Author/s:
Adelaar, A

Title:
Historical linguistics of the languages of Sumatra, Java, the Lesser Sunda Islands, and Moken-Moklen

Date:
2024

Citation:
Adelaar, A. (2024). Historical linguistics of the languages of Sumatra, Java, the Lesser Sunda Islands, and Moken-Moklen. Adelaar, A (Ed.). Schapper, A (Ed.). The Oxford Guide to the Malayo-Polynesian Languages of Southeast Asia, (1), pp.127-142. Oxford University Press.

Persistent Link:
https://hdl.handle.net/11343/322027

Chapter 10. Historical linguistics of the languages of Sumatra, Java, the Lesser Sunda Islands and Moken-Moklen

Alexander Adelaar

Sinophon Project, ~~Jan~~ Palacký University, Olomouc
Asia Institute, University of Melbourne

10.1. Introduction

This chapter is about the linguistic history of the languages of Sumatra, Java, the Lesser Sunda Islands and Moken-Moklen. It gives an overview of the various subgrouping hypotheses that have been proposed for the languages in this southwestern MPSEA fringe area, and it discusses the principles on which these hypotheses have been built. It does not include the Malayic languages, which are treated in Anderbeck (chapter 9, this volume) or Acehnese (North Sumatra), which is closely related to the Chamic languages and will be discussed in Brunelle (chapter 11, this volume). The languages will be discussed in a basically West - East order, and, as far as possible, in the context of larger genetic configurations to which they may belong.

Historical research on these languages has concentrated on phonology and lexicon, except for Edwards (2015) who also deals with grammatical developments in Enggano.

The chapter is mostly concerned with subgrouping proposals that have been made over the last four decades. Several smaller language configurations and isolates are discussed. They are detailed here for the sake of clarity:

Moken-Moklen are two closely related languages spoken in the Mergui Archipelago (on the border between Southwest Myanmar and Northwest Thailand). Their position within the Malayo-Polynesian subgroup remains uncertain (see also Smith on Malayo-Polynesian, chapter 2, this volume).

Non-Malayic languages of Sumatra include the Barrier Island languages, which supposedly form a subgroup, although the membership of two of them (Enggano, and to a less extent Mentawai) is problematic; the Batak languages, a close-knit and transparent subgroup in northern Sumatra; and Lampungic languages, a dialect chain in southern Sumatra. They also include Gayo, Nasal and Rejang, three Sumatran languages resisting subgrouping (see also Truong and McDonnell, chapter 28, including the map in Figure 28.1 ~~this volume~~).

Non-Malayic languages of Java (Javanese, Sundanese, Madurese). Together with Malay they were treated as a genetic subgroup by Nothofer (1975). However, they are no longer considered to be closely related (see also Vander Klok, chapter 31, this volume, including the map in Figure 31.1).

Bali-Sasak-Sumbawa or “BSS” is a coherent subgroup of languages spoken in respectively Bali, Lombok and West Sumbawa (see Shiohara and Arka, chapter 32, this volume, and the maps in Figures 31.2-3-4).

The following sections in this chapter deal in more detail with the genetic affiliations of the above languages. Section 10.2 explains the major subgroups and subgrouping devices that are currently used. Section 10.3 discusses Moken-Moklen; §10.4, the Barrier Island languages, Batak languages and Enggano, and §10.5, Gayo, Rejang, Nasal and Lampungic (that is, languages of Sumatra other than Malayic and Batak ones and Acehnese); §10.6, the languages of Java; and §10.7, the languages of Bali, Lombok and West Sumbawa. Section

10.8 is an appraisal of the use of *tuzuq 'seven' and of PMP *j as subgrouping evidence. Concluding remarks follow in §10.9.

10.2. Major subgroups and subgrouping devices

Robert Blust (2010b) and Alexander Smith (2017a) are currently the main protagonists in the classification of the languages in question. For a better understanding of the complicated issues that currently play a role in this classification, it is necessary from the outset to take notice of the subgroups that they propose and the diagnostic devices that they have been using.

All languages investigated in this chapter belong to Blust's Western Indonesian (henceforth WIn) subgroup. It is based on lexical evidence, including botanical and zoological terms. This evidence was initially rather limited but it has recently been augmented with more lexical evidence by Smith (2017a). Blust classifies a large part of the WIn languages into a further (smaller) group called the "Greater North Borneo" subgroup (henceforth GNB), which includes the languages of most of Borneo as well as the Malayic languages on the southern coasts of the South China Sea (see Anderbeck, chapter 9, this volume), the Chamic languages in Vietnam and Kampuchea (see Brunelle, chapter 11, this volume), the Moken and Moklen (which are spoken on the Mergui Archipelago and along the western coasts of Burma and Thailand), Sundanese (in Java) and Rejang (in Sumatra) (see also Smith on Malayo-Polynesian, chapter 2, this volume). Blust's main diagnostic device for establishing the GNB subgroup is that the languages in it have replaced *pitu, the original Austronesian numeral for 'seven', with *tuzuq, a word which originally meant 'to point with the index'. In contrast, other languages in the WIn group still have a reflex of *pitu for 'seven' (and by and large also kept the original meaning of *tuzuq).

Smith (2017b) rejects the WIn subgroup as defined by Blust and splits it into three primary Malayo-Polynesian subgroups, to wit Moken-Moklen, Sumatran and WIn "proper" (the latter maintaining GNB as a sub-division within it). Smith's main criterion for this further split is that the member languages in the resulting three subgroups each have their own reflex of PMP *j. Many Sumatran languages reflect it as *g* (or at least, as a consonant that has developed from an intermediate *g), Moken and Moklen reflect it as *y*, and other WIn languages reflect it as *d* (or *r*, which is a development thereof).

Other classifications that have been proposed in the last two decades are Zobel (2002) and Adelaar (2005c). Zobel (2002) distinguishes a Nuclear Malayo-Polynesian subgroup which encompasses Indonesian languages on the southwestern fringes of MPSEA, languages of central and southern Sulawesi, Chamorro and Palauan, and the CEMP languages. His classification is based on morphosyntactic evidence (see Smith on Malayo-Polynesian historical linguistics, this volume). Based on phonological and lexical evidence, Adelaar's (2005c) "Malayo-Sumbawan" subgroup excludes Javanese and puts BSS on a par with Malayic and Aceh-Chamic in one division, and Sundanese and Madurese in two other ones (see further §10.8).

10.3 **Moken and Moklen**

The westernmost languages, Moken and Moklen, have a rather unique phonological history.[1] Larish (1999) presents a subgrouping hypothesis in which Proto-Moken-Moklen is closely related to Malayo-Chamic. His Proto-Moken-Moklen–Aceh-Chamic-Malayic subgroup consists of two branches, Moken-Moklen and Aceh-Chamic-Malayic. Thurgood (1999:58) opposes such a relationship, pointing out the unusual Moken-Moklen change from *q to *k* and the different ways final high vowels became diphthongised (PMP *-i > Proto Malayic *-i, Proto Chamic *-ɛy; PMP *-u > Proto Malayic *-u, Proto Chamic *-ɔw; whereas in Moken-Moklen both vowels have merged to *-oi*. Moken-Moklen furthermore merged PMP *R and *l whereas Malayic and Chamic reflect these as respectively *r* and *l*.

<<Table 10.1 Comparative wordlist of PMP, Malayic, Chamic and Moken. Please put somewhere here>>

Blust (2010b) includes Moken-Moklen in the GNB group because they have a reflex of *tuzuq for 'seven'. Smith (2017b) contests this, considering the pair an isolate within the Malayo-Polynesian language group. He refers to Blust's own observation that *tuzuq reflexes meaning 'seven' are also found in some languages which cannot belong to GNB (such as Makassarese). He also points out that Moken-Moklen as well as some of the GNB languages in Borneo itself show irregularities in their reflexes of *tuzuq. Furthermore, in contrast to many GNB languages, Moken and Moklen do not merge PMP *j and *d but distinguish them as respectively *y* and *d* (compare PMP *qudip 'alive' > *kodip* and *daqan 'branch' > *dakan* vs. *qijuŋ > *yoŋ* and *qajəŋ 'charcoal' > *kayaŋ*). This is a reason for him not to include these languages in GNB but to afford them a separate branch status within Malayo-Polynesian.

10.4 Barrier Island languages, Batak languages and Enggano

The Barrier Island languages along Sumatra's west coast are Simeulue, Sikhule, Nias, Mentawai, and Enggano. They are often assumed to form a genetic subgroup (see Nothofer 1986), although the inclusion of Enggano is very uncertain as this language differs dramatically from other MPSEA languages. The major Batak languages are the Southern Batak ones Toba, Angkola, Mandailing and Simalungun, and the Northern Batak ones, Karo, Dairi-Pakpak, and Alas. Gayo is sometimes assumed to be closely related to the Batak languages, but the relationship has never been demonstrated and is unlikely (see below).[2] At a higher (more distant) level, the Barrier Island and Batak languages are also considered to form a genetic group on account of the large amount of lexical innovations that they exclusively share and on the assumption that they both reflect PMP *j as *g* or, in some cases, they reflect it as a consonant which must have been *g at an earlier stage (Nothofer 1986). In Simeulue, Sikhule and Nias, PMP *j became *x*; in Mentawai, *g*; and in Enggano, *h*.

10.4.1 Barrier Island languages and Enggano

[1] See Larish (1999, 2011) for information about differences between these closely related languages.

[2] McDonnell and Truong (chapter 28, this volume) already pointed to the lack of evidence for an inclusion of Gayo in the Barrier Island – Batak subgroup.

The comparative wordlist in Table 10.1[3] brings out the phonological differences between the assumed members of the Barrier Island subgroup. In Edwards' (2015:83-84) classification, Simeulue, Sikhule and Nias form the core of this subgroup, whereas Mentawai is a less integrated member, and Enggano shows a totally different phonological history from the other languages, even if it shares with Mentawai a development from *w to b. (This is somewhat different from Nothofer (1986:105), in which Sikhule, Nias and Mentawai are grouped closely together whereas Simeulue is somewhat more remote, and an even more remote Enggano is only provisionally included).

