



EXPANDING THE PAN CONSONANT INVENTORY¹

Peter Norquest

University of Arizona

[<norquesp@email.arizona.edu>](mailto:norquesp@email.arizona.edu)

Sean Downey

University of Maryland

[<sdowney2@umd.edu>](mailto:sdowney2@umd.edu)

Abstract

This paper provides evidence for three Proto Austronesian (PAN) phonemes that are preserved in several distinct languages and subgroups. These include distinctions between *p and *f, *l and *ʎ, and *k and *g. In addition, we assert that there is expanded evidence in Malayo-Polynesian for two currently recognized phonemes: *ʈ (PAN *C) and *c; evidence for the former has been restricted until now to the Formosan languages, and for the latter to a small group of languages in western Indonesia. These contrasts can be found in Nias (one of the Barrier Island languages off the northwest coast of Sumatra), Dohoi (a Northwest Barito subgroup of Borneo), the Western Central Malayo-Polynesian languages of Bimanese, Hawu, Dhao, Western Oceanic, and more sparsely in languages of North Sarawak, the Philippines, and Sumba. The findings presented in this paper highlight the importance of the above languages and subgroups for PAN reconstruction, and the new phonemes presented here are placed within the context of a wider PAN inventory which includes a total of seven places of articulation.

Keywords: Proto Austronesian, reconstruction, classification

ISO 639 language codes: nia, otd, bhp, hvn, nfa, kzi, kyi, ind, wew, lur, mvd, kod, akq, xbr, tgl, bik, ceb, mrw, uun, ssf, xsy, tay, sxr, xnb, dru, bnn, ami, ckv, pyu, pwn, sly, mlg, zlm, jav, bbc, mwt, kvh, mqy, ski, ksx

0.0 Introduction²

The Out-of-Taiwan (OoT) hypothesis was first proposed by Blust (1977) who used lexical evidence from Dyen (1963) to show that three conservative phonological distinctions were preserved in the Formosan languages, but in no other Austronesian languages. Under the OoT hypothesis nine first-order subgroups of Austronesian (An) are accepted for Taiwan (Blust 1999)³; conversely, all Austronesian languages spoken outside of Taiwan belong to a single Malayo-Polynesian branch which is defined by certain changes (*C > t,

¹ We would like to thank Steve Lansing for the use of his data collected during field trips to Nusa Tenggara in 2005 and 2007. Additional thanks to Leif Asplund, Brian Hallmark, Andrew Hsiu, Murray Cox, and Paul Sidwell for various comments on earlier drafts of this paper. Thanks also to the participants at the APLL5 and SEALS22 conferences for comments and suggestions on presentations which have been incorporated into this paper. Thanks to Robert Blust for making the ACD available online, without which much of this research would have been much more difficult. Finally, thanks to Linda Konnerth for offering valuable comments and suggestions on the final draft of this paper. This project was supported by the National Science Foundation (Awards #07254470 and #1030031).

² Abbreviations used in this paper include: (P)WCMP = (Proto) Western Central Malayo-Polynesian, PMS = Proto Macro-Sumba, PS = Proto Sumba, PHD = Proto Hawu-Dhao, PWOC = Proto Western Oceanic, PNS = Proto Northern Sarawak, PD = Proto Dayic, PSab = Proto Sabahan, B = Bimanese, H = Hawu, D = Dhao

³ See Ross (2009) for a more nuanced view of higher-order Austronesian subgrouping.

*N > n, and *S > h). These features were purportedly innovated after the original speech community that left Taiwan immigrated into the Philippines.

The OoT model predicts that Malayo-Polynesian is defined in part by a set of phonological mergers that occurred at least by the time of the settlement of the Philippines. However, according to the principles of historical reconstruction, if any distinction exists at any node of the tree which cannot be explained as the result of internal factors (environmental conditioning) or external ones (such as contact), such a distinction must be reconstructed for the proto-language.

In this paper we present evidence for five as yet unrecognized phonological distinctions in Proto Malayo-Polynesian (PMP); these include contrasts between *p and *f, *l and *ʎ, *k and *g (traditional *g reanalyzed below as a the voiced uvular stop *G), as well as evidence for a additional distinctions between *t and *ʈ (previously only recognized in the Formosan languages) and *c and *s, expanding the domain of traditional *c. We argue that there are crucial correspondences between the Western Central Malayo-Polynesian (WCMP) languages, Dohoi (a branch of northern West Barito), Nias (a Barrier Islands language) and Western Oceanic – as well as other MP subgroups -- which are not innovations, but retentions. They are not the results of parallel sound change, but rather the retention of distinctions that have been lost in many other (but not all) parts of the Austronesian-speaking world. The phonemes discussed in this paper are bolded below in Table 1 (traditional reconstructions are placed in parentheses; see below for further discussion); those suggested for the first time are also italicized:

Table 1: Proposed additions to and expansions of the PMP consonant inventory

p	t	ʈ (C)	c	k	q	ʔ
b	d	ɖ (j)	ɟ (z)	g	G (g)	
f	s					h
m	n		ɲ	ŋ		
	l	ʎ				
w	r		j (y)		R (R)	

The paper is organized into four major sections: Section 1 gives an introduction to the individual languages and subgroups which will form the core of the evidence provided in this paper. Section 2 discusses the evidence for the labiodental fricative *f, section 3 the retroflex series (*ʈ, *ɖ, and *ʎ), section 4 the palatal series (*c, *ç, and *ʎ), section 5 the evidence for the voiced velar stop *g, and section 6 the uvular series (*G and *R). Section 7 concludes.

Description of key languages and subgroups

The languages and subgroups that comprise the primary evidence in this paper are Nias, Dohoi, the WCMP languages Bimanese, Hawu and Dhao, and Proto Western Oceanic. According to the most widely accepted version of the Austronesian phylogenetic tree (Adelaar & Himmelmann 2005), Nias and Dohoi are both Western Malayo-Polynesian (WMP) languages; Bimanese, Hawu and Dhao are members of the western half of the Central Malayo-Polynesian (CMP) subgroup; and PWOC, a subset of Oceanic, is part of the Eastern Malayo-Polynesian (EMP) subgroup. With the exception of the South-Halmahera-West-New-Guinea (SHWNG) group, then, witnesses are therefore found in all major Austronesian subgroups outside of Taiwan:

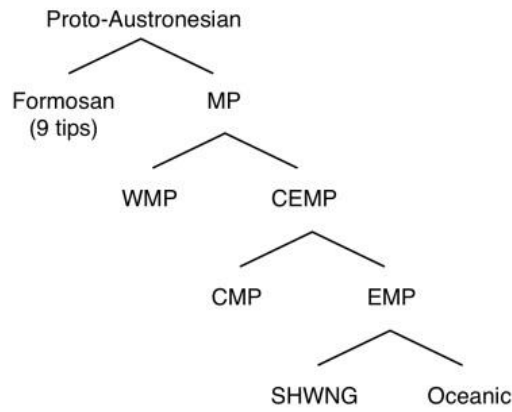


Figure 1: The Austronesian phylogenetic tree

The correspondences between them are presented in the following table, with our proposed reconstruction shown in a Revised Malayo-Polynesian (RMP) column:

Table 2: Consonant correspondences for five phonemic distinctions in PMP

PMP	RMP	Nias	Dohoi	Bima	Hawu	Dhao	PWOc
*p	*p	f-	-p-	p	p	p	*p
*p	*f	β-	-hp-	f	0	0	*β
*t	*t	t-	-t-	t	t	t	*t
*t	*t̪	d-	-ht-	d	d	d̪	*t̪
*s	*s	z-	-s-	s	h	s	*z
*s	*c	s-	-s-	c	h	c	*s
*k	*k	k-	-k-	k	k	k	*k
*k	*g	g-, -ʔ-	-hk-	h	0, -ʔ-	0, -ʔ-	*ɣ
*l	*l	l	-r-	l	l	l	*l
*l	*l̪	l	-r-	r	r	r	*l̪

Note that distinct reflexes in Nias are restricted to initial position, with the exception of the split between *k and *g; conversely, all distinctions in Dohoi are restricted to medial position. Phonemic splits occur in the WCMP languages and PWOc in all positions.

Western Central Malayo-Polynesian

Blust (2008:98) argues for a relationship between Hawu-Dhao and the Sumbanese languages, but states the following in his conclusion: “In any case, it is clear that the most strikingly distinctive innovations shared by Bimanese with Kambara, Hawu-Dhao, or Manggarai cannot be taken as evidence for an exclusive “Bima-Sumba” group, and with this conclusion comes the end to a myth that has lasted longer than one might have imagined possible.”

Blust is referring here primarily to the distinctions listed above in Table 2, and states that these are inexplicable parallel innovations in Bimanese and Hawu-Dhao that are not inherited from a common ancestor. The critical assumption he appears to make is that these cannot represent retentions from a shared ancestor since similar cases are not found higher up in the phylogenetic tree in Figure 1.

We agree with Blust that there is both evidence for a relationship between Hawu-Dhao and the Sumbanese languages on the one hand (a larger group which we call ‘Macro-Sumba’), and current lack of evidence for a close relationship with Bimanese on the other⁴. This being the case, we assert that the striking

⁴ There are at least a few instances of apparent shared lexical innovations, such as Bimanese *mami*, PMS *mami ‘ripe’; Bimanese *weki* ‘body’ PS *βeki ‘self’; Bimanese *habu*, Dhao *abo* ‘pound’; Bimanese *male*, Dhao *ka-male* ‘withered’. Further investigation is required to tell whether these are restricted to Bimanese and Macro-Sumba or shared by other CMP subgroups.

agreement between the reflexes in Bimanese and Hawu-Dhao indicates a case of shared retention as opposed to a collection of parallel, idiosyncratic splits; if it is true that Bimanese does not subgroup with Macro-Sumba, then this means that these distinctions must have been inherited from a more remote ancestor and been preserved independently.

1.1.1 Bimanese

Examples of the consonantal splits shown above in Bimanese are given below:

(1) Examples of phonemic distinctions in Bimanese

<u>Split</u>	<u>Gloss</u>	<u>PMP</u>	<u>Bimanese</u>	<u>Gloss</u>	<u>PMP</u>	<u>Bimanese</u>
*p/f	seven	*pitu	pidu	stingray	*paRih	fai
*t/ʈ	three	*telu	tolu	sugarcane	*tebuh	dobu
*s/c	breast	*susu	susu	spoon	*sidu	ciru
*k/g	scratch	*kaRaw	kao	louse	*kutu	hudu
*l/ʎ	five	*lima	lima	run, flee	*laRiw	rai

These splits are unconditioned and do not seem to be the result of borrowing from Indonesian or any other language. When borrowing occurs, it nearly always reflects the original consonants of the donor. Note also that unlike the examples of native vocabulary above ('three' and 'sugarcane'), where PMP *ə is inherited as Bimanese [o], the Indonesian vowel [ə] is reflected as [a] in Bimanese:

(2) Examples of Indonesian loanwords in Bimanese

<u>Gloss</u>	<u>Indonesian</u>	<u>Bimanese</u>	<u>Gloss</u>	<u>Indonesian</u>	<u>Bimanese</u>
coach	pəlati	palati	slander	fitnah	fitina
trumpet	tərompet	tarompe	dice	dadu	dadu
nature	sifat	sifa	opium	candu	candu
lute	kəcapi	kacapi	result	hasil	hasi
drawer	laci	laci	bread	roti	roti

It should be noted that the distinctions between [p] and [f], [t] and [d], [s] and [c], and [h] and [g] are neutralized in one particular environment. In addition to a set of derivational prefixes (most notably *ka-* and *sa-*), Bimanese retains what appear to be the remnants of an older nasal derivational prefix (or prefixes). Although apparently synchronically unproductive, their original function can often be inferred:

(3) Examples of post-nasal neutralization in Bimanese

<u>Gloss</u>	<u>Base</u>	<u>Gloss</u>	<u>Prefixed</u>	<u>Gloss</u>	<u>Base</u>	<u>Gloss</u>	<u>Prefixed</u>
to close	pula	closed	mpula	to break	foka	broken	mpoka
straighten	tiri	straight	ntiri	egg	dolu	to spawn	ntolu
insert	coŋge	stuck	ncŋge	nest	soʃu	to nest	ncobu
squint	giri	dazzled	taŋgiri	hook	hawi	to fish	ŋgawi

1.1.2 Hawu-Dhao

Proto Hawu-Dhao (PHD) inherited the same set of distinctions as Bimanese, with similar but non-identical reflexes. Examples are given below:

(4) Examples of phonemic distinctions in Hawu-Dhao

<u>Gloss</u>	<u>PMP</u> ⁵	<u>RMP</u>	<u>Bima</u>	<u>PHD</u>	<u>Hawu</u>	<u>Dhao</u>
seven	*pitu	*p-	pidu	*piqu	pidu	piqu
rice	*pajay	*f-	fare	*are	are	are

⁵ PMP forms are cited in accordance with the ACD. Note that PMP *e represents [ə].

three	*telu	*t-	tolu	*təlu	təlu	təlu
rope	*talih	*t̥-	dari	*ɖari	dari	ɖari
breast, milk	*susu	*s-	susu	*susu	huhu	susu
nine	*siwa	*c-	ciwi	*ceo	heo	ceo
scratch	*kaRaw	*k-	kao	*kao	kao	kao
tree	*kahiw	*g-	hadʒu	*afu	afu	afu
five	*lima	*l-	lima	*ləmi	ləmi	ləmi
flee, run	*laRiw	*l̥-	rai	*rai	rai	rai

Having accepted Blust’s (2008) classification of Hawu-Dhao with the Sumbanese languages, it is natural to ask if the latter have preserved any of the same distinctions. As it turns out, while most of these had merged in Proto Sumba, two can still be reconstructed. Proto Sumba retains a distinction between *s and *c (where *c has lenited to *ç), which is preserved in the languages of northwest Sumba; a distinction between *k and *ɣ is preserved in all Sumbanese languages. The following correspondences and reconstructions can therefore be posited for Proto Macro-Sumba (PMS), the name we have given to the parent of Proto Sumba and Proto Hawu-Dhao:

