the Rejang language of Sumatra in relation to Malay and the ‘ablaut’ languages of Northwest Borneo. In *Selected papers from the Eight International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 205–26. Taipei: Institute of Linguistics Preparatory Office, Academia Sinica.
- . 2000. Where did the Rejangs come from? In *Proceedings of the Tenth Annual Conference of the Southeast Asian Linguistics Society*, ed. by Marlys Macken, 247–62. Tempe, AZ: Arizona State University.
- . 2003. Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang. In *Issues in Austronesian historical phonology*, ed. by John Lynch, 37–64. Canberra: The Australian National University.
- . 2005. What the Rawas dialect reveals about the linguistic history of Rejang. *Oceanic Linguistics* 44(1): 12–64.
- . 2009. Out-of-Borneo subgrouping hypothesis for Rejang: Re-weighting the evidence. In *Austronesian historical linguistics and culture history: A festschrift for Bob Blust*, ed. by Alexander Adelaar and Andrew Pawley, 397–410. Canberra: The Australian National University.
- Mead, David. 1998. Proto-Bungku-Tolaki: Reconstruction of its phonology and aspects of its morphosyntax. Doctoral diss., Rice University.
- Mills, Roger F. 1991. Tanimbar-Kei: An Eastern-Indonesian subgroup. In *Currents in Pacific linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace* (Pacific Linguistics C-117), ed. by Robert Blust, 241–63. Canberra: The Australian National University.
- Modigliani, Elio. 1894. *L’Isola delle donne: Viaggio ad Engano*. Milan: Ulrico Hoepli.
- . 1898. Materiale per lo studio dell’Isola Sipòra. *Bolletino della Società Geografica Italiana* 11:256–99.
- Morris, Max. 1900. *Die Mentawai-Sprache*. Berlin: Conrad Skopnik.
- Neumann, Joshua H. 1951. *Karo-bataks nederlands woordenboek*. Medan: VAREKAMP & Co.
- Nothofer, Bernd. 1986. The Barrier Island languages in the Austronesian language family. In *FOCAL II: Papers from the fourth International Conference on Austronesian Linguistics* (Pacific Linguistics C-94), ed. by Paul Geraghty, Lois Carrington, and Stephen A. Wurm, 87–109. Canberra: The Australian National University.

- Pampus, Karl-Heinz. 1989. Zur dialektgeographischen Gliederung des Mentawai-Archipels. In *Mentawai: Identität im Wandel auf Indonesischen Außeninseln*, ed. by Wilfred Wagner, 61–102. Bremen: Übersee-Museum.
- . 1994. Sabirut~Taburit ‘Siberut’: Die Lautgesetze der Simalegi-Sprache im Mentawai-Archipel (Westsumatra). *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 144(1): 114–29.
- Pätsch, Gertrud. 1978. Das Nias im historischen Vergleich. *Language Typology and Universals* 31:58–72.
- Reid, Lawrence A. 1974. The Central Cordilleran subgroup of Philippine languages. *Oceanic Linguistics* 13(1/2): 511–60.
- von Rosenberg, Hermann. 1855a. Beschrijving van Engano en van deszelfs bewoners. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 3:370–86.
- . 1855b. Geografische en ethnografische beschrijving van het district Singkel, de landen liggende langs de Simpang Kanan en de Banjak-eilanden, benevens eene korte aantekening nopens de Simpang Kiri. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 3:397–476.
- Shorto, Harry Leonard. 1976. Gayo consonant correspondences. In *Southeast Asian linguistics studies*, vol. 2 (Pacific Linguistics C-42), ed. by Đãng Liêm Nguyễn, 199–218. Canberra: The Australian National University.
- Smith, Alexander D. 2017. The Western-Malayo-Polynesian problem. *Oceanic Linguistics* 56(2): 435–90.
- . 2020. Nasalization in Enggano historical phonology. *Oceanic Linguistics* 59(1/2): 347–65.
- Stokhof, Wim A. L. 1987. *Holle lists: Vocabularies in languages of Indonesia*, vol. 10/2, *Southern Sumatra* (Materials in Languages of Indonesia 35). Canberra: The Australian National University.
- Sundermann, Heinrich D. 1905. *Niassisch-Deutsches Wörterbuch*. Moers: J. W. Spaarmann.
- Syarfina, Tengku, Amrin Saragih, Amran Purba, Zufri Hidayat, Anharuddin Hutasuhut, Chairani Nasution, Sri Asrianti, and Juliana. 2015. *Kamus Bahasa Simalungun – Indonesia*. 2nd edition. Medan: Balai Bahasa Provinsi Sumatera Utara.
- Thomas, J. W., and E. A. Taylor Weber. 1887. *Niasch-maleisch-nederlandsch woordenboek*. Jakarta: Landsdrukkerij.
- Tim Kamus Balai Bahasa Banda Aceh. 2013. *Kamus Bahasa Haloban*. Banda Aceh: Balai Bahasa Banda Aceh.
- van der Tuuk, Herman Neubronner. 1861. *Bataksch-nederduitsch woordenboek*. Amsterdam: Frederik Muller.
- . 1971 (1864–1867). *A grammar of Toba Batak* (Pacific Linguistics Translation Series 13; Jeune Scott-Kemball, Trans.). The Hague: Koninklijk Instituut voor de Taal-, Land- en Volkenkunde.
- Waland, Jan. 1864. Het eiland Engano. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 14:93–124, 330–339.
- Warneck, Johannes G. 1977. *Toba-Batak – Deutsches Wörterbuch*. The Hague: Koninklijk Instituut voor Taal-, Land- en Volkenkunde.
- Yoder, Brendon E. 2014. *Phonological and phonetics aspects of Enggano vowels (SIL eBook 62)*. Dallas, TX: SIL International.
- Zobel, Erik. 2021. The phonological history of Nias. Paper presented at *Towards the next 40 years of Southeast Asian Studies in Frankfurt: Symposium in honour of Bernd Nothofer*, Frankfurt, 19 December.