Table 10.2. "Comparative wordlist of PMP, the Barrier Island languages, and Proto Batak."
<<to be inserted somewhere here>>

Edwards (2015) is an investigation into the phonological, lexical and morphosyntactic history of Enggano. The author provides a comprehensive body of evidence to show that it is a member of the Austronesian language family, rejecting the position of earlier scholars (Capell [1982]; Blench [2014]) who disputed this. His main arguments are that in spite of the unusual sound changes that Enggano has undergone since PMP (such as *w > *b*; *t > *k*; *s > k; *m > *b*; *n > *d*; *ŋ, j > *h*), these changes are often very regular; furthermore, Enggano also appears to have retained a large part of PMP verb morphology and pronouns. Compare the following genitive suffixes: *-ʔVu* '1SG.GEN' (*=(ŋ)ku); *-bu* '2SG.GEN' (< *=mu); *-dia* '3SG.GEN' (< *-nia); *-dai* '1PL.EXCL.GEN' (< *n-ami); *-ka* '1DU.INCL.GEN' (< *=ta); *kaʔa* '1PL.INCL.GEN' (< *=ta + plural extender); *-d(i)u* '2PL.GEN' (< *=miu); '3PL.GEN' (< *-da).

Edwards is doubtful about a common history between Enggano and other Barrier Island languages. He found the following three phonological changes that seem to suggest such a history:

(1) PMP *j > Enggano h, Simeulue, Sikule, Nias x (<-> Mentawai g, ɣ; Proto Batak *g[4]);
PMP *k > Enggano ʔ, Sikule, Nias ʔ, Ø, Simeulue k, ʔ, Ø (<-> Mentawai k, ʔ; Proto Batak *k);
PMP *w > Enggano b, Mentawai b (<-> Sikule, Nias, Simeulue w, Ø; Proto Batak *w, Ø);

He does not consider them a basis for including Enggano in a Barrier Island – Batak subgroup. In this language, not only *j but also *ŋ became *h*. Therefore, Edwards (2015:83-84) assumes that at an earlier point in Enggano phonological history *j had merged with *ŋ to *ŋ before the latter finally took its current shape *h*, and he does not think that this *ŋ could eventually have led to Simeulue, Sikule and Nias *x* or to Mentawai *g* or *ɣ* (or Proto Batak *g). He therefore finds it doubtful that the *j reflexes in all these languages went through a partly common history: a joint development from *j to *x is not obvious because the change is natural enough to have happened independently in these languages, whether *j was a velar stop (Wolff 1988:131-132), a velar fricative (Ross 1992:36) or a palatalised voiced velar stop (Blust 2013a:554). And even if this presumed joint development would be

[3] The information in this list is taken from Nothofer (1986:100-102), Edwards (2015), Kähler (1963:xxi-xxii), Blust and Trussel (ongoing).
[4] Edwards gives Toba Batak cognates but these Proto Batak phonemes (based on Adelaar 1981) are uncontroversial and phonologically more revealing.

taken as evidence for including Enggano in the Barrier Island group, it would clash with the inclusion of Batak and Mentawai in the same group because a change from *x to (Batak) *g* or (Mentawai) *ɣ* is phonetically unusual (Edwards 2015:83).

As to the other sound changes, the one from *k to *ʔ* or *Ø* is phonologically too general to have diagnostic value, and while *w > *b* is more meaningful, it should also be rejected as subgrouping evidence because it is found in only two Barrier Island languages (Enggano and Mentawai). The change is moreover not entirely unique and could be due to areal contact as it also occurs in Rejang and in Lampung in Sumatra (Edwards 2015:84; see also §5.3 for Rejang and Anderbeck 2007b for Lampung).

After a careful evaluation of the phonological, morphological and lexical evidence (including evidence from allegedly shared phonological irregularities), Edwards decides against subgrouping Enggano with the Barrier Island languages and to give the language an independent isolate status within Malayo-Polynesian. He is even doubtful that the other Barrier Island languages form a subgroup: Simeulue, Sikule and Nias may together qualify as one, but Mentawai would not integrate well in it (Edwards 2015:94).

The pathway taken from PMP *j to its current reflexes in Barrier Island and Batak languages also became an important detail in Smith's (2017b) classification, as the history of *j would be the cornerstone, two years later, of his distinction between a West Indonesian and a Sumatran subgroup. He does include Enggano in the Barrier Island subgroup, arguing that PMP *j first merged with *g to *g; later on it also merged with *ŋ to *g. The latter in turn evolved into *ɣ*, before it finally became modern Enggano *h*. In other Barrier Island – Batak languages, *j first merged with *g to *g in Batak languages or to *ɣ* and *g* in Mentawai, and it ended up as *x* in Nias, Sikule and Simeulue.

Smith brings the Enggano change from (PMP *ŋ, *j >) *ŋ > to *g > *h* in line with the fact that in Enggano, PMP *m and *n were also denasalised (they became *b* and *d* respectively, although they maintain nasal allophones). So the change from *ŋ to *g was part of a more general denasalisation process which affected all earlier nasals for which there is solid historical evidence.[5] In this scenario, the stage between PMP *j and Enggano *h* was not via *ŋ, which would have been phonologically implausible, but via *g, which solves that problem.

The developments of PMP *j in Enggano can be staged as follows:

(2) Stage 1: PMP *j > Proto Barrier Island *g

Stage 2: Proto Barrier Island *g merges with *ŋ > *g > *h*

In line with Smith's analysis, phonological developments in languages other than Enggano would be as follows:

(3) Stage 1: PMP *j > Proto Barrier Island *g

Stage 2: *g > Enggano h
*g > Proto Batak *g > Simalungun Batak *g*, N. Batak *g*, *-ŋ*, S. Batak *g*, *-k*
*g > Mentawai *g*, *ɣ*

[5] Incidentally, Enggano has a palatal nasal phoneme, but the status of palatal nasals is very uncertain both in Enggano (Edwards 2015:62) and in Malayo-Polynesian phonological history.

*g > *ɣ > Nias, Sikule and Simeulue *x*

In this way, Smith tries to demonstrate that Enggano and all Barrier Island and Batak languages went through a common stage in which PMP *j became *g, and to provide critical evidence for their inclusion in the Barrier Island – Batak subgroup. Smith's argument is a forceful one, but there remains the question as to whether the evidence of this one phonological development weighs up against the impressive body of evidence adduced by Edwards (2015) from several levels of linguistics and even from other disciplines.

10.4.2 Batak languages

Compared to the Barrier Island languages, the Batak languages form a transparent and close-knit subgroup, and the reconstruction of Proto Batak is a straightforward affair (Adelaar 1981). Of its members, Toba, Mandailing, Angkola, and Simalungun constitute a Southern branch, and Karo, Dairi-Pakpak and Alas, a Northern branch.

Phonological features that are typical for the various Batak languages are as follows:

(4)	PMP *ə >	N. Batak *ə*, S. Batak *o*
	PMP *-aw >	Simalungun *-ow*, other Batak *-o*
	PMP *-ay >	Simalungun *-ey*, other Batak *-e*
	PMP *-uy >	Simalungun *-uy*, other Batak *-e*, *-i*
	PMP *w-, *-w- >	N. Batak *u-*, *Ø-*, *-w-*, Southern Batak Ø
	PMP *q >	N. Batak *h, Simalungun Batak *Ø*, *-h*, other S. Batak Ø
	PMP non-final *k >	N. Batak *k*, S. Batak *h*, *-k*
	PMP *-j- >	all Batak languages *-g-*
	PMP *-j >	N. Batak *ŋ*, Simalungun *g*, other S. Batak *k*
	PMP no stress	N. Batak lacks stress, S. Batak has distinctive stress

<<Table 10.3. Genealogical tree of Batak languages to appear somewhere around here>>

Simalungun Batak shares linguistic features with both Northern and Southern Batak, and due to its seemingly intermediate position, Voorhoeve (1955) gave it the status of a third division within the Batak group. However, Adelaar (1981) classifies it as an early split-off within Southern Batak because it shares important innovative phonological developments with the Southern Batak varieties such as the changes from Proto Batak/PMP *k- and *-k- to *h*, from Proto Batak/PMP *ə to *o,* and from Proto Batak/PMP *c to *s*, as well as the development of distinctive stress. In contrast, commonalities between Simalungun and Northern Batak are often the result of borrowing or are retentions from Proto Batak (e.g. the maintenance of *-h (< PMP *q). Also significant is that Simalungun maintained Proto Batak final voiced stops, which is relatively more in alignment with the corresponding Southern Batak final voiceless stops than with the corresponding Northern Batak final nasal. Examples:

(5) P. Batak *kalak 'human being' >Simalungun, other S.Batak *halak*, N. Batak*kalak*;
P. Batak *aku 'I, me' >Simalungun, other S.Batak *ahu*, N. Batak *aku*;
P. Batak *kərbow 'buffalo' > S. Batak *horbo*, Simalungun *horbow*, N. Batak *kərbo*;

P. Batak *pərəh 'to squeeze out' > S. Batak *poro*, Simalungun *poroh*, N. Batak *pərəh*;
P. Batak *pucuk 'sprout, young leaf' > Simalungun, other S.Batak *pusuk*, N. Batak *pucuk*;
P. Batak *abab 'ashes' > S. Batak *abap*, Simalungun *abab*, N. Batak *abam*;
P. Batak *dələg 'mountain' > S. Batak *dolok*, Simalungun *dolog*, N. Batak *dələŋ*;
P. Batak *sərəd 'sting' > S. Batak sorot 'to sting', Simalungun sorod '(a) sting', N. Batak sərən

Stress is distinctive e.g. in S. Batak *bàgas* 'house' versus *bagàs* 'deep' and in Simalungun *bàyu* 'to braid, plait' and *bayù* 'new', *tòbu* 'sugarcane' and *tobù* 'sweet.