Table 3: Phonemic splits in Proto Macro-Sumba

PMS	Proto Sumba	PHD	Hawu	Dhao
*p	*p	*p	p	p
*f	*p	*ɸ	ɸ	ɸ
*t	*t	*t	t	t
*t̥	*t̥	*ɖ	ɖ	ɖ
*s	*s	*s	h	s
*c	*ç	*c	h	c
*k	*k	*k	k	k
*ɣ	*ɣ	*0, -ʔ-	0, -ʔ-	0, -ʔ-
*l	*l	*l	l	l
*l̥	*l̥	*r	r	r

In Proto Sumba, *f, *t̥ and *l̥ merged with *p, *t and *l in parallel with most other Austronesian languages. It is likely that *f passed through a stage of affrication before merging completely with *p (*f > *pf > *p). *t̥ and *l̥, on the other hand, apparently partook in a general constraint placed upon the entire retroflex series, where all members of this category merged with their alveolar counterparts. On the other hand, the same three phonemes underwent lenition in PHD; *f likely became a bilabial fricative before finally disappearing altogether (*f > *ɸ > *0), while *t̥ voiced to *ɖ and *l̥ delateralized, merging with *r. Examples of all five contrasts are given below:

(5) Examples of phonemic distinctions in Macro-Sumba etyma

Gloss	PMP	PMS	PSumba	PHD	Hawu	Dhao
warm, hot	*panas	*panas	*panas	*pana	pana	pana
rice	*pajay	*fare	*pare	*are	are	are
hood	*teduŋ	*təduŋ	*təduŋ	*tədu	tədu	tədu
egg	*qateluR	*təlu	*təlu	*ɖəlu	dəlu	ɖəlu
breast, milk	*susu	*susu	*susu	*susu	huhu	susu
nine	*siwa	*ciwa	*çiwa	*ceo	heo	ceo
scratch	*kaRaw	*kaʔo	*kaʔu	*kao	kao	kao
wood	*kahiw	*ɣaju	*ɣaju	*afu	afu	afu
five	*lima	*lima	*lima	*ləmi	ləmi	ləmi
flee, run	*laRiw	*lai	*lai	*rai	rai	rai

The contrast between Proto Sumba *s and *ç has been maintained in only the languages of northwestern Sumba (Wejewa, Laura and Lauli). As stated above, all of the Sumbanese languages distinguish between *k and *ɣ:

Table 4: Reflexes of Proto Sumba *s/ç and *k/ɣ

PSumba	Wejewa	Laura	Lauli	Kodi	Mamboro	Anakalang	Lewa	Mangili	Kambera
*s	z	s	s	h	s	s	h	h	h
*ç	0	0	0	h	s	s	h	h	h
*k	k	k	k	k	k	k	k	k	k
*ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ

Examples of Proto Sumba *s and *ç are given below:

(6) Examples of Proto Sumba *s and *ç

Gloss	PMP	PSumba	Wejewa	Laura	Lauli	Mamboro	Kambera
breast	*susu	*susu	zuzu	sussu	susu	susu	huhu
squeeze	---	*pasu	pazu	passu	pasu	pasu	pohu
one	*isa	*iça	ia	ia	ija	sisu	ha-u
heart	*pusuq	*puçu	pu:	pu:	pu:	pusu	puhu

There is some degree of variation in the reflexes of *ɣ; it is often deleted, and can also be substituted by the glides [j] or [w] depending on the vocalic context:

(7) Examples of Proto Sumba *k and *ɣ

Gloss	PMP	PSumba	Wejewa	Kodi	Mamboro	Anakalang	Lewa	Kambera
scratch	*kaRaw	*kaʔu	kaʔu	ka-jo	---	kau	kau	kau
tail	*ikuR	*ka-iku	kiku	kiku	kiku	kaiku	kiku	kiku
tree	*kahiW	*ɣaju	wazu	ɣai	jai	ai	e	ai
fish	*hikan	*iɣaŋ	ija	iɣja	ija	ijaŋ-u	ijaŋ-u	ijaŋ-u

1.2 Dohoi

Although the distinctions above occur in the WCMP languages in all positions, those in Dohoi occur only in intervocalic position in the final foot of the word (Inagaki 2005: 35). Hudson (1967: 53-4) comments on this in the following way:

“Dohoi is distinctive among all the Barito isolects for its medial preaspirated voiceless stops and affricate⁶. Whether this is a Dohoi innovation or the vestige of a proto-phoneme that has been lost in that environment in all other Barito isolects is indeterminate at present time. If it represents an innovation, it would appear that *T^v(P-DM1) (i.e., voiceless stop and affricate proto-phonemes) became preaspirated in medial position (i.e., /V_V/) in Proto-Dohoi. This holds true for the vast majority of Dohoi forms that can be shown to represent P-B forms.”

The situation which Inagaki (2005: 35-6) describes shows that in general, forms with underlying /-hC-/ are in free variation between [-hC-] and [-C-] at the surface level:

(8) Dohoi [-hC-] forms in free variation at the surface level

/nihpo/	[nipo] ~ [nihpo]	‘tooth’
/ŋamuhto/	[ŋamuto] ~ [ŋamuhto]	‘to shut the eyes’
/nohcot/	[nocot] ~ [nohcot]	‘to shiver’
/borahkaŋ/	[borakaŋ] ~ [borahkaŋ]	‘to sell’

⁶ Dohoi contrasts one more pair of phonemes not shown above: -hc- and -c-. These correspond with PAN *z and *y, respectively; for example: Dohoi ruhca < *luzaq ‘spit’, uhcan < *quzaN ‘rain’, mahcu < *zauq ‘far’, but kacu < *kayu ‘wood’.

However, forms with underlying [-hC-] are not in free variation if they contrast with minimal pairs having underlying [-C-], either native (9) or borrowed from Indonesian (10):

(9) Dohoi [-C-] forms in contrast with native [-hC-] forms

puti ‘banana’	puhti ‘white’
noto ‘(personal name)’	nohto ‘to see’
bakai ‘unfinished’	bahkai ‘monkey’

(10) Dohoi [-hC-] forms in contrast with borrowed Indonesian [-C-] forms

From <u>Indonesian</u>	Native <u>Dohoi</u>
kapan ‘when?’	kahpan ‘thick’
kita ‘we (incl)’	kihta ‘sap, resin’
toko ‘shop’	tohko ‘there are’

While it is apparent from other Northwest Barito languages that a distinction must be reconstructed between intervocalic plain and prenasalized voiceless stops, these categories have merged in Dohoi, which distinguishes instead between plain and preaspirated medial voiceless stops (data from Hudson 1967)⁷:

(11) Proto Northwest Barito plain intervocalic stops

	‘to wash’	‘centipede’	‘to laugh’	‘calf’	‘sibling’s child’
Proto-NWB	*mupuk	*ɟɔlipan	*Ka-tau	*bɔtih	*akun
Dohoi	mupʊʔ	ɟɔlipan	ŋa-tau	bɔtih	akun
Murung 1	mopuk	---	ka-tau	---	akɔ-m
Murung 2	mupuk	ɟalipan	kɔ-tauʔ	bɔtih	---
Siang	mupuk	ɟɔlipan	kɔ-tau	botih	akun

(12) Proto Northwest Barito prenasalized intervocalic stops

	‘spouse’s parent’	‘to swell’	‘banana’	‘elder sibling’	‘forehead’
Proto-NWB	*umpu	*lɔntɪŋ	*punti*ɔŋka	*liŋkɔu	
Dohoi	upuʔ	lɔntɪŋ	putiʔ	ɔkaʔ	likou
Murung 1	ɔmpʊʔ	---	---	ɔŋka	---
Murung 2	ɔmpʊʔ	lɔntɪŋ	puntiʔ	ɔŋkaʔ	liŋkou
Siang	---	lɔntɪŋ	punti	ɔŋka	liŋkou

(13) Proto Northwest Barito preaspirated intervocalic stops

	‘fire’	‘thick’	‘eye’	‘louse’	‘1sg’
Proto-NWB	*ahpui	*kahpan	*mahta	*kuhtu	*ahku
Dohoi	ahpui	kahpan	mahtaʔ	kuhtuʔ	ahkuʔ
Murung 1	---	---	mata-m	---	---
Murung 2	apui	kapan	mataʔ	kutuʔ	akuʔ
Siang	apui	kapan	mata	kutu	akuh

There is thus a three-way contrast that must be reconstructed for Proto Northwest Barito. It should be noted that Dohoi generally boyrrows Indonesian words with intervocalic plain stops as preaspirated, and prenasalized stops as plain (data from Inagaki 2005); however, what appear to be more recent borrowings of plain stops are also borrowed as plain. Likely examples of borrowings in all three categories include the following:

⁷ Hudson’s [ɰ] and [ɔ] are equivalent to Inagaki’s [r] and [o].

(14) Indonesian -C- loans into Dohoi

<u>Indonesian</u>		<u>Dohoi</u>	
ləpau	‘lounging-place’	rohpu	‘house’
lapik	‘lining, mat, base’	rahpi	‘sleeping mat’
upah	‘salary, reward’	uhpah	‘reward’
macam	‘kind, sort, type’	mahcam	‘(be) like’
akan	‘future’	ahkan	‘future’
pikir	‘think’	pihkir	‘think’
paku	‘(a) nail’	pahku	‘(a) nail’
pakat	‘discussion, agreement’	pahkat	‘friend’
pənakit	‘disease’	paṇahkit	‘disease’

(15) Indonesian -NC- loans into Dohoi

<u>Indonesian</u>		<u>Dohoi</u>	
pumpun	‘meeting place, center’	pupuŋ	‘peak, meeting’
tampar	‘slap’	tapar	‘slap’
kentut	‘fart’	kotut	‘fart’
lantij	‘house built on a raft’	ratij	‘raft’
rantang	‘hamper’	rataŋ	‘basket’
cinta	‘love’	sita	‘love’
aŋka	‘number’	aka	‘number’
leŋket	‘to cling’	reket	‘to cling’
naŋka	‘jackfruit’	naka	‘jackfruit’

(16) (recent) Indonesian -C- loans into Dohoi

<u>Indonesian</u>		<u>Dohoi</u>	
cepat	‘go fast’	cepat	‘fast’
ketel	‘kettle’	keter	‘cooking pot’
utang	‘debt’	utaŋ	‘debt’
garakan	‘mandated by’	garakan	‘omen’
tikus	‘rat, mouse’	tikus	‘rat, squirrel’

1.3 Nias

With one exception (*g), the distinctions between *p and *f, *t and *t̥, *l and *l̥, *c and *s, and *k and *g only occur in Nias in word-initial position. These have gone unrecognized in the past because they depend on the environment. Lase (2011: xxiv-xxv) describes these as "initial mutations", where the initial of a word undergoes a change when the word is in the middle or at the end of the sentence (i.e. in interphrasal position). It is not true for all lexical items, however, and it is our assumption that this environment preserves original alternations lost in other environments. The initial mutations that can occur in Nias are the following:

(17) Nias initial mutations

[f] > [β]	[b] > [mb]
[t] > [d]	[d] > [ndr]
[s] > [z]	[ʔ] > [g]
[k] > [g]	[ʔ] > [n]

Examples of non-mutating and mutating forms are given below:

(18) Non-mutating ‘fortis’ and mutating initials in Nias

<u>Non-mutating</u>		<u>Mutating</u>		
fao	[fao]	‘with’	fiso [βiso]	‘ear’
taboi	[taboi]	‘expel’	taroʔo [daroʔo]	‘hips’
sara	[sara]	‘one’	simbi [zimbi]	‘jaw’
kiə	[kiə]	‘urine’	kara [gara]	‘stone’
baxa	[baxa]	‘inside’	baho [mbaho]	‘ravine’
dəhə	[dəhə]	‘recover’	dela [ndrela]	‘bridge’
ʔato	[ʔato]	‘crowded’	ʔeha [geha]	‘cough’
ʔirə	[ʔirə]	‘tube’	ʔidanə [nidanə]	‘water’

The initials [b, d, g, mb, ndr, β, z, n] also occur independently in word-initial position, and contrast with the ‘mutating’ initials above:

(19) Non-mutating ‘lenis’ initials in Nias

basi	[basi]	‘reap’
dudu	[dudu]	‘release’
gaβu	[gaβu]	‘sand’
mbadu	[mbadu]	‘breath’
ndrohu	[ndrohu]	‘sober’
βəʔi	[βəʔi]	‘yes’
zizi	[zizi]	‘blunt’
niha	[niha]	‘person’

Our best understanding of the diachronic sources of these three categories of initials are shown below in (20):

(20) Fortis, mutating fortis, and lenis initials in Nias

<u>Fortis</u>	<u>Mutating</u>	<u>Lenis</u>
*p > f	*f > [β]	? > β
*t > t	*t > [d]	---
*c > s	*s > [z]	*j > z
*k > k	? > [g]	*G > g
*b > b	*b > [mb]	*mb > mb
*d > d	*d > [ndr]	*nd > ndr
*ʔ > ʔ	*g > [g]	---
*ʔ > ʔ	? > [n]	*n > n

There are a couple of cases in which the source of the synchronic initial is unclear. For example, we have been unable to locate Austronesian etymologies for any forms with non-mutating [β]. One plausible candidate would be *w, but it is apparent from forms such as Nias *waʔa* ‘root’ (< PMP *wakaR) and *walu* ‘eight’ (< PMP *walu) that Nias has inherited PMP *w as [w].

We have also been unsuccessful in uncovering etymologies for any forms with mutating [g] from surface [k], although many (about 50 percent) of these forms are borrowings from Indonesian words with initial [k], i.e. Nias *[g]afalo* ‘head’ < Indonesian *kəpala* and Nias *[g]atafele* ‘slingshot’ < Indonesian *katapel*. Forms with non-mutating [g] also include a fair number of Indonesian loans with either initial [k] or [g], i.e. Nias *galasi* ‘glass’ < Ind *gəlas*, Nias *gombəni* ‘company’ < Ind *kompəni*, suggesting that the absorption of Indonesian loanwords has added an extra layer of complexity within the overall phonological system, at least in the case of velar stops.

The source of mutating initial [n] is also presently opaque, although it’s frequent occurrence in what might be considered inalienably possessed objects such as kinship terms, body parts, and domestic items

[*n*]ama ‘father’ < PMP *ama, [*n*]ixu ‘nose’ < PMP *hiduŋ, [*n*]omo ‘home’ < PMP *Rumaq) suggests that at least one source of this alternation may be morphological.