N.B. Before moving on to the next section, it is worth mentioning some alternative viewpoints on the classification of the Barrier Island and Batak languages. Several scholars have referred to possible links between the Barrier Island + Batak languages and languages in Sulawesi, the Philippines, and even as far as Oceania.

Willms (1955) investigated lexical borrowing into Mentawai. Along with loanwords from other Barrier Island languages (Nias and Simeulue), he also listed 97 ones supposed to be from Sulawesi languages. Nothofer (1986:88) refers to unpublished work by Kähler, in which the latter points to a Philippine-Sulawesi substratum in Barrier Island languages and is particularly struck by the occurrence of shared vocabulary. Nothofer (1994b) argues that Sumatra, the Barrier Islands, the Batak area and parts of Sulawesi and the Philippines were initially a Paleo-Hesperonesian (early Malayo-Polynesian) language area which was subsequently largely supplanted by Hesperonesian (later Malayo-Polynesian) languages. He considers the vocabulary shared specifically by languages in the said Barrier Islands, Batak region and parts of Sulawesi and the Philippines as part of a substratum.

According to Capell (1982:5,6,15), Enggano is not an Austronesian language but one that has a structure "sui generis" (p.5). He nevertheless still classifies it together with other Barrier Island languages in his "Oceanic-type" languages (p.15). Mahdi (1988), proposes a primarily eastern and western split within the Austronesian family tree. He classifies Enggano together with Makuva (a language of Timor Leste called Lovaia in Mahdi 1988) in a Paleo-Hesperonesian subdivision of Eastern Austronesian, and Nias and Mentawai together with Philippine languages, Sulawesi languages and Palauan in Western Austronesian. One of his devices for distinguishing certain subgroups is the term for human being that their members share: Enggano and Makuva are both in a "hartanic" group because their term for human being derives from *qaRta. By the same token, Malay and the languages of Java are in an "Urangic" subgroup as they have reflexes of *uRaŋ as default terms for 'human being'.

10.5 The other Sumatran languages Gayo, Rejang, Nasal and Lampung(ic)

10.5.1 Gayo

In comparison to other languages associated with the Barrier Island – Batak group, Gayo has been rather neglected. The old language atlases from Esser (1938) and Salzner (1960) classify it in a Sumatra group which is basically geographically defined. On the basis of lexicostatistics, Dyen (1965a) classified it in a separate “Hesperonesian” division including Balinese, Sasak, “Dayak” and the Javo-Sumatran but excluding the Batak languages. However, Soravia (1984a) is in favour of a close genetic link between Gayo and Batak. Having written a grammatical sketch of Gayo as well as of the neighbouring Northern Batak language Alas (1984b), he also gives lexicostatistical evidence to support a Batak-Gayo subgroup; especially his percentages for Karo (48%) and Alas (55%) are very high. Nothofer (1986:87) notes that Lafeber (1922) mentioned a Batak-Gayo group. Nothofer also adds Gayo lexicon to his comparative lists, but he does not actually discuss its classification (Nothofer 1986:107). Blust (2013a:78) remarked that “Acehnese, Gayo and the Batak languages represent three quite distinct subgroups”. Smith (2017b), whose phonologically defined Sumatra group is rather different from that of Esser (1938) and Salzner (1960, does not mention Gayo.[6]

Taking the evolution of PMP *j as indicative, it appears that in five cases Gayo has lost *j whereas in four[7] other cases it agrees with various Batak languages in reflecting *ŋg*, final *k* or final *ŋ*. Compare the following six cognate sets[8]:

(6)	PMP	Gayo	Toba Batak	Karo Batak
	*pija ‘how much/many’	piyən, piən	piga	piga
	*qijuŋ ‘nose’	iuŋ	iguŋ	iguŋ
	*qaləjaw ‘day’	lô	---	---
	*qapəjuh ‘gall, bile’	pau	pogu	pegu
	*bajas ‘inside’ (Dempwolff 1934, 1937 1938)	was	bagàs	bagas
	*[hu]aji ‘younger sibling’	eŋ^{g}i	aŋgi	agi
	*pusəj ‘navel’	pusok (Blust online)	pusok	pusuŋ
	*kuñij ‘curcuma’	kuniŋ	hunik	kuniŋ
	(Proto West Malayo-Polynesian *ubaj ‘medicine’ [Adelaar 2005c:371])	uak[9]	obok ‘antidote’	---

The Ø reflex in *pi(y)ən*[10], *iuŋ*, *lô*, *pau* and *was* does not agree with reflexes of *j in other Sumatran languages. Theoretically, it could be analysed as the ultimate reduction of an intermediate (*g >) *h, in which case it is still somehow included in the BI-B subgroup, or even as the ultimate reduction of an intermediate *y, suggesting a genetic affiliation with Moken and Moklen, all things being equal. (However, in the latter case all things do not seem equal, as many other phonological changes in Moken and Moklen are rather idiosyncratic).

In other cases, the Gayo forms agree with Batak ones. However, the agreement is not with a single Batak variety but with two varieties which differ from one another, a factor

---

[6] Nor does Edwards, but his focus is by definition limited to the classification of Enggano (Edwards 2015:82).
[7] *ubaj is based on evidence from Sundanese (ubar), Balinese (ubad), and Malay (obat). Toba Batak *obok* may be related provided that *u and *a in *ubaj underwent vowel harmony in this language; however, this is a speculation, as such harmony is not usual in Toba Batak.
[8] Gayo also has *araŋ* ‘live coal’ which reflects PMP *qajəŋ (same meaning; Eades 2005). This must be a loanword from Malay, which also has *araŋ* ‘live coal’.
[9] I am grateful to Erik Zobel for drawing my attention to this Gayo reflex of *ubaj.
[10] The *y* in *piyən* is the orthographic expression of a non-phonemic glide.

which reduces the evidence for a close genetic relationship between Gayo and Batak. In *eŋᵍi* (exhibiting a “barred nasal”)[11] and *pusok*, Gayo has the reflexes *g* and *-k* (< *g), which it shares with Southern Batak languages such as Toba Batak. In kuniŋ, however, it has a final *ŋ*, which is in agreement with Northern Batak languages (e.g., Karo Batak). There is also much other lexical evidence which Gayo shares with Batak and on first sight suggests a common history between the two, but it all depends on whether the words in question are diagnosed as inherited or borrowed. In some of them there is a clear signal that borrowing from Batak is involved. For instance, Gayo *bətih* ‘to know’ corresponds with Karo Batak *mətəh* (< *bətəh*) and Toba Batak *um-boto* (same meaning): while it looks like an exclusively shared innovation, it must be borrowed from Batak as Gayo regularly lost initial and medial *b in inherited vocabulary, compare *atu* ‘stone’ (< *batu), *oros* ‘uncooked rice’ (< *bəRas), *inih* ‘seedling’ (< *binəhiq), *ulan* ‘moon’ (< *bulan), uak ‘medicine’ (< *ubaj).[12]

Whereas the association between Gayo and Batak needs a closer look, it has also become clear that in Gayo the inherited reflex of PMP *j is Ø and not *g, and that part of the lexicon that Gayo shares with Batak languages is demonstrably borrowed from the latter into the former. Therefore, there is no conclusive evidence for a close relationship between Gayo and the Batak languages.

10.5.2 Nasal

Anderbeck and Aprilani (2013) try to determine the classificatory position of Nasal. There is precious little lexical material for this language, but from what is available, it is apparent that much of Nasal’s lexicon is borrowed from the neighbouring Malayic and Lampungic languages.

There are some similarities between Nasal and Lampung varieties (Lampungic) which are not shared with Malayic. The most prominent of these is the debuccalisation of final stops, e.g. Nasal *hasuʔ* ‘smoke’ (< Proto Western Malayo-Polynesian *qasəp), *lawuʔ* ‘sea’ (< PMP *lahud ‘towards the sea’), *uwoʔ* ‘head hair’ (< PMP *buhək).

However, Anderbeck and Aprilani point out that other commonalities, such as the presence of a final diphthong *-uy and the distinction between *a and *ə in final closed syllables, carry little weight as they are retentions from PMP, or they are too widespread in western MPSEA languages to be significant, as with the change from *q to *h*. They also point out some important distinctions between Nasal and Lampungic. Among these are the maintenance of -*ay, which became *-i* in Lampungic (e.g. Nasal *hənay*, Lampungic *həni* ‘sand’), and the different reflexes of PMP non-initial *R in Nasal (l) and Lampungic (y or Ø), e.g. *qasira ‘salt’> Nasal *sila*, Lampungic *sia; *qatəluR ‘egg’ > Nasal *hatəlul* and Lampungic *hatəluy). Moreover, Nasal and Lampungic have very different histories in terms of lexical borrowing.

The authors conclude that Nasal is not part of the Lampungic group. They then investigate if the two are closely related and possibly derive from an exclusively shared ancestral language, but this proposition also runs into trouble because the only possible phonological support for this, debuccalisation, can be shown to be a post-Proto-Lampungic

[11] That is, an original homorganic nasal + voiced stop in which the velic closure is minimal (or even absent) and nasality is prevented from spreading to the following vowel (see further Truong and McDonnell, this volume).
[12] Then again, a case like *uluŋ* ‘leaf’ is ambiguous: on the one hand, it is a local innovation shared with Batak languages (which have *buluŋ* ‘leaf’); on the other, it has lost initial *b, which tends to happen in inherited languages rather than in loanwords. (The diagnostic value of *b- > Ø needs further investigation).

development which also affected many loanwords. It must have spread from Lampungic into Nasal.

They also compare Nasal to Rejang, which equally shows the change from *R to *l*. However, in spite of the highly emblematic value of this change, it did not happen under exactly the same conditions in both languages, yielding some otherwise regular sets in which Nasal has an l corresponding to Rejang *r*. The fact that Nasal does not seem to significantly agree with other Sumatran languages in its historical developments led the authors to conclude for the time being that Nasal was an isolate and a branch in itself within the Malayo-Polynesian language group (Anderbeck and Aprilani 2013:12). (For completeness' sake, the six Nasal reflexes of PMP *j are all over the place and not helpful as classification evidence either, showing *x*, *h*, *d*, or *-ʔ*, compare PMP *pija 'how much/many' > *pixo*; *pajay 'rice in field' > *pahay*; *qapəjuh 'gall, bile' > *əmpədu*; *[hu]aji 'younger sibling' > *adɪʔ*; *pusəj 'navel' > *pusuʔ*; and *quləj 'maggot' > <u>uluʔ</u> 'caterpillar').

Smith (2017b:454) also tries to find out if Nasal can be classified with other languages in or around Sumatra but he too is unable to find the necessary diagnostic evidence. He deems it unlikely that it is a member of the GNB subgroup because it has *pitu* for 'seven'. Culling from the limited data eight words which qualify at first sight as evidence for an inclusion in his WIn subgroup, he shows that half of them are clearly borrowed from (most likely) Malay, whereas the remaining ones are also suspect of being borrowed. He considers *hulon* 'person, human being', as evidence for the inclusion of Nasal into the WIn subgroup, provided that it is not a Malay loan.[13] However, he concedes that a single word is not sufficient evidence and that there is not enough support to include Nasal in the subgroup.

It is obviously early days for a classification of Nasal, considering the limited resources that are available and the current lack of historical research at all linguistic levels involving this language.

10.5.3 Rejang

According to Voorhoeve (1955), Rejang is closely related to Malay. However, Blust (1984) estimates that the similarities between the two languages are often the result of borrowing from Malay into Rejang. In a phonological history of the latter, he analyses the rather remarkable sound changes it has undergone. They include PMP *w > b, *R > l, ʔ or Ø, *-k and *-q > ʔ, the development of so-called barred nasals (see fn.11), and various complicated and staged vocalic changes leading to a large number of diphthongs and even triphthongs. Compare the following examples (taken from Blust 1984):

(7)

PMP	Rejang
*wahiR 'water	bioa
*wari 'day'	bilay
*ñawa 'breathe'	ñabəy
*baqəRu 'new'	bələw
*daRaq 'blood'	daleaʔ

[13] *qulun 'outsider' is one of the lexical innovations Smith (2017b) adds in support of the WIn group. He reasons that if Nasal *hulon* is a regular reflex of WIn *qulun, this language should be classified as WIn, but if Nasal has borrowed Malay *(h)ulun* 'slave', its classification remains undecided. However, borrowing from Malay is unlikely as a semantic change from 'slave' or 'servant' to 'person' would be odd.

*qasiRa 'salt'	siləy
*qapəju 'gall (bladder)'	pəgəw
*tuRun 'to descend'	tuʔun
*qapuR 'lime'	opoa
*danaw 'lake'	danuəw (danoa)
*buRuk 'rotten'	buʔuʔ
*ipən 'tooth'	epen
*uləj 'maggot'	olok

Blust makes typological comparisons with other Austronesian languages having undergone complicated vocalic changes such as Trukese and Kerinci. In Blust (2010b:69-70) he classifies Rejang in GNB because of *tujuaʔ* 'seven'. Smith (2017b:452) does not focus on the classification of Rejang but notes that it must be a GNB language because it has "an apparently native reflex of *tuzuq 'seven'". In a footnote he makes a further observation that although this language is quite different from other Sumatran languages, its retention of -q (< *-q) instead of instead of showing -h makes it difficult to "brush aside" *tujuaʔ* as a loanword.

In other words, its final syllable with diphthongised vowel and glottal stop is supposed to be a guarantee that it is not borrowed. But in fact, these features do not give that sort of guarantee as they are also found in loanwords. Compare loanwords like *Atjiaʔ*[14] 'Aceh', *Uleaʔ Talo* [*Allah Taʕālā*], which is Arabic for 'God The Elevated', *Mekeaʔ* 'Mecca' (Malay *Məkah*), *sugiaʔ* 'wealthy' (< Javanese *sugih*), *pluaʔ* 'a work gang' (most likely < Dutch *ploeg* [plux] 'team (of workers); soccer team'; Jaspan 1984). Apparently, the final *-h of source words became -ʔ in Rejang. As a result, the phonemic structure of the resulting loanwords can no longer be distinguished from that of native lexicon. Rejang tujuaʔ may be borrowed, especially in light of the fact that it co-occurs with *dlapeun* 'eight' (< *dua-alap-an, 'eight') and *smbileun* 'nine' (< *(ə)sa-ambil-an 'nine') numerals for which a borrowing from Malay seems an obvious explanation, as they bear traces of a suffix *-an, whereas Rejang has no suffixes. Compare Malay *səlapan* 'eight' and *səmbilan* 'nine' (see further §10.8).

The phenomenon of remodeling loanwords such that they cannot be distinguished from inherited vocabulary even if there is no structural motivation for it has been noticed in historical research involving various other language families. Aikio (2006) calls it "etymological nativization of loanwords" in his discussion of similar cases in the adoption of Finnish loanwords into Saami and of Spanish loanwords into Basque (among others). Nash (1997) and Evans (1998) use the term "correspondence mimicry" for manifestations of it in Australian languages.

Smith (2017b:455) also claims that Rejang merged PMP *d and *j (presumably to *d*), but this is inaccurate in two respects. First, in the Rawas dialect, final *j, does merge with *d and *t into *-t* (McGinn 2003), but in the Lebong dialect it does not, becoming -k instead (Jaspan 1984), as demonstrated below:

(8)	PMP	Rawas	Lebong
	*laləj 'mosquito'	dalət	daleuk
	*pusəj 'navel'	pusət	posok

[14] The 'tj' in this word is a feature of colonial spelling. It is written as 'c' in modern Indonesian spelling.

*quləj ‘maggot’	ulət ‘caterpillar’	olok ‘caterpillar’
*lahud ‘towards the sea’	laut ‘sea’	laweut ‘sea’
*tukəd ‘prop, support’	tokot ‘stick’	tokot ‘stick’
*kabut ‘fog’	kabut	kabuet
*həpat ‘four’	pat	pat

Second, in both the Rawas and other dialects (including Lebong), intervocalic *j became g (in two instances) or Ø (in three instances); *d remains d in this position:

(9)	PMP	Rawas	Lebong
	*ŋajan ‘name’	gén	gän
	*qapəju ‘gall, bile’	pəgəw	pəgaw
	*pajay ‘rice-plant’	pai	pai
	*qajəŋ ‘charcoal’	aʔaŋ	aʔaŋ
	*qijuŋ ‘nose’	yuŋ	?
	*qudaŋ ‘shrimp’	udaŋ	udaŋ
	*tiduR ‘to sleep’	tidua	tidua
	*qudip ‘to live; alive’	idup	idup

It appears that in Rejang, PMP *j did not merge with PMP *d except in final position, and only in the Rawas dialect. In the Lebong dialect *j and *d have been kept distinct all along in all positions. Following Smith’s own criteria, Rejang qualifies as much as a Sumatran language as it does as a WIn one. Note incidentally the partial analogy with *j reflexes in Gayo, which has four instances of Ø, and three instances of -g-/-k.

For the sake of completeness, subgrouping attempts by McGinn are mentioned here. McGinn (1999) is an attempt to link up Rejang with Malay and Mukah Melanau (Sarawak) on the basis of morphological comparisons. McGinn (2003) uses phonological arguments to subgroup Rejang with the Bukar Sadong dialects, a division of the Land Dayak group (see ‘Historical Linguistics of the languages of Borneo’ by Smith, this volume). Adelaar (2007a) is a critical review of this somewhat controversial paper.

10.5.4 Lampung

There are three main Lampung dialects, Lampung Api (or Pesisir), Komering and Lampung Nyo (or Abung). Anderbeck (2007b) made a thorough phonological and lexical reconstruction of Proto Lampung using lexical evidence from 23 subdialects. He offers a list of 15 “defining characteristics of Lampungic” (summarised on page 7). Some of these are not unusual among the languages in western MPSEA languages and therefore not so relevant for subgrouping. However, other ones are shared with Sundanese and are evidence for a direct relationship between both languages. The most diagnostic change among these is the one from PMP *R to Proto Lampungic y; examples:

(10) PMP *wahiR ‘water’ > Sundanese *cai*, Proto Lampungic **wai*
PMP *buRuk ‘rotten’ > Sundanese *buyuk*, Proto Lampungic *buyuʔ
PMP *təRas ‘hard’ > Sundanese *tɨas*, Proto Lampungic *tias
PMP *bəRas ‘unhusked rice’ > Sundanese *beas*, Proto Lampungic *bias
PMP *qulaR ‘snake’ > Sundanese *oray*, Proto Lampungic **ulay*

PMP *baRəq ‘swollen’ > Proto Lampungic *bayəh, Sundanese *bayah* (*a* in last syllable unexplained)

Compare also: Malay *kura-kura*, Javanese kura, Balinese *kəkua*, Toba Batak *hura-hura*), SUN *kuya*, Proto Lampungic *kuya ‘tortoise’

However, there are also cases in which either Lampungic or Sundanese lack the expected y:

(11) PMP *bəRat ‘heavy’ > Proto Lampungic *biat, Sundanese *bɨrat*

Anderbeck also mentions PMP *-iw to Proto Lampungic and Sundanese *-(y)u as a shared innovation, but this change occurs only in two instances. One of these (PMP *kahiw ‘wood’ > *kayu*) is not significant because the resulting form also emerges in many other languages in the area. The other (PMP *baRiw ‘beginning to spoil’ > Lampungic dialects *bayu* ‘spoiled’) does have a (Baduy) Sundanese counterpart *wayuq* ‘fermented drink’ but according to Nothofer (1975:215) it is likely to be a Javanese loanword.