1.4 Western Oceanic

Ross (1988) provides extensive discussion on a set of fortis/lenis contrasts in Western Oceanic for PAn *p, *s, and *k. Although limited examples are given of words with lenis reflexes (and fewer for fortis reflexes), the essential contrasts can be reconstructed as the following (we propose Proto Western Oceanic reconstructions based on the lower-level reconstructions in that book)⁸:

Table 5: Fortis and lenis consonants in PWOc

PWOc	PNNG	PPT	PMM	PSS
*p	*p	*p	*p	*p
*β	*β	*β	*β	*β
*t	*t	*t	*t	*t
*t	(*l) ⁹	*t	*t	*t
*s	*s	*s	*s	*s
*z	*z	*z	*c	*z
*k	*k	*k	*k	*k
*ɣ	*ɣ	*q	*ɣ	*ɣ

Ross argues that the lenis reflexes of the Proto Oceanic series *p, *t, *s, and *k are secondary developments and cannot be reconstructed in Proto Oceanic itself¹⁰. Due to this, and the fact that the data in Ross (1988) is somewhat limited, the Western Oceanic data is not as strong as that presented from other languages and subgroups; we include it nevertheless as inspiration for further research in this branch of Austronesian.

Examples of our Proto Oceanic reconstructions based on Ross (1988) are as follows¹¹:

(21) Split correspondences in Proto Western Oceanic

Gloss	POc	PWOc	PNNG	PPT	PMM
squeeze	*poRos	*poros	*poro	---	*poros
banana	*pudi	*βudi	*βudi	*βudi	*βudi
outrigger float	*saman	*saman	*sama	*saman	*[s]ama
paddle	*pose	*βoze	*[p/β]o[s/z]e	*βoze	*βoze
rat	*kusupeq	*ku[s]uβe	*ku[s/z]uβe	---	*kusuβe
tree	*kaju	*ɣaju	*ɣai	*qaju	*ɣaju

2.0 The distinction between *p and *f

We now begin a more detailed discussion of each phonemic split, beginning with the one between *p and *f. As shown above, there are two sets of reflexes associated with PAn *p:

⁸ PWOc = Proto Western Oceanic, PNNG = Proto North New Guinea, PPT = Proto Papuan Tip, PMM = Proto Meso-Melanesian, and PSS = Proto Southeast Solomonian. Reconstructions of PNNG are based on the following subgroups: PSCH (Proto Schouten), PHG (Proto Huon Gulf), PNg (Proto Ngero), PBEL (Proto Bel); those of PMM are based on PWZ (Proto Willaumez), PNI (Proto New Ireland), PNS (Proto North-West Solomonian), and PLN (Proto Lavongai/Nalik).

⁹ This correspondence set is limited to the Bel languages and the closely related Nenaya and Roinji languages as well as languages of southeast Malekula and Epi. As Ross (1988) does not offer any clear examples between fortis and lenis *t, this distinction will not be pursued in this paper.

¹⁰ For example, Ross (1988:50) argues that the variation between the PNNG forms for ‘bow’ (*pana(q)) and ‘shoot’ (*βana(q)) is due to morphological alternation, and this appears to be correct. However, the situation may be the reverse of what he suggests – rather than lenition having occurred in a prefixed form ‘to shoot’, the original initial may have been protected, instead undergoing *fortition* in the free form ‘bow’.

¹¹ Ross (1988) does not provide any examples of contrasts between *t and *l in PNNG, so examples cannot be included here.

Table 6: Reflexes of *p and *f

PMP	RMP	Nias	Dohoi	Bima	Hawu	Dhao	PWOc
*p	*p	f-	-p-	p	p	p	*p
*p	*f	β-	-hp-	f	0	0	*β

We propose that the second set of reflexes associated with PAN *p be reconstructed as *f. These reflexes occur in all positions in Bimanese, Hawu-Dhao, and Western Oceanic; they are restricted to word-initial position in Nias and word-medial position in Dohoi.

Examples of initial and medial *p are given below¹²:

(22) Examples of initial *p

Gloss	PMP	Nias	Bima	Hawu	Dhao	PWOc
causative	*pa-	fa-	pa-	pe-	pa-	---
warm/hot	*panas	---	pana	pana	pana	---
fruit bat	*paniki	---	panihi	niʔi	---	---
broken	*pataq	(a-fatə)	mpada ¹³	pada	---	---
squeeze	*peRes	---	pua	---	---	*poRos
how many	*pija	---	pila	pəri	pəri	(*βiza)
choose	*piliq	fili	---	pili	---	---
seven	*pitu	fitu	pidu	pidu	pidu	---
dragnet	*puket	---	puka	pəku	pəku	---
white	*putiq	(a-fusi)	---	pudi	puɟi	---
sharp pain	*hapejes	([β]əxə)	pili	pəɟa	pəɟa	---

(23) Examples of medial *p

Gloss	PMP	Dohoi ¹⁴	Bima	Hawu	Dhao	PWOc
fathom	*depa	---	ndupa	rəpa	---	---
fold	*lepet	---	lipi	ləpa	ləpe	---
pinch	*qapit ¹⁵	---	api	ŋ-api	ŋ-api	*kapi
mat	*tepiR	---	dipi	dəpi	dəpi	---

The Nias forms for ‘broken’ and ‘white’ have been placed in parentheses because, being affixed, their root initials are in an intervocalic environment where the distinction between [f] and [β] is neutralized. There is also a disagreement between the Nias form for ‘sharp pain’ and the PWOc form for ‘how many’ on the one hand, and the PWMP forms on the other.

Examples of initial and medial *f are given below:

(24) Examples of initial *f

Gloss	PMP	Nias	Bima	Hawu	Dhao	PWOc
paddy	*pajay	[β]axe	fare	are	are	---
to shoot	*panaq	---	fana	---	---	*βana
leucoderma	*panaw	---	fano	ano	---	---
bait	*paniŋ	---	(pani)	ani	ani	---
pandanus	*paŋdan	[β]andra	fanda	---	---	---
thigh	*paqa	[β]aha	---	---	---	*βaqa

¹² Indonesian parallels are given to the right when the phonotactics of some or all forms are consistent with a possible loan scenario.

¹³ “exhausted”

¹⁴ We have unfortunately not been able to locate any Dohoi forms with medial [-p-] which have cognates in the other languages used here.

¹⁵ “tongs, anything used to hold things together by pinching”

chisel	*paqet	[β]ahə	(paa)	(pae)	(paa)	---	(Ind <i>pahat</i>)
hoarse	*paRaw	(a-fau)	fao	---	---	---	
stingray	*paRih	---	fai	---	---	*βari	
mango	*pahuq	---	foʔo	(pau)	(pao)	---	(Ind <i>pauh</i>)
turtle	*peɽu	[β]ənu	fonu	əɽu	əɽu	*βoɽu	
good	*pia	---	---	ie	ia	---	
ten	*sa-ŋa-puluq	[β]ulu	mpuru	ŋ-uru	ŋ-uru	---	
navel	*pusej	[β]usə	---	əhu	əsʉ	---	
heart	*pusuq	---	---	uhu	usu	---	

There are unfortunately not many instances of medial *p, and all are confined to WCMP with the exception of ‘four’ (below) in which there is variation: Dohoi and PWOC pointing to *əfat, but Bimanese and Hawu-Dhao pointing to *əpat. The PHD form *əpa might be explained as being inherited from a previous *əmpa (cf. Iban, Malay, Balinese and Sasak *əmpat*), since the regular reflex of PAN *mp is *p (i.e. PAN *əmpu ‘grandchild’, PHD *əpu). However, the Bimanese reflex of PAN *mp is normally [mp] (cf. Bimanese *ompu* ‘grandchild’), so it may be necessary to reconstruct disjunct forms.

(25) Examples of medial *f

<u>Gloss</u>	<u>PMP</u>	<u>Dohoi</u>	<u>Bima</u>	<u>Hawu</u>	<u>Dhao</u>	<u>PWOC</u>	
thick	*ma-kapal	kahpan	---	me-aa	ma-aa	---	
dream	*h-in-ipi	nuhpi	nifi	nii	nii *m-niβi		
thin	*ma-nipis	mihpi	(nipi)	me-nii	ma-nii	---	(Ind <i>nipis</i>)
centipede	*qalu-hipan	(ɟoripan)	rifa	---	---	*qaliβan	
gall	*qapeju		folu	ədu	ədu	---	
lime (for betel)	*qapuR	---	afu	ao	---	---	
sweep	*sapuh	---	cafi	hai	---	---	
fire	*hapuy	ahpui	afi	ai	ai	*(j)afi	
four	*[h]jepat	ohpat	(upa)	(əpa)	(əpa)	*βati	(Ind <i>əmpat</i>)
reward	*upaq	uhpa	ufa	---	---	---	

Some of the exceptions above are explainable. Nias *a-fau* ‘hoarse’ is intervocalic, so its underlying form has been neutralized. The Hawu and Dhao forms for ‘mango’ are likely borrowings from Indonesian *pauh*, and the same is true for Bimanese *nipi* ‘thin’ from Indonesian *nipis* (compare Bimanese *naji* and Indonesian *mənanjis* ‘weep’, where the expected prenasalized form in Bimanese would be *ntaŋi*). The Nias forms for ‘chisel’ and ‘sharp pain’ have reflexes of *f where the WCMP forms have reflexes of *p; it is possible that the cases of ‘chisel’ in Bimanese and Hawu-Dhao are loans from Indonesian. This may also be the case in the Hawu-Dhao forms for ‘sharp pain’, although the vowels in the Bimanese form make this explanation less plausible.

There is no apparent explanation for the discrepancy between Bimanese *pani* and PHD *ani ‘bait’. In the case of Dohoi *ɟoripan* ‘centipede’, it should be noted that Hudson (1967) occasionally transcribed preaspirated medials as plain and some of these were corrected by Kazuya in his publications; in this case, Kazuya does not list a form for ‘centipede’ which could be used to cross-check Hudson.

In addition to the items above, the following words have been identified in individual languages which have either a *p or an *f reflex, but don’t have cognates in other languages (many of these were identified in Blust (2008) or in his online Austronesian Comparative Dictionary). Some of them with Indonesian equivalents may actually be loans:

(26) Further examples of *p

(a)	<u>Gloss</u>	<u>PMP</u>	<u>Bimanese</u>	
	to rasp	*parud	paru	(Ind <i>parut</i>)
	knife	*pisaw	pisu	(Ind <i>pisau</i>)
	other side	*hipaR	ipa	
	bitter	*paqit	paʔi	
	shoot, sprout	*qapucuk	pucu	(Ind <i>pucuk</i>)
	dew	*hapun	apu	
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	
	board	*papan	papa	(Ind <i>papan</i>)
	splash	*picik	pihi	
(c)	<u>Gloss</u>	<u>PMP</u>	<u>Dhao</u>	
	fulfill	*pa-penuq	pa-pənu	

(27) Further examples of *f

(a)	<u>Gloss</u>	<u>PMP</u>	<u>Nias</u>
	dove	*punay	[β]une
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Dohoi</u>
	tooth	*ŋipən	ŋihpo
(c)	<u>Gloss</u>	<u>PMP</u>	<u>Bimanese</u>
	astringent	*apeled	firi
	temple (anat)	*pispis	mpifi
	blow	*upi	ufi
	fern	*paku	fahu
	base of tree	*puqun	fuʔu
(d)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>
	vulva	*puki	uʔi

Within the Flores-Lembata group of WCMP, there are two closely-related dialects of the same language spoken in the villages of Seranggorang and Nilanapo which seem to distinguish between *p and *f intervocalically: [-p-] being the reflex of the former, and [-hp-] of the latter in a way which bears a striking similarity to Dohoi. Although our data is limited, we can offer the following examples:

(28) Examples of *-p- in Seranggorang-Nilanapo

<u>Gloss</u>	<u>PMP</u>	<u>Seranggorang</u>	<u>Nilanapo</u>
what	*apa	aape	ape
pat, light slap	*pik	api ¹⁶	api-ŋ
four	*epat	pa	pa

(29) Examples of *-f- in Seranggorang-Nilanapo

<u>Gloss</u>	<u>PMP</u>	<u>Seranggorang</u>	<u>Nilanapo</u>
fire	*hapuy	ahpe	ahpe
thin	*ma-nipis	mihpi	mihpi
tooth	*ipen	ihpe	ihpe

¹⁶ *api* < Proto-Lembata *gəpik ‘wing’

Regarding these two series (*p and *f) in the WCMP languages, Blust (2008: 93-4) states that “[...] much of this agreement is due to common conditioning. In particular, stops that otherwise undergo lenition in medial position tend to resist it when following a stressed schwa (PAN *ə), because this gave rise to phonetic gemination. Moreover, it is likely that the vocalic syncope in *ma- ‘stative’ before PMP bases that began with a labial stop either happened early in the history of the Central Malayo-Polynesian languages, or was recurrent, giving rise to prenasalized stops that also resisted lenition, but were subsequently reduced to simplex stops in some daughter languages.”

However, it seems evident from the evidence above that neither of these conditions holds true absolutely. Not only are there exceptions to the rule that *p was protected after a schwa, there is little evidence for the hypothesis that *p remained [p] only after stative *ma-.

Blust goes on to say “[...] the conclusion that *p lenition was independent in Hawu and Bimanese also follows from the simple fact of subgrouping: because Sumba-Hawu is a fairly clearcut group and Kambera shows no evidence of *p lenition, the simplest hypothesis is that *p lenition in Hawu must have occurred independently of similar changes in Bima.”

This makes the crucial assumption that reflexes of *f are the result of *p lenition. If, on the other hand, *f is assumed to be original, then it is languages in which *p and *f merged that were innovative.

Finally, “[t]he remaining phonological innovations that are shared exclusively by Hawu and Bima also fail to bear close scrutiny. Rather, parallel sound changes evidently have been unusually common in this part of Indonesia, and there are no clear grounds for using such evidence to propose a subgroup larger than that of Sumba-Hawu.”

We assert that the evidence presented here is actually demonstrative of shared retentions, and it is rather that parallel mergers have actually occurred in many other Austronesian languages, particularly Western Malayo-Polynesian but also including the Formosan languages.