Anderbeck points out that the Sunda-Lampung subgroup is further strengthened by the circumstance that the Sundanese and Lampung speaking areas are adjacent, and by the existence of the following two cognates which are shared among Lampungic, Sundanese and Javanese and were adduced by Nothofer (1985:297): Proto Malayo-Javanic *pa-waRi ‘dry in the sun’ > Sundanese poe, Javanese pé, Lampung paway; Proto Malayo-Javanic *pitək ‘gadfly’ > Sundanese pitɨk, Javanese pitaq, Lampung pitoq. But this lexical evidence is neither much, nor is it exclusive in the case of *poe*/*paway* in view of the presence of Javanese pé; cognates are furthermore found in Sulawesi languages.[15]

Another feature, debuccalisation of final (voiceless) occlusives, is typical of Lampungic varieties, but Anderbeck shows that it is not diagnostic for reconstruction and subgrouping purposes because one of the 23 Lampungic subdialects, Paku, escaped this form of levelling. Examples of debuccalisation are as follows (Anderbeck 2007b:14):

(12)	Proto Lampungic	Komering (Paku)	Komering (Kayu Agung Asli	Nyo (KotaBumi)
	*pahit ‘bitter’	pahit,	pahiʔ	pahiʔ
	*hasəp ‘smoke’	hasop	hasoʔ	asəʔ
	*sərəp ‘needle’	sawop	soʁoʔ	səʁəʔ
	*huət ‘husk of rice’	huət	huoʔ	uəʔ

Note incidentally that the establishment of a Sunda-Lampung subgroup would be another strain on the use of a reflex of *tuzuq for ‘seven’, as Lampung varieties maintained a reflex of *pitu whereas Sundanese has tujuh.

[15] Erik Zobel (p.c.) draws attention to the fact that there are also cognates in Sulawesi languages, to wit Tialo *puagi*, Pendau *puai* (Himmelmann 2001) and Proto Bungku-Tolaki *puai (Mead 1998).

10.6. The languages of Java

Many historical linguistic observations had been made about the languages of Java in the nineteenth and early twentieth century, and Dempwolff (1934–1938) made systematic use of Javanese material in his reconstruction of PMP. However, it was not until 1965 that these languages became the subject of a more detailed classification, when Dyen (1965a) included them in a lexicostatistical study which had the entire Austronesian language family in its scope. He grouped Javanese together with Sundanese, Malay and Madurese. (Some other languages which he closely affiliated with specifically Malay were Minangkabau, Kerinci, Lampung and Acehnese). Dyen's Javo-Sumatran became the basis of a thesis, published ten years later, by Nothofer, who used Javanese, Sundanese, Malay and Madurese for a reconstruction of the phonology and lexicon of what he called Proto-Malayo-Javanic. It was a solid thesis, but it was also a product of its time, when very little research had been done on the history of the languages concerned, and lexicostatistics was still considered as a scientific basis for subgrouping. Furthermore, Nothofer followed Dyen's method of reconstructing additional proto phonemes whenever a sound correspondence left exceptions (rather than "resigning" to the fact that there are exceptions to most sound correspondences). These additional proto phonemes were indicated with capital letters of given a number (*R_1, *R_2, *R_3 etc.). The advantage of this diversification of proto phonemes was that irregular sound changes were made explicit and given their own symbol. A somewhat less fortunate consequence was that the reconstructed phoneme system looked rather unbalanced and had become unrealistically large. For instance, Proto Malayo-Javanic has one *r and three *R's, two *w's, *z + *j, *b + *B, a retroflex *ḍ and alveolar *D, and seven final diphthongs. Many of these proto phonemes are the result of undetected borrowing. Furthermore, Proto Malayo-Javanic initial and final glottal stop (*q) was reconstructed on the basis of Sundanese, which does not allow word-initial or word-final vowels and has a default glottal stop in these positions in all roots. Proto Malayo-Javanic geminate consonants were reconstructed based on Madurese, in which they are contrastive. (Malay and Javanese also have geminated consonants, but these are not phonemic). Compare the word pairs in Table 10.4:

<<**Table 10.4 Comparative wordlist of PMP, PMJ, Sundanese, Javanese, Malay, and Madurese**
to be inserted somewhere here>>

Another drawback was that Proto Malayo-Javanic also included phoneme series that are the result of South Indian influence, such as a retroflex series, which was based on the retroflexes ḍ, ḍh and ṭ which exist in Old and New Javanese and in Madurese, and a series of aspirated voiced stops. The latter were based on the aspirated stops (bh, ḍh, jh, gh) that currently exist in Madurese and used to occur in Old Javanese, but not in modern Javanese.

Blust (1981c) made some other critical observations. For one thing, the Malayo-Javanic (or Dyen's Javo-Sumatran) subgroup included languages that did not belong in the group (e.g. Javanese, which even by lexicostatistical standards is not as close to other Malayo-Javanic languages as they are to one another). At the same time, it failed to include languages that do belong. Blust (1981c:459) singles out Iban. This language also maintains the distinction between *b and *w when occurring between *a's, whereas Nothofer (1975:81) considered it lost in Proto Malayo-Javanic since PMP *b and *w had merged to *w* in Malay, Javanese and Sundanese, and to *b* in Madurese. Compare the following cognate sets (Blust 1981c:459):

(13)	PMP *tabaR 'tasteless, unsalted, unsweetened' >	Iban tabar 'id.', ML, SUN tawar, JV tawa 'tasteless, powerless, sweet, MAD tabar 'uncooked'
	PMP *laban 'opposed, against' >	Iban laban 'id.', ML, JV, SUN lawan, MAD laban 'opponent'
	PMP *ñawa 'soul, breath' >	Iban ñawa 'life, existence', ML, JV, SUN ñawa, MAD ñaba 'soul, spirit'
	PMP *awaŋ 'atmosphere' >	Iban awan (final -n unexplained) 'air', ML, JV, SUN awaŋ-awaŋ, MAD abaŋ-abaŋ 'airspace'

Furthermore, the distorting effects of lexical borrowing from Javanese and Malay into other languages were not sufficiently recognised. Nor was it recognised that languages may have different retention rates of core vocabulary, and that high retention rates have a skewing effect on the rate of basic vocabulary shared among languages. However, the words in question are retentions and therefore are not significant for subgrouping. Blust suggests that high retention rates of core vocabulary may be the reason why Malay and Madurese have so much basic vocabulary in common.

Blust also argued for the use of qualitatively based lexical innovations, especially numerals. In Malay and Sundanese, PMP *pitu 'seven' was replaced by tujuh and its cognates, PMP *wwalu 'eight' by Malay dəlapan and Sundanese dalapan (< *dua + *alap-an), and PMP *siwa 'nine' by Malay səmbilan, Sundanese sambilan (< *(ə)sa + *ambilan); in Javanese and Madurese, *siwa was replaced by saŋa 'nine'.

He proposed an alternative classification of languages relative to their genetic distance to Malay. It is based on the *tuzuq device discussed above and the metathesis in PMP *qudip 'to live; alive': in Malay and languages closely related to it the latter became (h)idup. Four degrees of relatedness are distinguished. In the first degree are Malay, Minangkabau, Kerinci and "Middle-Malay" (Pesemah and Seraway in South East Sumatra); (2) in the second degree, Salako and Iban (Malayic languages from western Borneo); (3) in the third degree, Sundanese, Embaloh ('Maloh', in Borneo) and Rejang (Sumatra); and in the fourth degree, Acehnese and Cham and Jarai (Vietnam, Kampuchea).

In his evaluation of the position of Madurese, Blust should also have highlighted that in this language PMP *j had become l (as seen, for instance, in èlòŋ 'nose' < *PMP *qijuŋ).[16] The development has not gone unnoticed in the literature but nor has it been given due recognition in the wider context of subgrouping evidence. This phonological evidence is relatively much stronger than an elevated lexicostatistical percentage in need of further scrutiny, and it gives extra weight to Blust's arguments against unqualified use of the latter.

---

[16] Other instances are *alèʔ* 'younger sibling' (< *(hu)aji); *pəlləs* 'smarting like when hit with thin rattan' (< *hapəjəs 'smarting, stinging pain'); *bálá* 'to announce, say' (< *bajaq 'to tell, inform; ask, inquire').

Notwithstanding the critical observations outlined above, Nothofer (1975) still stands out as a very systematic and well-documented work which was based on insights and methodological principles that were innovative at the time of its appearance.

In his reply, Nothofer (1985) accepts Blust's objections in general outline and refines some of his subgrouping conclusions in Nothofer (1975). But he also justifiably opposes Blust's assertion that until relatively recently Sundanese and Malay had never been in close contact, and he rejects his arguments based on numerals. He still maintains the Malayo-Javanic subgroup, while acknowledging that Javanese is more distantly related to other members. He adduces new lexical support for the subgroup, which now also includes Lampung and Iban. As discussed above, he also provides some evidence for a close relationship between Sundanese and Lampung.

As already indicated, in Adelaar's (2005c) Malayo-Sumbawan, Javanese is excluded; it puts Malayic, Chamic (+ Acehnese) and the BSS languages on a par in one primary branch, and Sundanese and Madurese in two other ones. It is rejected by Blust (2010b), who includes Sundanese and Malayic in GNB, (see §10.8 for an evaluation of these alternatives).

In conclusion, it is clear that the languages of Java do not show the same close kinship to one another as do the Batak -, Lampungic - or BSS languages. The phonological features and lexicon they have in common are also the result of a very long history of contact. This contact has been both global (e.g. from the Indian subcontinent) or local. It has clearly also affected BSS languages. Locally, Malay and Javanese have for a very long time been hegemonic languages and major sources of influence, both on other Austronesian languages (in Java, but also elsewhere in MPSEA) as on one another. Other factors such as a high common retention rate of basic vocabulary (between Malay and Madurese) as suggested by Blust may also play a distorting role in subgrouping attempts, especially in the use of lexicon.

At the current state of research, the default assumption is that the languages of Java simply not do not form a genetic subgroup, something that becomes even more likely if their common history and similarities due to contact are taken into account. In terms of genetic distance, it has become clear that Javanese is further removed from Malay, Sundanese and Madurese than the latter are from one another, and Malay (possibly also Sundanese) is more significantly related to various languages outside Java than it is to languages within. Further research based on more than phonology and lexicon alone will be needed to clarify the picture.