3.0 The retroflex series

Evidence for a series of retroflex initials can be found in the WCMP languages Bimanese, Hawu and Dhao. This includes a voiceless retroflex stop *ʈ and a retroflex lateral *ɭ. Additional evidence for *ʈ can also be found in initial position in Nias, as well as medial position in Dohoi. With this series established, traditional PAN *j can be interpreted phonetically as a voiced retroflex stop *ɖ. Finally, there is tentative evidence that a retroflex nasal *ɳ may have been preserved in Malay and the languages of northwest Borneo.

3.1 Evidence for *ʈ

The two series of reflexes of PMP *ʈ are repeated below:

Table 7: Reflexes of PMP *ʈ and *ɭ

PMP *ʈ	RMP *ʈ	Nias t-	Dohoi -t-	Bima t	Hawu t	Dhao t
*ʈ	*ɭ	d-	-ht-	d	d	ɖ

There is no apparent conditioning environment for the latter set of reflexes; given the correspondences above, it appears that a distinction needs to be reconstructed between *ʈ and *ɭ. Examples of *ʈ are given below:

(30) Examples of PMP *ʈ in initial position

Gloss	PAN	PMP	Nias	Bima	Hawu	Dhao
natural cockspur	---	*tara	([ndr]ara)	tara	tara	(ɖara)
head covering	---	*teduŋ	---	toɖu	təɖu	təɖu
three	*telu	*telu	təlu	tolu	təlu	təlu
pierce	---	*tubak	---	tuba	təbu	---

knock, peck	*tuktuk	*tuktuk	---	tutu	tutu	---	
pole for boat	*tuku	*tuku	---	tuku	tuku	ta-tuku	
old (people)	*tuqaS	*tuqah	---	tua	---	tua	(Ind <i>tua</i>)
knee	---	*tuqud	---	ta-tuʔu	tuu	tuu	

(31) Examples of PMP *t in medial position

Gloss	PAn	PMP	Dohoi	Bima	Hawu	Dhao
calf	---	*b[e/i]ties	botih	wisi	---	---
cut, chop	*Setek	*hetek	(n-ohtok)	---	əta	əta

Additional cases in Bimanese and Hawu-Dhao are given below:

(32) Examples of initial *t in Bimanese

Gloss	PAn	PMP	Bima	
not	---	*ati	ati	
flick, snap fingers	*betik	*betik	beti	
explode	---	*betuʔ	wotu	
roof thatch	---	*qatep	ati	
low, of tide	---	*ma-qeti	moti ‘sea’	
brain	---	*qutek	iti	
alum	---	*tawas	tawa	(Ind <i>tawas</i>)
punting pole	---	*teken	tiki ‘staff, cane’	
pillar, mast, post	---	*tiqaŋ	tija	(Ind <i>tiaŋ</i>)
oyster	---	*tiRem	tire ¹⁷	(Ind <i>tiram</i>)

(33) Examples of initial *t in Hawu-Dhao

Gloss	PAn	PMP	Hawu	Dhao	
split	*betak	*betak	wəta	---	
star	*bituqen	*bituqen	---	hətu	
earthworm	---	*kali-wati	kelati	---	
flea	*qatimela	*qatimela	teməla	---	
placenta	---	*tabuni	tewuni	---	
round shield	---	*tamiŋ	tami	---	(Ind <i>tameŋ</i>)
middle	---	*teŋaq	təŋa	---	(Ind <i>təŋah</i>)
knock, beat	---	*tetuk	dətu	dətu	
to blow	---	*tiup	tiu	tiu	
year	---	*taqun	təu	təu	(Ind <i>tahun</i>)
look upward	---	*tiŋadaq	teŋara	teŋara	
bend, curve	---	*-tuk	ŋətu ¹⁸	---	

Upon comparison with PAn reconstructions, there appears to be a strong correlation between what we reconstruct here as *t and what is normally reconstructed as PAn *C. This is surprising, as evidence for *C has thus far been restricted to the Formosan languages:

¹⁷ This is likely an Indonesian loanword since the normal reflex of *R in Bimanese is [0].

¹⁸ ‘bow, curtsy’

(34) Examples of initial *ʈ

Gloss	PAn	PMP	Nias	Bima	Hawu	Dhao	
afraid	*[C/t]akut	*takut	(a-taʔu)	dahu	me-daʔu	ma-ɖaʔu	
rope	*CaliS	*talih	(tali)	dari ¹⁹	dari	ɖari	(Ind <i>tali</i>)
bury	*Canem	*tanem	[d]anə	---	pe-dana	pa-ɖane	
earth	---	*taneq	[d]anə	dana	---	---	
feces	*Caqi	*taqi	[d]ai	(taʔi)	dei	ɖei	(Ind <i>tahi</i>)
indigo	---	*taRum	---	dau	dao	---	
sea	---	*tasik	---	dasi ²⁰	dahi	ɖasi	
person	*Cau	*tau	---	dou	dau	ɖau	
sugarcane	*CebuS	*tebuh	---	doɓu	dəɓu	ɖəɓu	
swallow	---	*telen	(tələ)	---	dəla	ɖəle	(Ind <i>təlan</i>)
mat	---	*tepiR	---	dipi	dəpi	ɖəpi	
elder	---	*tua	[d]ua ²¹	dua ²²	---	---	
palm wine	---	*tuak	---	(tua)	due	ɖua	(Ind <i>tuak</i>)
index finger	*Cuzuq	*tuzuq	[d]uru	(turu) ²³	---	---	

(35) Examples of medial *ʈ

Gloss	PAn	PMP	Dohoi	Bima	Hawu	Dhao	
tree, log	---	*bataŋ	bahtaŋ	(bata)	---	---	(Ind <i>bataŋ</i>)
blind	*buCa	*buta	---	mbuda	ɓədu	bədu	
send	*pa-kaCu	*pa-katu	---	ŋgadu	---	pa-aɖu	
louse	*kuCu	*kutu	kuhtu	hudu	udu	uɖu	
eye	*maCa	*mata	mahta	mada	mada	maɖa	
die	*m-aCay	*m-atay	mahtoi	made	made	maɖe	
white	---	*putiq	puhti	---	pudi	puɖi	
ghost	*qaNiCu	*qanitu	(otu) ²⁴	---	nidu	niɖu	(Ind <i>hantu</i>)
liver	*qaCay	*qatay	ahtoi	ade	ade	---	
cucumber	---	*qatimun	---	dimu	dimu	---	
egg	*qiCeluR	*qatelur	(kotoruh)	dolu	dəlu	ɖəlu	
black	---	*ma-qitem	(mitom)	---	mədi	məɖi	
hundred	*RaCus	*Ratus	rahtus	(ratu)	---	---	(Ind <i>ratus</i>)
porcupine	---	*taRutuŋ	(tohotuŋ)	dudu	dudu ²⁵	ɖuɖu	

Of the four Dohoi exceptions above, two of them (‘ghost’ and ‘black’) descend from original prenasalized stops: Proto West Barito *həntu and *mintəm, respectively. As noted above, the phenomenon of Dohoi preaspiration is confined to the onset of the final syllable of the word, which explains the reflex in ‘egg’. As for ‘porcupine’, the WCMP forms allow for the reconstruction of an initial *ʈ, but there is a discrepancy with Dohoi as to the place of articulation of the second stop. Since this example is drawn from Hudson’s data and Kazuya does not provide an example of ‘porcupine’ to provide a crosscheck, it may be an example of Hudson failing to record the preaspiration. Bimanese ‘tree, log’ and ‘hundred’ are probably borrowed from Indonesian (the [r] in Bimanese *ratu* confirms this in the second case).

Additional examples of *ʈ are given below:

¹⁹ ‘strap’

²⁰ ‘beach’

²¹ ‘grandfather’

²² ‘parent’s sibling’

²³ ‘point’

²⁴ ‘corpse’

²⁵ The Hawu-Dhao forms both mean ‘thorn’

(36) Examples of *t̚ in Nias

<u>Gloss</u>	<u>PAn</u>	<u>PMP</u>	<u>Nias</u>
taro	---	*tales	[d]alə
ear	*Caliŋa	*taliŋa	[d]aliŋa
finger, toe	---	*taŋan	[d]aŋa
bone	*CuqelaN	*tuqelan	[d]əla

(37) Examples of *t̚ in Bimanese

<u>Gloss</u>	<u>PAn</u>	<u>PMP</u>	<u>Bima</u>
blunt tip	---	*tampak	dampa
blunt	---	*tumpul	dumpu
callus, blister	*beCuʔ	*betuʔ	wodu ²⁶
testicles	*buCuq	*butuq	wudu ²⁷
kill	*ka-aCay	*ka-atay	hade
outsider	---	*qaRta	ada ²⁸

(38) Examples of *t̚ in Hawu-Dhao

<u>Gloss</u>	<u>PAn</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>
beat on with both hands	---	*tambak	daba	---
come, arrive	---	*tekas	dəka	ɖəka
ear (wax)	---	*tilu	dilu	ɖilu
above	*i aCas	*di atas	dida	deɖa
haft of knife	---	*utiŋ	udi	---

(39) Examples of *t̚ in Dohoi

<u>Gloss</u>	<u>PAn</u>	<u>PMP</u>	<u>Dohoi</u>
sister	---	*betaw	bohtou
that	*i-Cu	*i-tu	ihtuh ‘this’, ahtuh ‘that’

There is a subset of forms in which *t̚ corresponds to PAn *t:

(40) Correspondences between PAn initial *t and PMP *t̚

<u>Gloss</u>	<u>PAn</u>	<u>PMP</u>	<u>Nias</u>	<u>Bima</u>	<u>Hawu</u>	<u>Dhao</u>
fish poison	*tuba	*tuba	[d]uβa	duwa	---	---
freshwater eel	*tuNa	*tuna	---	duna	dəno	ɖəno
southeast wind	*timuR	*timur	---	---	dimu	ɖimu

(41) Correspondences between PAn medial *t and PMP *t̚

<u>Gloss</u>	<u>PAn</u>	<u>PMP</u>	<u>Dohoi</u>	<u>Bima</u>	<u>Hawu</u>	<u>Dhao</u>
raw	*ma-qetaq	*ma-qetaq	---	mada	mada	---
stone	*batu	*batu	bahtu	wadu	wadu	waɖu
see	*(k)ita	*(k)ita	---	eda	-ədi	-əɖi
we (excl)	*kita	*kita	ihto	(ndai-ta)	di	əɖi
vomit	*um-utaq	*um-utaq	ŋ-uhta	---	mədu	məɖu
seven	*pitu	*pitu	pihtu	pidu	pidu	piɖu
penis	*qutiN	*qutin	---	---	udi	---

²⁶ ‘swell up slowly’

²⁷ ‘penis’

²⁸ ‘slave’

If we are accurate in interpreting PAN *C as *t̚, then it must be assumed that in the examples above, *t̚ either shifted to *t in the Formosan languages, or *t shifted to *t̚ in PMP²⁹.

There are also five cases in which PAN *t̚ corresponds to PMP *t:

(42) Correspondences between PAN *C and PMP *t

Gloss	PAN	PMP	Nias	Bima	Hawu	Dhao	
foxtail millet	*beCej	*betej	---	witi	---	---	
weep	*Caŋis	*taŋis	---	---	taŋi	---	(Ind <i>taŋis</i>)
know	*Caqu	*taqu	---	---	tou	təu	(Ind <i>tahu</i>)
guts	*C-in-aqi	*tinaqi	---	---	tenei	tanei	
roast, burn	*CuNuh	*tunu	tunu	---	tunu	tunu	(Ind <i>tunu</i>)

Of the above, it is possible that the forms for ‘weep’, ‘know’, and ‘roast, burn’ are Indonesian loans.

3.1.1 Evidence from Seranggorang-Nilanapo

As in the case of *f (see (28) and (29) above), the retroflex stop *t̚ has been preserved in medial position (in the reflex [-ht-] as opposed to the [-t-] reflex of *t) in Seranggorang-Nilanapo. Forms such as the following show that medial [-t̚] is contrastive with [-ht-] in Seranggorang-Nilanapo in a way similar to Dohoi:

(43) Seranggorang-Nilanapo reflexes of medial *t̚

Gloss	PAN	PMP	Seranggorang	Nilanapo
head	---	---	ette	ette
bad	---	---	dateŋ	datə
black	*ma-qitem	*ma-qitem	mite	mitaŋ

(44) Seranggorang-Nilanapo reflexes of medial *t̚

Gloss	PAN	PMP	Seranggorang	Nilanapo
louse	*kuCu	*kutu	uhtu	---
die	*maCay	*matay	mahta	mahta
stone	*batu	*batu	wahto	wahto

Note that Seranggorang-Nilanapo agrees with the Dohoi and WCMP forms for ‘stone’ in presenting evidence for a retroflex stop, contrary to evidence from the Formosan languages for an alveolar stop. The reflexes for ‘black’, on the other hand, indicate PMP *t̚ in contrast with the Hawu and Dhao forms in (35) which have reflexes of *t̚.

3.2 PAN *d̚

With evidence presented for the reconstruction of the voiceless retroflex stop above, attention can now be turned to its voiced counterpart. The PAN phoneme which seems most naturally to fill this gap is PAN *j. The distributional properties of *j support the idea that it was originally a retroflex. It is well-known that the acoustic properties of retroflexes are strongest in the vowel formants preceding the stops themselves (Ladefoged and Maddieson 1996: 28), and that word-initial retroflexes are therefore most prone to neutralization; this is exactly the pattern of PAN *j, which occurs only in medial and final position.

In Bimanese, the reflex of PAN *d̚ is *r* when following a low vowel, and *l* elsewhere, with the additional constraint that if the word begins with an *l*, the reflex becomes *r* via dissimilation (‘batten of loom’ in (45) and ‘day’ in (46)). In Macro-Sumba, *d̚ merged with *d and lenited to [r]:

²⁹ This may have happened, for example, in the case of *pitu ‘seven’, due to anticipatory assimilation of the retroflex feature from the retroflex lateral in ‘eight’ *waŋu (see section 3.3 below) in a counting sequence. It is presently unclear what the explanation might be for the discrepancy between Formosan *t̚ and RMP *t̚ in the other forms above.