10.7. Balinese-Sasak-Sumbawa

Finally, Mbete (1990) reconstructed Proto Bali-Sasak-Sumbawa on the basis of Balinese, Sasak (spoken in Lombok) and Sumbawa (West Sumbawa). These genetically closely related languages form the Bali-Sasak-Sumbawa subgroup or "BSS", which is divided further into a Balinese branch on the one hand, and a Sasak-Sumbawa branch on the other. BSS was outlined for the first time in 1938 in Esser's linguistic atlas. The latter had previously classified Balinese with Javanese and Madurese, in line with the prevailing viewpoints at the time and instigated by the fact that Balinese had much of its lexicon in common with Javanese. However, this common lexicon was much more a feature of high register Balinese than of low Balinese, which in every respect shows a language that is much more akin to Sasak and Sumbawa. Esser's (1938) subgrouping was taken over in the linguistic atlases of Salzner (1960) and Wurm and Hattori (1981-3). Note that it was basically impressionistic: it was not based on systematic linguistic research but rather on philological insight in Balinese and Javanese. However, it was confirmed by Dyen (1982:32) lexicostatistical research, whose "Balic" covers the same languages. As to the external classification of BSS, as indicated previously, Adelaar (2005c) groups it together with Malayic, Chamic (and Acehnese) in a Malayo-Sumbawan macrogroup. Blust (2010b) includes these language groups in his WIn subgroup, disagreeing with the close affiliation between them as implied in Adelaar (2005c) (see §10.8).

Mbete's 1990 PhD thesis is an internal classification of the BSS languages and a reconstruction of the phonology and basic lexicon of Proto BSS. Primarily using the Bali Dataran dialect of Balinese, the Ngeto-ngete dialect of Sasak, and the Sumbawa Besar dialect of Sumbawa, he applies the comparative method and uses an approach which is strictly from bottom up. He divides the subgroup into a Balinese branch and a Sasak-Sumbawa branch.

Mbete's evidence for his BSS subgroup consists of 41 exclusively shared lexical innovations vis-à-vis PMP[17] (Mbete 1990:74-76) and a handful of formal innovations involving metathesis (p.74). He also provides shared lexicostatistical percentages for Balinese, Sasak, Sumbawa, Javanese, Bimanese and Manggarai showing that the first three languages score significantly higher in shared vocabulary with one another than they do with the latter three (Mbete 1990:66).

As Mbete's approach is from the bottom upwards, he does not use sound changes from PMP to BSS languages to test his subgroup, although he discusses these changes in a later chapter (on "Reflexes of P[roto] A[ustronesia]n phonemes in [Proto Bali-Sasak-Sumbawa]"). Most PBSS sound correspondences are not particularly impressive as they are either PMP retentions or tend to be similar to ones that happened in Malayic and other languages in the region (such as the loss of *w-, and the change from *j to *d* and –*t*). Somewhat more diagnostic are individual irregular sound changes such as the loss of last syllable shown in *kənawan*, the denasalization shown in *tajəp*, and the vowel raising in *taek*, or the reduplication in *kakən. (More significant are various lexical innovations). Compare the following sound developments:

(14) PMP *wada 'exist' > PBSS *adaq > Balinese *adə*, PSS *adaq > Sasak, Sumbawa *adaq*
PMP *waRi 'sun; day' > PBSS *ari > Balinese *ai*
PMP *qijuŋ 'nose' > PBSS *iduŋ > Balinese, Sasak *iduŋ*, *iruŋ*, Sumbawa *iduŋ*
PMP *(ŋ)ajan 'name' > PBSS *adan > Balinese adan, PSS *aran > Sasak, Sumbawa *aran*
PMP *tumid 'heel' > PBSS *tumed > Sasak *tumet* 'heel', Balinese *tumed* 'push with the foot'
PMP *ka-wanan 'right side' > *PBSS* *kənawan > Balinese kənawan, Sasak kawan, Sumbawa kanan
(?PMP) *pəniŋ 'dizzy' > PBSS *pinəŋ > Balinese peneŋ, PSS *pinəŋ
PMP *nahik 'to climb, go up' > PBSS *taek > Sasak taek, Sumbawa (ən-)tek
PMP *tazim, *tazəm 'sharp point; to whet' > PBSS *tajəp 'sharp'> Balinese, Sasak tajep, Sumbawa tajam (< Malay?)
PMP *kaʔən 'food' > PBSS *kakən > Sasak *kakən*, Sumbawa *kakan*

[17] Mbete uses Proto Austronesian (PAn) as a reference point, but for reasons of phonological transparency and given the topic of the current volume, PMP is used instead.

Mbete's Sasak-Sumbawa microgroup is based on phonological, lexical and lexicostatistical evidence. His main phonological evidence for distinguishing Balinese as one branch versus Sasak-Sumbawa as another is as follows:

Phonological changes:

(15) PBSS *-a > Balinese -ə (not in all dialects)
PBSS voiced final occlusives > PSS devoicing of final occlusives
PBSS *-d- > PSS *-d-, *-r-;
PBSS *r, *R > PSS *r;
PBSS *R > Balinese Ø-, -Ø-, -h;
PBSS *-q > Balinese -h, PSS -q
PBSS *-w- > Balinese Ø, PSS -w-
PBSS *əRV > Balinese VV
PBSS heterorganic nasal + stop clusters become homorganic
PBSS heterorganic clusters of unlike stops are reduce to last stop

Examples:

(16) PBSS *səRəb 'to glance, spy, lurk' > Balinese *saəb*, PSS *sərəp
PBSS *uRad 'vein, nerve' > Balinese *uad*, PSS *urat > Sasak, Sumbawa *urat*
PBSS *əmpug 'split, peeled' > Balinese *əmpug*, PSS *əmpuk
PBSS *bəgaq 'just, only, enough' > Balinese *bəgah*, PSS *bəgaq
PBSS *adan 'name' > Balinese *adan*, PSS *aran
PBSS *siwa 'nine' > Balinese *siə*, PSS *siwaʔ > Sasak, Sumbawa *siwaʔ* (with irregular ʔ accretion)
PBSS *ada > Balinese *adə*, PSS *adaʔ > Sasak *araʔ*, Sumbawa *adaʔ* (with irregular -ʔ accretion)
PBSS *Ratus 'hundred' > Balinese *-atus*, PSS *ratus > Sasak *ratus*, Sumbawa *ratis*[18]
PBSS *pəRut 'to shrink' > Balinese *puut*, PSS *pərut
PBSS *bəRas 'rice (uncooked)' > Balinese *baas*, PSS *bəras
PBSS *kaRuŋ 'male (of animals)' > Balinese *kauŋ*, PSS *karuŋ
PBSS *akaR 'root' > Balinese *akah*, PSS *akar
PBSS *diŋdiŋ 'wall' > Balinese *diŋdiŋ*, PSS *dindiŋ
PBSS *tabtab 'to knock hard' > Balinese *tabtab*, PSS *tatap > Sasak, Sumbawa *tatap*

Mbete's lexical evidence consists of 31 lexical innovations (Mbete 1990:79-80).
Finally, Mbete gives the following lexicostatistical percentages for basic vocabulary shared among Balinese, Sasak and Sumbawa: Sasak-Balinese, 51%; Sumbawa-Balinese, 49%; Sasak-Sumbawa, 64%. They show a significant gap between Balinese on the one hand versus Sasak and Sumbawa on the other.

[18] In Sumbawa, last syllable *u often changes to i, as in (PMP) *niuR 'cocospalm, coconut' > *ñir*; *tahun 'agricultural cycle; year' > *tin*; *Ratus' hundred' > *ratis;* *tuhəd 'knee' (> **əntut*) > *əntit*.

As indicated before, Adelaar's (2005c) Malayo-Sumbawan subgroup has Malayic, Chamic and the BSS languages on a par in one primary branch, and Sundanese and Madurese in two other ones. It is rejected by Blust (2010b) on various grounds. Some of these concern the quality of Adelaar's arguments. Blust concedes that there are many phonological developments shared between the Malayic, Chamic and BSS languages, but he considers most of these insignificant because they are also frequently observed in languages outside the group (e.g. PMP *q > *h*, *R, *r > *r*, *h > Ø). Other phonological developments are not properly shared: they do not yield a representative proto form at the Malayic-Chamic-BSS level because they do not occur in all languages that are diagnostic for such a subgroup (as with reflexes of *-ay, *-uy, *-iw), or, if they do occur, they differ in detail to the extent that they must have followed independent pathways (as with PMP *w-). Moreover, Blust considers the fact that BSS languages lack a reflex of *tuzuq meaning 'seven' as crucial evidence for rejecting their inclusion in the same subgroup as Malayic and Chamic.

Blust's meticulous assessment mostly conclusive, except for the undue value he attaches to Chamic *tuzuq reflexes, which are likely to be borrowed from Malay (see §8).

10.8. The use of *tuzuq 'seven' and of PMP *j as subgrouping evidence: an appraisal
The use of these reflexes is a recurrent theme in the subgrouping proposals in this chapter. While in general, these reflexes are useful diagnostic devices for subgrouping, it has also become increasingly clear that neither of them have absolute diagnostic value with regard to the languages under discussion, Concerning reflexes of *tuzuq, Blust acknowledges that Macassarese *tuju* 'seven' (in Sulawesi) must be borrowed from Malay. However, borrowing may also apply to *tuzuq reflexes in Rejang and even in Chamic languages, Sundanese, Moken-Moklen, and Simeulue.

There are some indications that this is in fact the case. These reflexes (often together with the reflexes of 'eight' and 'nine') have been replaced or are in the process of being replaced by a Malay loanword in various languages.

• In Simeulue, inherited *ito*/*itu* and *tujuh* (< Malay) exist side by side (Blust 2010b:69).
• Gayo has inherited numerals but systematically uses Malay numerals alongside them (Eades 2005:86)].
• In Rejang, *tujuaq* 'seven' occurs along with reflexes of *dua-alap-an, 'eight' and *(ə)sa-ambil-an 'nine', which are clearly borrowed from Malay. This increases the likelihood that *tujuaq* is too: while its phonological appearance suggests that it is inherited, there is in fact no guarantee that it is, as shown by the evidence of other loanwords (§8.5.3).
• The situation in Chamic is comparable to that in Rejang. There is no direct evidence that *tuzuq reflexes are borrowed but the fact that they co-occur with reflexes of *dua-alap-an, 'eight' and *(ə)sa-ambil-an 'nine' (and in the Chamic case, also reflexes of *(ə)sa-alap-an 'nine'), makes it likely that they are. Like Rejang, Chamic languages lack reflexes of the roots *alap and *ambil. They also lack reflexes of the suffix *-an (in fact, they lack any suffixes). Moreover, in Chamic, reflexes of *dua + *alapan, *(e)sa + *alapan and *(e)sa + ambil-an have very unstable meanings. This is abundantly demonstrated in Thurgood (1999:38-39), where it is shown, among others, that in Rade and Jarai, reflexes for 'nine and 'eight' were reversed, and in some other cases, *dua + *alapan, *(e)sa + *alapan and *(e)sa + ambil-an compete in denoting the same numerical values .[19] On the other hand, in standard and

[19] Thurgood himself is convinced that the Chamic forms for 'seven', 'eight', and 'nine' are evidence for subgrouping Chamic with Malayic (p.38), pointing out that some of the confusion regarding *(əsa)-alapan and

literary forms of Malay, the morphemic structure and semantics of these compound numerals are transparent, and it is clear that their Chamic counterparts were borrowed from Malay. As far as *tuzuq reflexes are concerned, they formally betray neither inheritance nor borrowing, but their occurrence in the same numeral series as reflexes of *dua *alap-an and *(e)sa *ambil-an only increases the likelihood that they were borrowed as well.

Malayic and Chamic are obviously closely related, and it may be the case that this relationship excludes other Malayo-Polynesian languages, as is proposed e.g. by Blust (2010b) and Thurgood (1999). However, the absence of *tuzuq reflexes for 'seven' in BSS languages is *by itself* not a sufficiently robust argument for denying them a coordinate position in a subgroup with Malayic and Chamic. Nor is the presence of these reflexes in Chamic languages critical evidence for a close relationship between Chamic and Malayic, as it is likely that they were borrowed from Malay.

Sundanese *tujuh* could also be borrowed from Malay, but here the evidence is less compelling as the forms in question could theoretically also have developed independently in this language. It has *dalapan* 'eight' (< *dua *alap-an 'two taken away') and salapan 'nine' (< *(ə)a *alap-an 'one taken away'). Note also Old Javanese *ḍolapan* 'eight', which occurs sporadically alongside *wwalu* and is clearly borrowed from Malay.

Reflexes of *tuzuq 'seven' can be the result of contact and outside influence, even if they show inherited phonology. One can think of various ways in which languages may acquire them other than through membership in the GNB subgroup. They may be loan-words (in which case their phonological shape may or may not give them away); they may be calque translations; or they may be the result of independent parallel developments.

The limited use of *j reflexes as evidence for subgrouping has also become obvious. Here again, the device is not without value, but it is clearly co-dependent on other evidence. Moreover, the evidence of *j is complicated by the fact that languages often do not have a uniform reflex: for instance, Gayo has Ø and g; Rejang has *-g-*, *-Ø-*, *-k* and *-t*; Nasal has *x*, *h*, *d*, and *-ʔ*; Madurese has *l* and *-ʔ*.[20] Furthermore, closely related languages may still have different reflexes of *j, as in South Sulawesi languages, where Buginese and Tamanic have *s*, and Macassarese has *r* (< *d) These South Sulawesi *j reflexes may be relatively recent. Mills (1975) tried to reconcile them by reconstructing Proto South Sulawesi *z, but this is unsatisfactory: a change from *z to *s* can be accounted for, but one from *z to *d*/*r* is odd.

Neither Smith's WIn subgroup nor his Sumatran subgroup are clearly defined by a uniform reflex of *j. About the merger of *j and *d in WIn languages, Smith (2017b:455-456) writes that "there are no languages that unambiguously belong to Western Indonesian (either Greater North Borneo, or Barito) that do not reflect this merger, and it can likely be reconstructed to the immediate ancestor of these languages." It is not immediately clear if he is including WIn languages in Sumatra and Java in his statement, but in the same context he does mention Rejang as another WIn language in which the merger occurred. However, the merger occurred only to a very limited extent (§5.3). Furthermore, if his statement was meant to include all WIn languages, it ignores Madurese, which reflects PMP *j as *l* and *-ʔ*. In Sumatran languages, another concern is the hitherto missed *j > Ø reflex which is predominant in Gayo and is also seen in a part of the Rejang cognates.

---

*dua-alapan also exists in Minangkabau and Kerinci. However, his merely shows that the latter languages must have borrowed these numerals from the same literary Malay as the Chamic languages did.

[20] Note that in Madurese, PMP final *l remained *-l*, in contrast to *-j > *-ʔ*. It shows that in this language final *j did not first change to *l before ending as *-ʔ*, and that *j and *l only merged in intermediate position.

Finally, in both the case of GNB and of WIn, the inclusion of outlier languages such as Moken, Moken in the GNB group (only based on sharing a *tuzuq reflex for 'seven', as done by Blust), and Rejang in the WIn group as defined by Smith (only based on a merger of *j and *d to *d/-t*, which moreover is questionable) is unsatisfactory because it is based on insufficient evidence. Including them seems to do no more than rescuing the subgroups involved, which would otherwise lack a critical exclusive innovation to define them.

To summarise, evidence based on *tuzuq or on the shape of a *j reflex is insufficient by itself. Moreover, if not given absolute diagnostic value, various pieces of other and possibly more promising subgrouping evidence may fall into place, such as the change from PMP *R to *y* linking Sundanese to Lampungic, or the change from PMP *j to Rejang *g* and Ø, allowing this language to align with other Sumatran languages exhibiting *g* (and Ø) for *j. There would also no longer be a compelling reason to include Moken-Moklen with the GNB languages. Without further evidence, that inclusion is rather a stretch, given the idiosyncratic phonological history of Moken-Moklen and its eccentric position in relation to Borneo.

10.9. Concluding remarks

The classification of the languages focussed on in this chapter remains largely unsolved. Three subgroups are uncontroversial, namely Batak, Lampungic, and BSS. Subgrouping Lampungic and Sundanese at a higher cladistic node seems promising but needs further investigation. As to the Barrier Island languages, Simeulue, Sikule and Nias form a cohesive subgroup, but further inclusions of Enggano, and even of Mentawai remain problematic. The classifications of Enggano, Mentawai, Nasal, Gayo, Rejang, Moken-Moklen and the languages of Java remain uncertain and need further investigation. A close relationship between Batak and Gayo is by no means obvious without further investigation.

The evidence of the innovation of a numeral like *tuzuq or of a complex phonological reconstruction like PMP *j is powerful but still needs to be assessed in context. Basically, one is unlikely to find subgrouping devices that have absolute diagnostic value.

The following statement may not be original but still needs to be made. The only promising approach to a classification of the languages in this chapter is to systematically conduct research from the bottom upwards while investigating data at all linguistic levels (including morphosyntax and semantic taxonomies). The results of this research should also be brought into a wider perspective involving contact history and recent research in other disciplines such as archaeology, human genetics, anthropology, and ethnobotany.

Acknowledgements: I am indebted to Bradley McDonnell, Laurie Reid, Antoinette Schapper and Erik Zobel, for their valuable feedback on earlier drafts of this paper.

Iso 639-3 language codes

Alas (Batak)	btx	Bimanese	bhp	Enggano	eno
Angkola Batak	akb	Buginese	bug	Finnish	fin
Balinese	ban	Bukar Sadong	sdo	Gayo	gay
Banjar Malay	bjn	Bungku	bkz	Iban	iba
Basque	eus	Chamic	cmc	Jakarta Malay	bew
Batak	bya	Dairi(-Pakpak)	btd	Jarai	jra
Besemah	pse	Embaloh	emb	Javanese	jav

Karo Batak	btx	Mentawai	mwv	Sasak	sas
Kerinci	kvr	Minangkabau	min	Seraway	pse
Komering	kge	Moken	mwt	Sikhule	skh
Lampung Api	ljp	Moklen	mkm	Simalungun Batak	bts
Lampung Nyo	abl	Mukah Melanau	mel	Simeulue	smr
Macassarese	mak	Nias	nia	Spanish	spa
Madurese	mad	Old Javanese	kaw	Sumbawa	smw
Makuva	lva	Pendau	ums	Sundanese	sun
Malay	zlm	Rade	rad	Tialo	txm
Malay	zsm	Rejang	rej	Toba Batak	bbc
Mandailing Batak	btm	Saami	smi	Tolaki	lbw
Manggarai	mqy	Salako	knx		

Abbreviations

BSS	Bali-Sasak-Sumbawa
JV	Javanese
MAD	Madurese
ML	Malay
N.	northern
OJV	Old Javanese
P.	Proto
PBSS	Proto Bali-Sasak-Sumbawa
PMJ	Proto Malayo-Javanic
PMP	ProtoMalayo-Polynesian
PSS	Proto Sasak-Sumbawa
S.	southern
SUN	Sundanese

References

Adelaar, Alexander. 2007. *Issues in Austronesian historical phonology.* John Lynch ed. Canberra: Pacific Linguistics (Australian National University), 2003. Review: *Bijdragen tot de Taal-, Land- en Volkenkunde 163/1:139-146*.

Adelaar, Alexander. 2005. Malayo-Sumbawan, *Oceanic Linguistics* 44/2:357-388.