(45) Examples of PMP *ɖ (*j)

Gloss	PMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao
to paddle	*aluja	[g]aluxa	---	---	---	---	---
batten of loom	*balija	---	---	lira	*lira	---	---
dry, evaporate	*maja	---	---	---	*mara	---	ɖasi mara ³⁰
foam	*bujaq	---	---	---	*βura	---	---
name	*ŋajan	---	aran	ŋara	*ŋara	ŋara	ŋara
rice	*pajay	[β]axe	paroi	fare	*pare	are	are
how much/many	*pija	---	pira	pila	*pira	pəri	pəri
when?	*p-ijan	---	mira	---	*piran	pəri	pəri
nose	*qijuŋ	[n]ixu	uruŋ	ilu	*iru	---	---
ant	*sejem	---	---	sa-sili	---	---	---
to smell	*hajek	---	-arok	---	---	---	---
younger sibling	*huaji	[n]axi	---	ari	*ari	ari	ari

In words where *ɖ followed schwa, it was reinterpreted as an implosive in PMS, likely because of the common feature of retraction shared by retroflexes and implosives (see Haudricourt 1950, as well as Heine 1978 in which an identical shift of *ɖ > [ɖ] is posited between Proto-Sam and the Cushitic language Boni). With the exception of ‘day’, these forms also have a reflex of [l] in Bimanese:

(46) PMP *ɖ > PMS *ɖ

Gloss	PMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao
day	*qalejaw	---	ondou	liro	*ləɖo	loɖo	loɖo
gall	*qapeju	[g]aβəxu	---	folu	*pəɖu	əɖu	əɖu
sharp pain	*hapejes	[β]əxə	---	pili	---	pəɖa	pəɖa
stinging pain	*hapejiq	---	---	---	---	pəɖi ³¹	pəɖi

Interestingly, the same change with the same conditioning environment occurred in the languages of northern Borneo. Examples from Proto Northern Sarawak (PNS) and Proto Sabahan (PSab) are given below (Borneo data are taken from Blust (1995b, 1997a, 1997b, 1998, 2000, 2001, 2002b, 2006b, 2007a, 2010)):

(47) Examples of PMP *ɖ > *ɖ in languages of northern Borneo

Gloss	PMP	PNS	PSab
inform	*bajaq	*badaʔ	---
foam	*bujaq	*budaʔ	*budaʔ
chills	*dajem	*dadəm	---
younger sibling	*huaji	*(t)-(w)adi	(*adi)
when	*ijan	*idan (PK)	*ŋ-idan
how (much/many)	*kuja	*kuda	*kuda
name	*(ŋ)ajan	*(ŋ)adan	*ŋadan
riceplant	*pajay	*padaj	*padaj
how (much/many)	*pija	---	*pida
charcoal	*qajeŋ	*adəŋ	*adəŋ

(48) Examples of PMP *ɖ > *ɖ in languages of northern Borneo

Gloss	PMP	PNS	PSab
sting, smart	*hapejes	*pəɖəs	*pəɖəs
stinging pain	*hapejiq	*pəɖiʔ	*pəɖiʔ
day	*qalejaw	*əɖaw	*əɖaw

³⁰ ‘low tide’

³¹ PHD ‘itch’

gall bladder	*qapeju	*pədu	*pədu
ant	*sejem	---	*sədəm

Although PAN *ɖ was restricted to non-initial position as discussed above, there is reason to believe that the word ‘two’ had an original and exceptional retroflex initial in PWCMP. The evidence for this comes from PMS, in which ‘two’ can be reconstructed with an implosive (*ɖua), and from Proto Lembata, in which ‘two’ is reconstructed with a retroflex initial that is normally only reconstructed word-medially (the reflex of PAN *j):

(49) PWCMP evidence for initial *ɖ in ‘two’

<u>Gloss</u>	<u>PMP</u>	<u>PMS</u>	<u>PLembata</u>
two	*duha	*ɖua	*ɖua

3.3 PAN *l

Having postulated the existence of a retroflex stop series in PAN, we turn to the question of whether evidence exists for other retroflex phonemes. There is indeed evidence for a retroflex lateral.

3.3.1 PPS evidence for *l

The WCMP languages exhibit another split series of reflexes for PAN *l; Paz (1981) reconstructs a distinction between front and back laterals for Proto Philippines:

Table 8: WCMP reflexes of *l and *ɭ

PMP	RMP	PPh	Bima	Dhao	Hawu
*l	*l	l	l	l	l
*ɭ	*ɭ	ɭ	r	r	r

Since *r can be reconstructed independently for PAN, these series can be reconstructed as alveolar *l and retroflex *ɭ respectively. In comparison with the WCMP data above, the correspondence between *l and *ɭ is nearly exact. Examples of alveolar *l and retroflex *ɭ in Proto Philippines (PPh) are given below:
 Examples of alveolar *l are given below:

(50) Examples of *l

<u>Gloss</u>	<u>PMP</u>	<u>PPh</u>	<u>Bima</u>	<u>Hawu</u>	<u>Dhao</u>	
return	*baliw	---	ɓali ³²	ke-ɓali ³³	---	(Ind <i>balik</i>)
buy	*beli	*bəlɨʔ	weli	wəli	həli	
seaward	*lahud	---	ka-lau ³⁴	lou	ba-ləu ³⁵	(Ind <i>laut</i>)
sail	*layaR	---	lodʒa	lai	lai	(Ind <i>lajar</i>)
fold	*lepet	---	lipi	ləpa	ləpe	
five	*lima	*limáʔ	lima	ləmi	ləmi	(Ind <i>lima</i>)
day	*qalejaw	---	liro	lodɔ	lodɔ	
egg	*qateluR	*ʔiklúg	dolu	dəlu	ɖəlu	
three	*telu	(*ta-t[úʔ])	tolu	təlu	təlu	
ear(wax), deaf	*tuli, *tilu ³⁶	*tulɨʔ	---	dilu	ɖilu	

Additional examples of *l are given below:

³² ‘restore’

³³ ‘again’

³⁴ ‘unobstructed view’

³⁵ ‘south’

³⁶ This comparison assumes vocalic metathesis in either Proto Philippines or PWCMP.

(51) Examples of *l

(a)	<u>Gloss</u>	<u>PMP</u>	<u>PPh</u>		
	tongue	*dilaq	*dílaʔ		
	back	*likud	*likúd		
	neck	*liqeR	*liʔáǵ ^j		
	nit	*lisehaq	*lisáʔ		
	ear	*taŋila	*talíŋaʔ		
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>		
	broad, wide	*abelaj	wela		
	count	*bilaŋ	bila	(Ind <i>bilaŋ</i>)	
	lamp, torch	*ilaw	ilo		
	lightning	*kilat	kila	(Ind <i>kilat</i>)	
	exceed	*lalaw	lalo		
	walk, go	*lampañ	lampa ³⁷		
	sky	*laŋit	laŋi	(Ind <i>laŋit</i>)	
	step, stride	*laŋkaŋ	laŋga	(Ind <i>laŋkah</i>)	
	passageway	*lawañ	lawa	(Ind <i>lawañ</i>)	
	calm (water)	*linaw	lino		
	eel sp.	*linduŋ	lindu		
	leech	*lintaq	linta	(Ind <i>lintah</i>)	
	boil over	*luab	lua ³⁸	(Ind <i>luap</i>)	
	outside	*luqar	lua	(Ind <i>luar</i>)	
(c)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>	
	saliva	*iluR	ilu	ilu	
	men's house	*kamali	kemali ³⁹	---	
	fold	*leku(q)	ləku	ka-ləko	
	cave	*liaŋ	lie	---	(Ind <i>liaŋ</i>)
	sour	*nilu	me-nilu	ma-nilu	(Ind <i>nilu</i>)
	dog flea	*qati-mela	teməla	---	
	wing	*qelad	əla	əla	
	to swallow	*telen	dəla	dəle	

Examples of retroflex *l̥ are given below:

(52) Examples of *l̥ in WCMP

<u>Gloss</u>	<u>PMP</u>	<u>PPh</u>	<u>Bima</u>	<u>Hawu</u>	<u>Dhao</u>	
reverse	*balik	---	wari	wari	hari ⁴⁰	
widow(er)	*balu	---	mbaru	(balu)	(balu)	(Ind <i>balu</i>)
moon	*bulan	*búlan	wura	wəru	həru	
ear of grain	*buliR	(*búlig ⁴¹)	wuri	wuri	huri	
body hair	*bulu	*bu bú	wuru	wuru ⁴²	---	
in(side)	*dalem	*dáləm ⁴³	---	dara	dara	
road	*zalan	*da an	---	fara	fara	

³⁷ 'road, run'

³⁸ 'pour'

³⁹ 'house(hold)'

⁴⁰ 'again'

⁴¹ 'bunch, cluster'

⁴² '(root) fibers'

⁴³ 'depth, bottom'

skin	*kulit	---	huri	k-uri	ka-ʔuri	
rat	*labaw	---	ka-rawo	---	ma-raho	
man, husband	*laki	*laʎaki	rahi	(laʔi)	(laʔi)	(Ind <i>laki</i>)
ginger	*laqia	(*laʔújaʔ)	rea	---	---	
run away, flee	*laRiw	---	rai	rai	rai	
ten	*sa-ŋa-puluq	*púʎuʔ	mpuru	ŋuru	ŋuru	
pestle	*qahelu	*háʎʎuŋ	aru	aru	---	
head(waters)	*qulu	---	uru	uru	uru	
error	*salaq	---	sara	(hala)	(sala)	(Ind <i>salah</i>)
torch	*suluq	*suʎúʔ	---	huru	suru	
forest	*halas	*háʎas ⁴⁴	ara	---	ɖua ara ⁴⁵	
snake	*hulaR	*ʔuʎɔg	---	---	əru-əru ⁴⁶	
rope	*talih	---	dari	dari	ɖari	
eight	*walu	*waʎúʔ	waru	aru	aru	

Additional examples of *ʎ are given below:

(53) Examples of *ʎ

(a)	<u>Gloss</u>	<u>PMP</u>	<u>PPh</u>	
	house	*balay	*baʎəj	
	rolled together	*balun	*báʎun ⁴⁷	
	deaf	*bəŋəl	*biŋəʎ	
	itch	*gatel	*gatəʎ	
	mortar	*lusuŋ	*ʎasúŋ	
	moss	*lumut	*ʎúmut	
	leech	*qali-matək	*aʎimatək	
	sour	*qaRsem	*kaʎsəm	
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>	
	astringent	*apeled	firi	
	revolve, turn	*biliŋ	wiri	
	trivet	*dalikan	riha ‘hearth’	
	dig	*kali	ŋari (?)	
	sesame	*ləŋa	riŋa	
	Venus	*mantalaq	ntara ‘star’	
	to plant	*mula	mura	
	hand	*qalima	rima	
	centipede	*qalu-hipan	rifa	
	bamboo floor	*saleR	sari	
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>
	pale, white	*balaR	wara	---
	housefly	*ləlej	lara	lara
	withered	*ma-layu	me-rafu	---

The correspondence between the Proto Philippines and the WCMP forms is generally quite good – there are only three discrepancies above (‘three’, ‘ear of grain’, and ‘ginger’). In the case of ‘three’, it may be

⁴⁴ ‘wild; snake’
⁴⁵ ‘palmwine’
⁴⁶ ‘worm’
⁴⁷ ‘provisions’

the case that the lateral in Pre-Proto Philippines *təlu assimilated to the initial retroflex initial in ‘two’ *ɖuha in counting.

3.5 PAn *ŋ?

Blust (2006) describes a small set of forms found in the Philippines, southern Sulawesi, northwest Borneo, and Malay which have liquid reflexes in the former two but sibilant reflexes in the latter two. Examples are given below:

(54) Examples of liquid-sibilant correspondences in WMP

Gloss	Tagalog	Bikol	Cebuano	Maranao	Kelabit	Malay
foam	buláʔ	---	buláʔ	bolaʔ	---	busa
rotten	bulók	bulók	bulúk	---	---	busuk
to water	dilíg	---	---	---	---	diris
red	---	---	---	rigaʔ	siaʔ	---
long hair	lúgaj	lúgaj	---	---	---	suraj
compel	pílit	pírit	pílit	---	---	pisit
disorder	guló	---	---	---	---	rusoh
sell	---	---	---	saliw	m-asiw	---
provoke	ulót	urót	---	---	---	us[o/u]t
flower	---	---	búlak	---	busak	---
blind	búlag	---	bulúg	---	busəɾ	(bular)
tickle	---	gírok	gilúk	---	g-əl-isək	---

It is difficult to know how to interpret this data, and if it represents a legitimate proto-phoneme. Since the Philippine reflexes would normally reflect *j (*ɖ), some kind of retroflex is suggested. The gaps in the inventory include *ʂ, and *ŋ. *ʂ may seem possible at first glance, given the sibilant reflexes in Malay and the northwest Bornean languages, but the change from *ʂ to a liquid seems implausible. Although none of the reflexes include a nasal, *ŋ can be considered if the following changes are assumed:

Philippines/southern Sulawesi: ŋ > ʎ > l, r

Malay/NW Borneo: ŋ > z̥ > ʂ > s

Given the limited nature of the evidence, it is difficult to make a definitive statement, and this must remain a tentative hypothesis. One interesting corollary hypothesis, however, is that this would explain the unique change in the East Formosan languages of *j (*ɖ) to *n (see Blust (1999: 46)). Although there are no Formosan examples cited for the items above, *if* *ŋ was part of the PAn consonant inventory, then the following sequence of changes might be posited, in which East Formosan *ɖ merged with original *ŋ before shifting to *n:

*ɖ > *ŋ > n

3.6 Interim discussion on the retroflex series

One of the biggest reasons that the unity of the retroflex series has not been recognized is because of the different patterns of merger amongst its members. While *ɖ is reconstructed from a heterogeneous series of late mergers, *ʎ and *ʎ appear to have merged quite early in most PMP languages with *t and *l, respectively. In the case of the *l/*ʎ merger, it appears to have been motivated by the depalatalization of *ʎ to dental *ʎ (see below), which would have put a severe strain on perceptual contrast between the laterals. The same may have been true of the *t/*ʎ merger under pressure from *c. The western Indonesian evidence given in section 4.2.2 below suggests that the *c/*s merger occurred later than the *t/*ʎ merger, perhaps through a stage where it became a dental affricate. In this scenario, the following set of shifts and mergers would have occurred (in the non-Formosan languages):

c > tʂ > ʂ
 t, t̚ > t > t̚
 ʎ > ɺ > n
 l, l̥ > ɭ > l

Overall, it appears that within Austronesian, evidence for the retroflex series has been preserved best at two geographic extremes – in the northwest in Taiwan and the Philippines, and in the southeast in WCMP. The cumulative reflexes for the retroflex series in the Formosan languages are given below:

Table 9: Reflexes of retroflex consonants in the Formosan languages

PAn	*t̚	*d̚	*l̥/ɭ
Pazeh	s	z, -t	r
Saisiat	s	z	ɻ
Thao	θ	ð	r
Atayal	ts	ʔ	l
Saaroa	ts	ɸ	ʔ
Kanakanavu	ts	l	ʔ, l
Rukai	ts	g	ɭ
Bunun	t	ʔ	h-, ʔ
Amis	t	n	ɭ
Kavalan	t	n	r, ɣ
Paiwan	ts	d	ɭ
Puyuma	t̚	d̚/d	ɭ

The most conservative Formosan language appears to be Puyuma, which has maintained retroflex reflexes in all cases. It is notable that even though the distinction between *l̥ and *ɭ is not maintained in Taiwan, the Formosan languages still provide indirect evidence for the latter (as noted in Ross (1992)) since in the majority of cases, *l̥ and *ɭ seem to have merged as *ɭ, with modern languages showing a combination of lateral and rhotic reflexes.