Adelaar, Alexander. 1981. Reconstruction of Proto-Batak phonology, *Studies in Malay dialects* (J.T. Collins ed.). Nusa Series 10; 1-10; Jakarta: Badan Penyelenggaraan Seri Nusa.

Aikio, Ante (2006). 'Etymological nativization of loanwords: a case study of Saami and Finnish', in: Ida Toivonen and Diane Nelson (eds), *Saami Linguistics.* Amsterdam: John Benjamins, 17-51.

Anderbeck, Karl. 2007. An initial reconstruction of proto-Lampungic: phonology and basic vocabulary, in: David Mead (ed.), *Studies in Philippine languages and cultures: 10-*

ICAL Sumatra papers (Studies in Philippine Languages and Cultures 16), 41–165. Manila: Linguistic Society of the Philippines and Summer Institute of Linguistics.

Anderbeck, Karl & Herdian Aprilani. 2013. The Improbable Language: Survey Report on the Nasal Language of Bengkulu. Sumatra. SIL International.

Blench, Roger. 2014. The Enggano: archaic foragers and their interactions with the Austronesian world. Unpublished draft. http://www.rogerblench.info/Language/Austronesian/Enggano/Enggano%20and%20its%20hist.

Blust, Robert A. 2013. *The Austronesian languages*. A-PL 008. Asia-Pacific Open Access Monographs. Canberra: College of Asia and the Pacific, The Australian National University.

Blust, Robert A. 2010. The Greater North Borneo hypothesis, *Oceanic Linguistics* 49/1: 44-118.

Blust, Robert A. 1984. 'On the history of the Rejang vowels and diphthongs', *Bijdragen tot de Taal-, Land- en Volkenkunde* 140/4:422-450.

Blust, Robert A. 1981. The reconstruction of Proto-Malayo-Javanic: an appreciation, *Bijdragen tot de Taal-, Land- en Volkenkunde* 137/4:456-469.

Blust, Robert A. & Stephen Trussel (online) *Austronesian comparative dictionary.* (Data on computer file).

Capell, Arthur. 1982. 'Bezirkssprachen im gebiet des UAN', in Rainer Karle, Martina Heinschke, Peter Pink, Christel Rost, and Karen Stadtlander (eds): *'Gava'. Studies in Austronesian languages and cultures dedicated to Hans Kähler.* 1-14. Berlin: Dietrich Reimer.

Dempwolff, O. 1934. *Vergleichende Lautlehre des Austronesischen Wortschatzes. I. Band. Induktiver Aufbau einer indonesischen Ursprache.* (*Beihefte Zeitschrift für Eingeborenensprachen*, 15.) Berlin: Dietrich Reimer Verlag.

Dempwolff, O. 1937. *Vergleichende Lautlehre des Austronesischen Wortschatzes. III. Band. Deduktiver Anwendung des Urindonesischen auf austronesische Einzelsprachen.* (*Beihefte Zeitschrift für Eingeborenensprachen*, 17.) Berlin: Dietrich Reimer Verlag.

Dempwolff, O. 1938. *Vergleichende Lautlehre des Austronesischen Wortschatzes. III. Band. Austronesisches Wörterverzeichnis.* (*Beihefte Zeitschrift für Eingeborenensprachen*, 19.) Berlin: Dietrich Reimer Verlag.

Dyen, Isidore. 1982. The present status of some Austronesian subgrouping hypotheses, in *Papers from the Third International Conference on Austronesian Linguistics, Vol.2: Tracking the travellers* (Amran Halim, Lois Carrington and S.A. Wurm eds), 31-35, Canberra: Pacific Linguistics C-75.

Dyen, Isidore. 1965. *A lexicostatistical classification of the Austronesian languages.* International Journal of American Linguistics, Memoir 19. Baltimore: Waverly Press.

Eades, Domenyk. 2005. *A grammar of Gayo: a language of Aceh, Sumatra* (Pacific Linguistics 567). Canberra: Pacific Linguistics.

Edwards, Owen. 2015. The Position of Enggano within Austronesian. *Oceanic Linguistics* 54(1):54–109. https://doi.org/10.1353/ol.2015.0001.

Esser, S. J. 1938. Languages, in *Atlas van tropisch Nederland*. Batavia: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Evans, Nicholas (1998). 'Iwaidja Mutation and its Origins', in: Anna Siewierska and Jae Jung Song (eds), *Case, Typology and Grammar. In Honor of Barry J. Blake.* Amsterdam: John Benjamins, 115-149.

Himmelmann, Nikolaus P. 2001. *Sourcebook on Tomini-Tolitoli languages: General information and word lists*. PL 511. Canberra: Pacific Linguistics.

Jaspan, M.A. 1984. *Materials for a Rejang-Indonesian-English dictionary.* Materials in Languages of Indonesia, No. 27 (W.A.L. Stokhof ed.). PL-D58. Canberra: Pacific Linguistics.

Hans Kähler. 1963. "Texte von der Insel Simalur". Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg. Berlin: Dietrich Reimer Verlag.

Lafeber, Abraham. 1922. *Vergelijkende klankleer van het Niasisch*. Leiden PhD thesis. The Hague.

Larish, Michael. 1999. *The position of Moken and Moklen in the Austronesian language family.* PhD Thesis, University of Hawai'i.

Larish, Michael. 2011. 'Moken and Moklen', in A. Adelaar and N. Himmelmann (eds): *The Austronesian languages of Asia and Madagascar*, 513-533. New York: Routledge.

Mahdi, Waruno. 1988. *Morphonologische Besonderheiten und historische Phonologie des Malagasy.* Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg Band 20, Berlin-Hamburg: Dietrich Reimer Verlag.

Mbete, Aron Meko. 1990. *Rekonstruksi protobahasa Bali-Sasak-Sumbawa.* PhD Thesis from the Universitas Indonesia (Fakultas Pascasarjana). Jakarta.

Mead, David (1998). *Proto–Bungku-Tolaki: Reconstruction of its Phonology and Aspects of its Morphosyntax*, Ph.D. Thesis, Rice University.

McGinn, Richard. 2003. Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang. In John Lynch (ed.): *Issues in Austronesian historical phonology.* 37-64. Canberra: Pacific Linguistics.

McGinn, Richard. 1999. 'The position of the Rejang language of Sumatra in relation to Malay and the 'ablaut' languages of Northwest Borneo'. In Elizabeth Zeitoun and Paul Jen-kuei Li (eds): Selected papers from the eighth International Conference on Austronesian Linguistics. Symposium Series of Linguistics, No 1. 205-226. Taipei: Academia Sinica.

Mills, Roger F. 1975. *Proto-South-Sulawesi and Proto-Austronesian phonology* (2 volumes). Ph.D. thesis, University of Michigan (University Microfilms International 1978).

Nash, David (1997). 'Comparative flora terminology of the central Northern Territory Area', in: Patrick McConvell and Nicholas Evans (eds), *Comparative Linguistics and Australian prehistory.* Oxford, UK: Oxford University Press, 187–206.

Nothofer, Bernd. 1994. 'The relationship of the languages of the Barrier Islands and the Sulawesi Philippine languages', in *Contact-induced language change in the Austronesian-speaking area* (T. Dutton and D. Tryon eds.); 1-41. Berlin: Mouton-de Gruyter.

Nothofer, Bernd. 1986. The Barrier island languages in the Austronesian language family. In Paul Geraghty, Lois Carrington & Stephen A. Wurm (eds.), *FOCAL II: Papers from the Fourth International Conference on Austronesian Languages*, 89–107. Canberra: Research School of Pacific and Asian Studies, Australia National University.

Nothofer, Bernd. 1985. The subgrouping of the languages of the Javo-Sumatra Hesion: a reconsideration*, Bijdragen tot de Taal-, Land- en Volkenkunde* 141/2,3:288-302

Nothofer, Bernd. 1975. The Reconstruction of Proto-Malayo-Javanic. The Hague: Martinus Nijhoff.

Ross, Malcolm D. 1992. 'The sound of Proto-Austronesian: an outsiders' view oft he Formosan evidence', *Oceanic Linguistics* 31:23-64.

Salzner, Richard. 1960. *Sprachenatlas des Indopazifischen Raumes.* Wiesbaden: Harrassowitz.

Smith, Alexander D. 2017a. The Western Malayo-Polynesian problem, *Oceanic Linguistics* 56/2:435-490.

Smith, Alexander D. 2017b. *The languages of Borneo: a comprehensive classification*: Unpublished Ph.D. Dissertation. Department of Linguistics, University of Hawai'i at Mānoa.

Soravia, Giulio. 1984a. *A sketch of the Gayo language.* Gruppo Linguistico Catanese No 1. Catania: Facoltà di Lettere e Filosofia, Università di Catania.

Soravia, Giulio. 1984b. *A study of the Alas language.* Catania.

Thurgood, Graham. 1999*. From ancient Chamic to modern dialects, 2000 years of language contact and change*. Oceanic Linguistics Special Series. Honolulu: University of Hawaii Press.

Zoetmulder, P. 1980. *Old Javanese - English dictionary (2 Parts).* KITLV publication, The Hague: Nijhoff.

Voorhoeve, 1955 *Critical survey of studies on the languages of Sumatra.* KITLV Bibliographical Series I. The Hague.

Willms, Alfred. 1955. Lautliche und syntaktische Untersuchungen über die Mentawai Sprache. *Afrika und Übersee* 40. Hamburg.

Wolff, John U. (1988). 'The PAn consonant system', in Richard McGinn (ed.): *Studies in Austronesian linguistics*. 125-47. Athens, Ohio: Ohio University, Center for International Studies.

Wurm, S.A.; Hattori, S. (eds) (1981-3) *Linguistic atlas of the Pacific area (Parts I and II).* Canberra: Pacific Linguistics.

Zobel, Erik (2002). 'The position of Chamorro and Palauan in the Austronesian family tree: Evidence from verb morphosyntax', in Fay Wouk and Malcolm Ross (eds.), The history and typology of Western Austronesian voice systems. Canberra: Pacific Linguistics, 405–434.