4.0 Palatals

A distinct palatal series has already been recognized in PAn, which includes the members *c, *z (*j), *ɲ, and *N (*ʎ). These will be covered below in this order.

4.1 The PAn palatal stop

The reflexes of PAn *s and the *c proposed in this paper are given below:

Table 10: reflexes of *s and *c

PAn	RMP	Nias	Bima	Hawu	Dhao	PWOc
*s	*s	z-	s	h	s	*s
*s	*c	s-	c	h	c	*z

Although Dohoi distinguishes intervocalic [-c-] and [-hc-], one might expect it to pattern with the other key languages listed above in distinguishing *s from *c. However, comparative evidence shows that Dohoi [-c-] is a reflex of PMP *y while [-hc-] is a reflex of PMP *z:

(55) Dohoi [-c-] < *y and [-hc-] < *z

Gloss	PMP	Dohoi	Gloss	PMP	Dohoi
spit	*luzaq	ruhca	wood	*kayu	kacu
rain	*quzan	uhcan	crocodile	*buqaya	pacu (vocalic metathesis)
far	*zauq	ma-hcu			

There is more variation in this PWOC lenis series than the other two examined by Ross (*β < *f and *γ < *g), and as a result reconstructions are less secure.

(56) Examples of initial *s

Gloss	PMP	Nias	Bima	PSumba	Dhao	PWOC
error	*salaq	(sala)	sara	*sala	(sala)	--- (Ind salah)
bamboo floor	*saleR	[z]alo	sari	*ka-[s]al[e]	---	---
python	*sawa	[z]awa	sawa ⁴⁸	---	---	---
douse fire	*sebu	---	suwu	---	səbu ⁴⁹	---
ant	*sejem	[z]ixə	sa-sili	---	---	---
elbow	*siku	[z]iʔu	(cihu) ⁵⁰	*siyu	siʔu	---
torch	*suluq	[z]ulu	---	*[s]ulu	suru	--- (Ind suluh)
breast, milk	*susu	---	susu	*susu	---	*zuzu ⁵¹ (Ind susu)

(57) Examples of medial *s

Gloss	PMP	Bima	PSumba	Dhao	PWOC
iron	*besi	besi	*bəçi	bəsi	--- (Ind bəsi)
paddle	*beRsay	wese	*bōse	---	*βoze
contents	*isi	isi	*içi	---	--- (Ind isi)
cook	*nasu	---	*ma-na[s]u	nasu	---
navel	*pusej	---	*pə[s]u	əsu	---
breast, milk	*susu	susu	*susu	---	*zuzu (Ind susu)
ripe	*ta-esak	ntasa	---	ma-ɖasa	---
sea	*tasik	dasi	*tasik	ɖasi	---

Proto Sumba forms with an *[s] do not have a Wejewa or other northwest Sumba cognate, crucial to the distinction between Proto Sumba *ç and *s; these are therefore unfortunately ambiguous and merely listed for completeness. The two Proto Sumba exceptions above, ‘iron’ and ‘contents’, both occur before a final [i] which indicates that *s palatalized to *ç in this environment.

(58) Examples of initial *c

Gloss	PMP	Nias	Bima	PSumba	Dhao	PWOC
climb	*sakay	---	---	*[s]ayi	caʔe	*saye
bifurcation	*saŋa	---	caŋa ⁵²	*[s]aŋa	ka-caŋa ⁵³	*saŋa
comb	*saRu	---	cau	*[s]ai	---	---
nine	*siwa	siwa	ciwi	*çiwa	ceo	---
spoon, ladle	*suduk	---	ciru	*[s]uru	curu	---

(59) Examples of medial *c

Gloss	PMP	Bima	PSumba	Dhao	PWOC
wet	*baseq	bəca	*mba[s]a	(basa)	--- (Ind basah)
one	*isa ⁵⁴	ica	*iça	əci	---

⁴⁸ ‘snake’

⁴⁹ ‘smoke, vapor’

⁵⁰ Since all other examples of ‘elbow’ indicate *s instead of *c, this may have been an innovation in Bimanesic; however, see the Proto Dayic form below.

⁵¹ While *susu appears to be the dominant form, *suzu and *zuzu are also reconstructible based on various reflexes in daughter languages.

⁵² ‘split’

⁵³ ‘branch of a river’

⁵⁴ See also Nias *sara* ‘one’.

who	*i-sai	---	---	cee	*sai	
sated	*ma-besuR	mbocu	*ka-mbəçu bəcu	---		
salty	*ma-qasin	maci ⁵⁵	*maçi	(masi)	---	(Ind (m)asin)
mortar	*esuŋ	nocu	*ŋoçu	---	---	
heart	*pusuq	---	*puçu	(usu)	---	
smoke	*qasu	---	*maçu	---	*qasu	

Note that the following forms in Nias corroborate the hypothesis that in at least some instances, Nias [s] derives from an earlier [c] (the result of the palatalization of [t] before [i]):

(60) Nias si < *ci

<u>Gloss</u>	<u>PMP</u>	<u>Nias</u>		
white	*putiq	a-fusi	<	a-fuci
calf	*bities	[mb]isi	<	[mb]ici

These contrast with sequences of [ti] in Nias, which are largely, if not exclusively, the result of Indonesian loans, i.e. Nias *bufati* ‘regent’ < Ind *bupati*, Nias *gariti* ‘curly’ < Ind *kəritij*, Nias *lati* ‘train’ < Ind *latih*, etc.

Additional examples of *s and *c are given below:

(61) Examples of PMP *s

(a)	<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>	
	comb	*nusa	nisa	(Ind <i>nusa</i>)
	salt	*qasiRa	sia	
	pierce, stab	*rusuk	rusu	
	regret	*selsel	sinci	
	3sg	*si-ia	sia	

(62) Examples of PMP *c

(a)	<u>Gloss</u>	<u>PMP</u>	<u>Nias</u>
	comb	*sujud	suxu
	split	*silaq	sila
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>
	mercy, pity	*ma-qasiq	meci ⁵⁶
	sweep	*sapuh	cafi
	insert	*sipsip	cici
(c)	<u>Gloss</u>	<u>PMP</u>	<u>PWOc</u>
	nit	*lisehaq	*lisa

4.2 Proto Dayic

There is an additional correlation between PPS and Proto Dayic, a Northern Sarawak subgroup in northwest Borneo, where the following sound correspondences are observed:

Table 11: Correspondences between PPS and PD

PPS	PD
c	0/h
s	s

Where Proto Dayic forms are unavailable, Kelabit (K) forms are used instead. Examples of *s:

⁵⁵ ‘sweet’

⁵⁶ ‘love’

(63) Proto Dayic reflexes of *s

<u>Gloss</u>	<u>PMP</u>	<u>PDayic</u>
oar	*beRsay	bəsay (K)
heart	*pusuq	*pusu?
pierce, stab	*rusuk	n-usuk (K)
breast	*susu	tusu (K)
error	*salaq	sala? (K)

(64) Proto Dayic reflexes of *c

<u>Gloss</u>	<u>PMP</u>	<u>PDayic</u>
wet	*baseq	*baha?
sated	*besuR	*baur
mortar	*esuŋ	iuŋ (K)
who	*i-sai	*iih
mercy, pity	*ma-asiq	*m-ahe?
salty	*ma-qasin	m-ain (K)
navel	*pusej	*puəd
elbow	*siku	*ijuh
nine	*siwa	*iwa?
comb	*sujud	udud (K)

A crucial question can now be raised about the relationship between traditional PAN *c and the *c proposed in this paper. Since traditional *c is of such relatively low frequency⁵⁷ and retained in only a small set of western Indonesian languages, the extension of the domain of *c promises to solve a vexing problem in Austronesian comparative studies, summarized in the following table from Blust (2009b: 557):

Table 12: Arguments for and against assigning *c to PMP or PAN

Against	For
1. found only in western Indonesia	1. *c or massive unconditioned splitting
2. spread by borrowing from Malay	2. found in monosyllabic roots
3. not in basic vocabulary	3. patterns with other palatals (*z, *ɲ)

The particularly interesting point in the above table is that *c is “found in monosyllabic roots” in western Indonesia. We would add that it is not only found in monosyllabic roots, but in the prefixes attached to these roots as well. In order to illustrate this, we collected all of the words with reconstructed *c in the Austronesian Comparative Dictionary (Blust 1995e); the following are all of the instances in which *c appears outside of the final syllable. Forms in which there is an identifiable root are listed in (65), and those in which there is none in (66):

(65) PWMP words reconstructed with initial *c with identifiable roots

<u>Gloss</u>	<u>PWMP</u>	<u>Root</u>	<u>Root meaning</u>
take with the fingers	cu=bit	=bit	hook, clasp, grasp with fingers
shelter	ci(n)=duŋ	=duŋ	shelter, protect
quick, swift	ce=kas	=kas	swift, agile; energetic
choke, strangle	ce=kel	=kel	cough
seize, grasp	ce=kep	=kep	seize, grasp, embrace
catch with the hands	ci=kep	=kep	seize, grasp, embrace

⁵⁷ We have unfortunately not been able to uncover any cognate sets which include both forms with traditional PAN *c and forms with *c reconstructed from our data which would allow a correlation between the two to be (dis)proven.

sticky, stick to	ca(η)=ket	=ket	adhesive, sticky
cover with a hollow container	cu(η)=kub	=kub	cover
bent, curved	ci(η)=kuk	=kuk	bent, crooked
close, cover	cu(η)=kup	=kup	enclose, cover
to shine, of heavenly bodies	ci=lak	=lak	shine
sink, submerge	ce=leb	=leb	sink, disappear
soak, steep in water	ce=lep	=lep	sink, submerge
curved area; corner, angle	ce=luk	=luk	bend, curve
dive, immerse	ce(R)=neb	=neb	dive; sink, disappear underwater
immerse, submerge	ce=ŋeb	=ŋeb	dive, submerge
gape, open the mouth	ci=ŋaŋ	=ŋaŋ	amazed, gaping
seize in the mouth or beak	ce=ŋap	=ŋap	open, of the mouth
catch one's breath	ci=ŋap	=ŋap	open, of the mouth
to smack	ca(m)=pak	=pak	clap, slap; split, break
crack, split, break	ce=pak	=pak	clap, slap; split, break
pincers, tongs	ca(R)=pit	=pit	press, squeeze together; narrow
narrow	cu=pit	=pit	press, squeeze together; narrow
dull sound	ce=puk	=puk	thud, clap, break
part, separate, divide	ce=ray	=ray	separate
spotted, speckled	cu=rik	=rik	spot, freckle
mark, line, dash	cu=rit	=rit	scratch a line

(66) PWMP words reconstructed with initial *c without identifiable roots

<u>Gloss</u>	<u>PWMP</u>
rice variety (?)	candur
betel nut case	calapa?
small bird with shiny black plumage: <i>drongo</i>	cawi
stench	ceŋis
shrill, of the voice	cerik

Of the five words in (66) above, ‘stench’ and ‘shrill’ “look” like they could be prefix+root combinations, although no roots have yet been identified. Of the remaining three, ‘rice variety’ is a questionable reconstruction, ‘betel nut case’ is an item that could have been borrowed widely through trade, and ‘drongo’ is a faunal term (faunal terms are often ideophonic or known to be phonologically idiosyncratic in other ways (Smoll 2011)).

The next set of words is comprised of reduplicated roots:

(67) PWMP words reconstructed with *c with reduplicated roots

Gloss	PWMP
hissing, rustling sound	cik
skewer	cuk=cuk
insert	cu(l)=cul
sip, suck	cup=cup
suck noisily	cut=cut

Finally, the following are forms in which *c occurs in the penultimate syllable of the word. (68) lists all forms with identified roots, and (69) lists exceptions:

(68) PWMP words reconstructed with medial *c with identifiable roots

<u>Gloss</u>	<u>PWMP</u>	<u>Root</u>	<u>Root meaning</u>
muddy, waterlogged	ba=cak	=cak	muddy
muddy	bi=cak	=cak	muddy
step, tread on	i=cak	=cak	step, tread, trample
drive in, as a post, nail, or wedge	pa=cek	=cek	insert, stick into a soft surface
press in by force	pu=cek	=cek	insert, stick into a soft surface
splatter, fly out in all directions	be=cik	=cik	fly out, splash, splatter
splash, splatter	bi=cik	=cik	fly out, splash, splatter
fly off, of solid bits or droplets	le=cik	=cik	fly out, splash, splatter
splash, splatter	peR=cik	=cik	fly out, splash, splatter
splash, spray, sprinkle	pi=cik	=cik	fly out, splash, splatter
sound of splashing, etc	ri=cik	=cik	fly out, splash, splatter
speckled, stained	re=cik	=cik	fly out, splash, splatter
dent, dented	pi=cuk	=cuk	penetrate
hoe, chop up soil	ba=cuk	=cuk	penetrate
enter, penetrate	ha=cuk	=cuk	penetrate
squeeze, squirt out	be=cit	=cit	squirt out
squeeze, squirt out	pe(R)=cit	=cit	squirt out
squirt or gush out	pu(R)=cit	=cit	squirt out
squeeze out	pi=cit	=cit	squirt out
squeeze, squirt out	le=cit	=cit	squirt out
squeeze out, slip out	le=cut	=cut	squirt, squeeze, or slip out
squeeze, squirt out	lu=cut	=cut	squirt, squeeze, or slip out
spurt out	pu=cut	=cut	squirt, squeeze, or slip out

(69) PWMP words reconstructed with medial *c without identifiable roots

<u>Gloss</u>	<u>PWMP</u>		
hook used in lifting	gancu	sexually impotent	baciR
crush, press in	picak	budge, shift, move aside	icud
dove, pigeon	acaŋ	mountain peak	qapucuk
disturb	kacaw	mouth (of an animal)	mu(n)cuŋ
rotten, spoiled	bucek	slip or slide down	lu(ŋ)cuR
smooth, slippery	licin	take away from, dispossess	pucut
creep, crawl	icir		

There are more counterexamples in this group of words, but they are still outnumbered by those with identifiable roots. Several of the above forms also “look” like they could be prefix+root combinations; this list also includes ‘dove, pigeon’, another faunal term.

Given the evidence above, it can now be postulated that *c was originally more widespread than previously thought, not restricted to western Indonesia, and that it occurred in several basic vocabulary items. The question then becomes how to interpret the data from western Indonesia.

What seems to have happened is that evidence for *c has been preserved most conservatively in WCMP, Nias, Proto Dayic, and (erratically) in PWOC. Conversely, *c generally merged with *s in the languages of western Indonesia if they occurred in monomorphemic forms; it was preserved if the forms were bimorphemic, indicating a sensitivity to root structure (and the implication that it was still active when *c shifted to *s). It should be noted that this provides potentially valuable subgrouping evidence for a group that includes several Malayic languages, as well as languages spoken across the southern Indonesian chain of islands stretching from Sumatra to Lombok.

The fact that what is currently reconstructed as PMP *s derives from an earlier *c and *s may also serve to explain why the reflex of PAN *s in some languages (such as Kakanabu, Bunun, Amis, or Manggarai) is *c* ([ts] or [tʃ]) – the existence of original *c meant that there was a pre-existing category into

which *s could merge. Whereas the normal pattern of merger in the majority of languages was *s, *c > s, the reverse pattern, *s, *c > c occurred in a minority of languages.

4.3 The voiced palatal stop

Blust (2009: 547) interprets PAN *z as a voiced palatal affricate, while Wolff (2010: 34) suggests that it was “a very forward voiced stop.” We consider it most parsimonious to assume that PAN *z was the voiced counterpart of *c, namely the voiced palatal stop [ʝ].

4.4 The palatal fricative

Blust interprets PAN *s phonetically as the palatal fricative [ç] and *S as [s] (Blust 2009b: 547). Wolff (2010) agrees about *S but interprets *s as [c]. We suggest that PAN *s was [s], that *S was [ç], and suggest that *s remained stable in many languages with *ç shifting to *h* in all but some of the Formosan languages.

One of the reasons why Blust interprets *s as [ç] is because “in languages that preserve *ɲ as a palatal nasal and that have active systems of nasal substitution *s* is replaced by *ɲ* in certain word-formation processes [...]” (2009b: 579). However, this could also be because of the fact that Proto-Austronesian *s was a dental sibilant with a domed (as opposed to grooved) tongue articulation, which would be a phonetic driver for a palatal nasal even if *s was not phonologically a member of the palatal series. Where a grooved tongue configuration would pull the tip of the tongue toward the alveolar ridge, a domed configuration would push it into a dental position, promoting interdentalization (Rukai) as well as affrication (Kanakanavu, Bunun and Amis) which would be an intermediate stage on the way to becoming a plain stop (Thao, Paiwan). This slot would then be available to be filled by the palatal fricative [ç] as it shifted forward.

(70) Formosan reflexes of PAN *s and *S (*ç)

PAN	Pazeh	Thao	Sai	Ata	Saa	Knknvu	Rukai	Bunun	Amis	Kav	Puy	Pai
*s	z	t	h	h, x	s, 0	c	θ, s, 0	c	c	s	s	t
*S	s	ʃ	ʃ	s	0	s	s	s	s	s	0	s

4.5 The palatal lateral

Evidence for PAN *N has been preserved, especially in word-medial position, most robustly in the Formosan languages – many of which have also retained a phonetic realization that supports its reconstruction as a palatal lateral *ɺ; this is supported by the fact that *ɲ has merged with *ɺ in all Formosan subgroups except for Tsouic (Blust 2009: 579).

The question of how exactly PAN *ɺ merged with *n in PMP is a thorny one. It would be simplest to argue that *ɺ first shifted to *ɲ which then merged with *n, but that leaves the question of why there are still etyma that can be reconstructed with *ɲ which did not undergo this shift. Since there is evidence that *t and *s were dental [t] and [s̪] (contrasting with alveolar *d, *n, and *l) (Blust 2009b: 552), it can be suggested that *ɺ passed through a stage where it depalatalized to a dental *l, remaining distinct from *ɲ (as well as alveolar *l) before its ultimate merger with *n.

There is reason to think that the shift to *n was not abrupt. This general trend toward de-palatalization is apparent in the variation which occurs in the handful of words that can be reconstructed with initial *ɺ and in which various non-Formosan languages preserve an initial [l]:

(71) Examples of PMP variation indicating original *ɺ

Gloss	PMP
swim	*[l/n]aŋuy
fish with poisonous dorsal spines	*[l/n]epuq
wound	*[l/n]uka
mosquito	*[l/n]amuk

Examples of languages that retain [l] in these forms, selected from Wolff (2010), are given below:

(72) Examples of PMP forms which show reflexes of *ɺ

<u>Gloss</u>	<u>Tagalog</u>	<u>Selayar</u>	<u>Malagasy</u>	<u>Malay</u>	<u>Javanese</u>	<u>T. Batak</u>	<u>Moken</u>
swim	laŋoj	laŋe	lano	---	laŋuj	---	(naŋoj)
pois. fish	---	---	---	lepu	---	---	lepuuk
wound	---	loko	loka-na	luka	---	luha	loka?
mosquito	lamok	---	---	(ŋamuk)	lamuk	---	(ŋamok)

There is another small group of forms which are more ambiguous, where Formosan languages provide evidence for *ʎ, but non-Formosan languages reconstruct to *l:

(73) Examples of variation between Formosan *ʎi- and PMP *li-

<u>Gloss</u>	<u>PAn/PMP</u>
sap	[N/I]iteq
water leech	qa[N/I]i-meCaq
den, pen	[N/I]ibu
bail out	[N/I]imas

Two interpretations of these data are possible. Either *ʎ was reinterpreted as *l before [i] in the PMP languages, or *l palatalized before [i] in the Formosan languages. A choice between these two options can't be made without more data, but this variation strengthens the interpretation of PAn *N as a palatal lateral more globally.

5.0 The distinction between *k and *g

Similar to the cases of *p and *t, there are two sets of reflexes associated with PAn *k:

Table 13: Reflexes of *k and *g

PMP	RMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao	PWOc
*k	*k	k	-k-	k	k	k	k	*k
*k	*g	g-, -ʔ-	-hk-	h	ɣ	0, -ʔ-	0, -ʔ-	*ɣ

We propose that the second set of reflexes associated with PAn *k be reconstructed as *g. These reflexes occur in all positions in Bimanese, Hawu-Dhao, and Western Oceanic; and in this case Nias shows this split in both initial and medial positions, although Dohoi continues to exhibit it in medial position only.

The development of the plain voiced stops in many WCMP languages is completely symmetrical, whereas it was asymmetrical in many other Austronesian languages. In this way, the original three-way contrast between voiceless, voiced, and implosive stops⁵⁸ was maintained. The majority of other Austronesian languages, conversely, experienced mergers of the bilabial and alveolar plain and implosive voiced stops on the one hand, and the plain voiced and voiceless velar stops on the other (the remaining palatal voiced stop then often merging with the voiced alveolar stop, although its development was more varied). We use the typical development of the WCMP languages as an example:

<u>WCMP</u>			<u>Other languages</u>		
ɓ	>	ɓ	ɓ	>	b
b	>	β	b	>	b
ɗ	>	ɗ	ɗ	>	d
d	>	r	d	>	d
k	>	k	k	>	k
g	>	ɣ	g	>	k

Examples of *k are given in the tables below:

⁵⁸ On the subject of Austronesian implosive stops, see Norquest & Downey (in preparation).

(74) Examples of *k in initial position

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
bite	*kaRat-i	---	---	*kati	kadi	kaɟi	
scratch	*kaRaw	---	kao	*kaʔu	kao	kao	
stand	*kideŋ	---	kiɟi	---	kəɟi	kəɟi	
belly	*kempuŋ	---	---	*kambu	kabu ⁵⁹	---	
pinch together	*kepit	---	kapi	*kəpit ⁶⁰	---	---	(Ind <i>kapit</i>)
lightning	*kilat	[mb]a-kila	kila	---	---	---	(Ind <i>kilat</i>)

(75) Examples of *k in medial position

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
split	*bekaɟ	---	---	*bəka	bəka	bəka	
open	*bukas	buko	---	*bukas	boka	bβoke	
joint, node	*buku	[mb]uku -	--	*ka-βuku	---	---	
bent	*ma-bukuɟ	---	mbuku	*boko	buku	---	
tie, adhere to	*dekit	---	ɟiki	*dəkɟ	---	---	
dragnet	*puket	---	puka	*pukat	pəku	pəku	(Ind <i>pukat</i>)
come, arrive	*teka	---	---	*təka	dəka	ɟəka	

*g lenited to *ɣ in the WCMP languages (with reflexes which were often null or in variation with the glides [j] or [w] depending on vocalic environment) as well as in PWOC. Examples of *g in initial and medial position are given below:

(76) Examples of *g in initial position

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	PWOC
eat	*kaən	a ⁶¹	ŋa-ha	*ŋa-ɣaŋ	ŋa-ʔa	ŋa-ʔa	*ɣani
elder sibling	*kaka	gaʔa	---	*ɣaya	aʔa	aʔa	---
tree	*kahiw	[g]eu	hadzu	*ɣaju	afu	afu	*ɣaju
send	*(pa)-katu	---	ŋgadu	---	---	pa-ʔaɟu	---
see	*kita	---	eda	*ita	ŋ-əɟi	ŋ-əɟi	---
1pl (incl)	*kita	---	ndai-ta	*ɣita	di	əɟi	*ɣita
cooking pot	*kuden	---	---	*urəŋ	əru	əru	---
skin	*kulit	[g]uli	huri	*ka-lit	k-uri	ka-uri	---
louse	*kutu	[g]utu	hudu	*utu	udu	uɟu	---

⁵⁹ 'chest'

⁶⁰ 'pinch between, insert'

⁶¹ See also əʔa 'you eat', [g]iʔa 'he eats'

(77) Examples of *g in medial position

Gloss	PMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao	PWOc
monkey	*bakay	[mb]aʔe	bahkai	---	---	---	---	---
grime, dirt	*daki	[ndr]aʔi	---	---	*rayi	raʔi	---	---
grass	*zuket	[ndr]uʔu	---	---	*rut	fuʔu	fuʔu	---
fish	*hikan	---	---	---	*iyan	---	iʔa	*iyan
1sg	*i-aku	---	ahku	n-ahu	*jauwa	jaa	jaʔa	*[i]au
1pl (excl)	*i-kami	jaʔami ⁶²	ihkai	n-ami	*jamə	---	---	*ya[m]i
2pl	*i-kamu	---	ihkam	ndai-mu	*yəmi	mu	miu	*ya[m]u
2sg	*i-kahu	---	ihko	---	*yaʔu	au	au	*yo[e]
tail	*ikuR	[g]iʔo	ihku	---	(*ka-iku)	---	---	---
(finger)nail	*kuku	---	---	uhu	*[k/ɣ]uɣu	kuʔu	kuʔu	*ku[k/ɣ]u
man, husband	*laki	---	---	rahi	*layi	laʔi	laʔi	---
steal	*nakaw	---	nahkou	---	---	naʔo	naʔo	*pa[i]naɣo
wound	*nuka	---	---	---	*nuɣa	noʔe	nuʔa	---
fern	*paku	---	pahku	fahu	*payu	---	---	---
flying fox	*paniki	---	---	panihi	*paniyi	niʔi	---	---
climb	*sakay	---	---	---	*[s]aye	haʔe	caʔe	*[s/z]aye
elbow	*siku	[z]iʔu	(siku)	cihu	*siɣu	hiʔu	siʔu	---
fear	*ma-takut	a-taʔu	---	dahu	*ma-ndaut	me-daʔu	ma-ɖaʔu	*matayut
root	*wakaR	waʔa	---	---	---	---	---	*wa[k/ɣ]ar

Additional examples of *k and *g are given below:

(78) Examples of *k

- (a)

<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>
little finger	*kiŋkiŋ	kiŋgi
mangrove	*bakaw	wako
- (b)

<u>Gloss</u>	<u>PMP</u>	<u>PSumba</u>	
mangrove crab	*kali-maŋu	*kalimaŋu	
hook	*kawit	*kait	(Ind <i>kait</i>)
breadfruit	*kuluR	*kulu	
go	*lakaw	*lako	
- (c)

<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>	
men's house	*kamaliR	kemali	---	
grandfather	*baki	---	baki	
fold	*lekuɣ	ləku	ləko	
tie	*hiket	əki	əki	(Ind <i>ikat</i>)

(79) Examples of *g

- (a)

<u>Gloss</u>	<u>PMP</u>	<u>Nias</u>
intestine	*bituka	[mb]etuʔa
- (b)

<u>Gloss</u>	<u>PMP</u>	<u>Dohoi</u>
back	likud	rihku
- (c)

<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>
hook	kawil	hawi

⁶² 'you'

(d)	<u>Gloss</u>	<u>PMP</u>	<u>PSumba</u>	
	octopus	kuRita	*βita	
(e)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>
	thick	*ma-kapal	me-aa	ma-aa

5.1 Flores-Lembata

The distinction between *k and *g is also preserved in Proto-Manggarai (in initial position) and some of the more remote languages of the Flores-Lembata group. Besides the already-mentioned Seranggorang-Nilanapo dialects, other languages include Komodo, Proto-Manggarai, Sika and Kedang. Examples are given below:

(80) Examples of *g in initial position

<u>Gloss</u>	<u>PMP</u>	<u>Komodo</u>	<u>PManggarai</u>	<u>Sika</u>	<u>Ser-Nil</u>	<u>Kedang</u>
eat	*kaən	ahaŋ	*ɣaŋ	a	a	(ka)
1pl (excl)	*kami	hami	*ɣami	ami	ome	(ke)
2sg	*kahu	---	*ɣau	au	---	o
thick	*kapal	---	---	apar	---	(kapal)
tree	*kahiw	---	*ɣazu	ai	---	ai
1pl (incl)	*kita	---	*ɣita	ita	ite	te
(finger)nail	*kuku	---	*ɣuku	---	---	---
skin	*kulit	---	---	ulit	uli	---
louse	*kutu	hutu	*ɣutu	utu	uhto	utu

(81) Examples of *g in medial position

<u>Gloss</u>	<u>PMP</u>	<u>Komodo</u>	<u>Sika</u>	<u>Ser-Nil</u>	<u>Kedang</u>
1sg	*aku	---	?au	---	---
fish	*hikan	---	i?aŋ	i?aŋ	i?a
tail	*ikuR	---	?iur	i?u	---
man, husband	*laki	lahi	la?i-n	te-lae	---
steal	*nakaw	---	nao	---	---
climb	*sakay	---	ha?e	---	---
fear	*takut	---	(b-lau)	---	tau?

5.2 North Sarawak

Certain PNS languages have also preserved evidence for *g intervocalically via lenition: specifically Proto Dayic and Kiput. (Kelabit (K) forms are given in the absence of Proto Dayic reconstructions):

(82) Reflexes of PNS medial *g in Proto Dayic and Kiput

<u>Gloss</u>	<u>PMP</u>	<u>PDayic</u>	<u>Kiput</u>
large intestine	*bituka	*bətuaḥ	tufih
tail	*ikuR	*iur	cəu?
male	*laki	*də-la?ih	laaj
back	*likud	---	cut
steal	*nakaw	mə-no (K)	---
fern	*paku	*pa?uh	paaw
pain, sickness	*sakit	a?it (K)	seet
elbow	*siku	*ijuh	ticəw
fear	*takut	*ta?ut	---

Blust (2009b: 604) states that “Kelabit [...] reflects *k as [ʔ] between unlike vowels provided that the first is not schwa, but as *k* elsewhere,” implying a conditioned loss of *k intervocalically in Kelabit (and, by

implication, Proto Dayic). However, forms such as the following from the first part of Amster's 1995 Kelabit dictionary contradict this:

(83) Examples of Kelabit [k] between unlike vowels, the first of which is not schwa

<u>Gloss</u>	<u>Kelabit</u>
k.o. snail	akəp
agreement	akun
skin disorder	bakε
barren	bakəh
k.o. metal tool	bikuŋ
open	bukaʔ
stunted	bukəŋ

6.0 The uvular series

Although there has been a general longstanding acceptance of PAn *q as a voiceless uvular stop, the uvular series itself has always appeared generally impoverished, the only other member being *R which is sometimes interpreted variously as a uvular trill or voiced fricative. We propose that in addition to *q, the PAn inventory also included a voiced counterpart *G as well as uvular liquid *R (the latter of which nevertheless varied substantially in place of articulation in various Austronesian subgroups and daughter languages).

6.1 Uvular *g

If the evidence presented here for the reconstruction of a novel PAn *g is accurate, then how should original PAn *g be reinterpreted? Traditional PAn *g is a phoneme which has been problematic because of the comparatively small number of forms in which it occurs as well as its general absence from any given core vocabulary list. Blust (2009b: 574-5) provides several examples in which *g is reconstructed, including examples that include the WCMP languages Manggarai and Sika, all of which have a [g] reflex: Manggarai *gəgər* 'shiver with chills, tremble' < PAn *gerger, *gərɪt* 'scratch, claw; scream' < PAn *geriC, *gagar* 'to like, have an appetite for (fighting, talk, sex)' < PMP *gagar 'bold', *pagal* 'hobble; heavy block hung from buffalo's neck to impede his movements' < PMP *pa(ŋ)gal 'neck shackle on domestic animal'; Sika *gəmi* 'pinch, shut, close (as in mouth)' < PMP *gemi 'hold on by biting'.

We propose that it be reinterpreted as an original voiced uvular *G. This solution is attractive because it fills a gap in the reconstructed PAn phoneme inventory in which the voiceless uvular stop *q has always lacked a voiced counterpart; its relative infrequency may also be explained by the fact that uvular is a marked place of articulation.

Under this hypothesis, what is here reconstructed as PAn *g underwent two general developments: (1) devoicing and merger with *k (the case in the majority of Austronesian languages), or (2) lenition, sometimes to the point of deletion (as in the case of several WCMP languages and PWOC). *G then filled this gap by shifting to the velar place of articulation:

*g > *k/ɣ
 *G > *g

In addition to the above, this hypothesis is also based on the following additional pieces of indirect evidence. First, as argued in the previous section, *g patterns with the other voiced stops *b, *d and *ʝ in undergoing lenition in the WCMP languages and PWOC. What we interpret here as *G patterns with the implosive series (Norquest & Downey, in preparation) in resisting lenition, but is unlikely to be an implosive itself due to a general typological constraint on posterior implosives (Greenberg 1970, Maddieson 1984) as well as the fact that *G may occur word-finally. Second, within the comparatively small part of the Austronesian lexicon in which *G occurs, a disproportionate number of forms appear to be ideophonic reduplicants (*Gapgap 'stammer', *Guŋguŋ 'deep resounding sound', *təŋtəŋ 'pound, beat', etc), in which case a marked member of the phoneme inventory may be employed for effect. Finally, the

majority of disjunct/doublet forms listed in the ACD involve crossover with *k (*gisgis ~ *kiskis ‘scrape’, *GƏRGƏR ~ *kƏRKƏR ‘shiver, tremble’, *Gawaj ~ *kawaj ‘tentacles’, etc) which, without witnesses from the key languages listed in this paper, may also potentially be reconstructed with *g (i.e. *gisgis, *gƏRGƏR, *gawaj, etc.). This indicates a consonant that was close enough phonetically to *g to be in variation with it across subgroups; the most plausible gap in the PAn consonant inventory presented here is that of the voiced uvular stop.

6.2 PAn *R

Wolff reconstructs *R as post-velar *ɣ. According to him (2010: 33), “[t]he change of *ɣ to [r] and [l] involves a change widespread in the world’s languages, where a voiced velar spirant becomes a uvular trill and subsequently a tongue-tip trill, which may then merge with /l/.” Blust, on the other hand, states (2009: 582-3) that “it is reasonable to suggest that [...] *R [was] an alveolar trill, a hypothesis that is supported by known sound changes, since the shift of an alveolar trill to a uvular trill is well-attested in the history of French and other European languages, while a shift in the opposite direction is unknown.”

We posit a third interpretation: that *R was originally a uvular trill [ʀ], of which retroflex [ɽ] was sometimes an allophone, and which contrasted with the alveolar trill *r. We consider it unlikely that *R was originally [ɽ], as [ʀ] and [ɽ] would have been in close competition with each other phonetically; the existence of a retroflex series on the other hand may have provided an impetus for variation with [ɽ].

Nias, Bimanese, and PMS are consistent in their reflexes of PAn *r; however, there is a split in the reflexes of PAn *R in which they show unexpected reflexes of *r. The three sets of reflexes are the following:

Table 14: Rhotic reflexes

PMP	RMP	Nias ⁶³	Bima	PSumba	Hawu	Dhao
*R ₁	*R	0	0	0/?	0	0
*R ₂	?	r	r	r	r	r
*r	*r	r	r	r	r	r

Given the other split correspondence sets provided in this paper, it’s tempting to wonder if the split in reflexes of *R indicate yet another reconstructible phoneme. However, unlike the cases of *p/f, *t/ʈ, *l/l̥, *s/c and *k/g, the two reflexes of *R don’t show a strong correlation and instead appear to be somewhat random. The best explanation in this case is that items with the second reflex (*R₂) are borrowings from (or cross-contaminations with) Indonesian. Examples of *r and *R are given below:

(84) Examples of PMP *r

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
sound of grating	*kerit	---	---	*kərit	---	---	(Ind <i>kərit</i>)
insufficient	*kura	---	kura	---	---	---	
to rasp	*parud	---	---	---	---	paru	(Ind <i>parut</i>)
slice, cut fine	*qiris	---	---	*iris	---	---	(Ind <i>hiris</i>)
group	*quruŋ	---	uru	---	---	---	
ramie	*rami	---	---	*rame ⁶⁴	---	rame ⁶⁵	(Ind <i>rami</i>)
red	*rara	---	---	*rara(k)	rara	---	
slip/slide off	*rusrus	ruru	---	---	---	---	(Ind <i>rurut</i>)
stab	*rusuk	---	rusu	---	---	---	
write	*surat	[z]ura	sura	*surat	---	---	(Ind <i>surat</i>)
cockspur	*tara	[ndr]ara	tara	*tara ⁶⁶	tara	ɽara	

⁶³ *R in Nias additionally colored contiguous non-high vowels to [o].

⁶⁴ ‘k.o. high thicket’

⁶⁵ ‘hemp’

⁶⁶ ‘thorn, barb’

k.o. large bamboo	*teriŋ	---	---	*təriŋ	---	---	
directly through	*terus	---	---	*turus	---	---	(Ind <i>tərus</i>)

The majority of forms have null reflexes for *R⁶⁷. However, a small residue of forms have reflexes of *r for PAN *R:

(85) Examples of PMP *R

<u>Gloss</u>	<u>PMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PSumba</u>	<u>Hawu</u>	<u>Dhao</u>	
new	*baqeRu	---	bou	*mbəʔu	wiu	hiu	
lung	*baRa	[mb]o	---	---	wa	haa	
swollen	*baReq	a-bao	---	*(m)baʔə	ʒai	bai	
brave	*baRani	(barani)	mbani	*ʒani	ʒani	bani	(Ind <i>bərani</i>)
rotten	*baRiw	---	mbai	*mbai	---	---	
hibiscus	*baRu	---	wau	*βau	wau	---	
rice	*beRas	([mb]əra)	(ʒura)	*βias	---	---	(Ind <i>bəras</i>)
give	*beRay	---	mbei	---	wie	hia	
paddle	*beRsay	---	wese	---	wohe	---	
wild taro	*biRaq	[mb]io	wia	*βia	---	---	
bear	*biRuaŋ	([mb]erua)	(birua)	---	---	---	(Ind <i>bəruaŋ</i>)
blood	*daRaq	[ndr]o	raa	*raa	raa	raa	
earth	*daReq	---	rae	---	rai	rai	
bathe	*diRuq	---	ndeū	---	fiu	diu	
bone	*duRi	[ndr]oi ⁶⁸	ruwi	*ruī	ruī	ruī	
west	*habaRat	(bara laowi) ⁶⁹	waa ⁷⁰	(*warat)	waa	haa	(Ind <i>barat</i>)
housepost	*hadiRi	---	rii	---	ge-rii	aga-rii	
needle	*zaRum	---	ndau	*roŋ	fau	fau	
bite	*kaRat-i	---	---	*kati	kadi	kaɟi	
scratch	*kaRaw	---	kao	*kaʔu	kao	---	
run	*laRiw	o-loi	rai	*lai	rai	rai	
heavy	*ma-beRat	a-bua	---	*mbuəto	---	bia	
red	*ma-iRaq	ojo	---	*mia	mea	mea	
come	*maRi	---	mai	*mai	mai	mai	
dry	*ma-Raŋaw	---	maŋo	---	maŋu	maŋo	
man	*ma-Ruqanay	---	mone	*mone	mone	mone	
k.o. tree	*naRa	---	(nara)	(*nara)	---	---	
hoarse	*paRaw	a-fau	ka-fao	---	---	---	
stingray	*paRih	---	fai	*pai	---	---	
slave	*qaRta	---	ada	*ata	---	---	
current	*qaRus	---	(aru)	(*aris) ?	---	---	(Ind <i>arus</i>)
salt	*qasiRa	[n]asio	sia	---	---	---	
bind, raft	*Rakit	---	(raŋki)	(*rakit)	---	---	(Ind <i>rakit</i>)
hundred	*Ratus	otu	(ratu)	(*rata) ⁷¹	---	---	(Ind <i>ratus</i>)
root	*Ramut	---	amu	*amu	amo	amo	
thousand	*Ribu	---	(riwu)	(*riβu) ⁷²	(riwo)	(riho)	(Ind <i>ribu</i>)
house	*Rumaq	(ruma)	uma	*uma	əmu	əmu	(Ind <i>rumah</i>)

⁶⁷ Note that as pointed out in Blust (2008), the presence of a rhotic can often be inferred in PHD on the basis of *i in PHD.

⁶⁸ ‘thorn’

⁶⁹ ‘northwest’

⁷⁰ ‘below’

⁷¹ ‘thousand’

⁷² ‘10,000’

comb	*saRu	---	cau	*sai ?	---	---	
indigo	*taRum	---	dau	---	dao	---	
porcupine	*taRutuŋ	---	dudu	---	dudu	ɖuɖu	
tin, lead	*timeRaŋ	---	---	(*tambura)	(teməra)	(taməru)	
vein	*uRat	[n]uo	ka-ʔuʔa				
day	*waRi	(hari)	ai	---	---	---	(Ind <i>hari</i>)

Additional items in individual languages are given below:

(86) Regular reflexes of *R

(a)	<u>Gloss</u>	<u>PMP</u>	<u>Nias</u>	
	ember	*baRah	[mb]o	
(b)	<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>	
	typhoon	*baRiuh	ɓadʒu	
	squeeze	*peRas	pua	
	evening	*Rabihi	awi	
(c)	<u>Gloss</u>	<u>PMP</u>	<u>PSumba</u>	
	octopus	*kuRita	*βita	
(d)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>
	rear	*buRit	wui	hui ⁷³
	fog	*Rabun	awu	ahu

(87) Irregular reflexes of *R

(a)	<u>Gloss</u>	<u>PMP</u>	<u>Bima</u>	
	rotten	*buRuk	mburu	(Ind <i>buruk</i>)
	young girl	*daRa	ndara ⁷⁴	(Ind <i>dara</i>)
	wild duck	*ŋaRaŋ	ŋara	
	oyster	*tiRem	tire	(Ind <i>tiram</i>)
(b)	<u>Gloss</u>	<u>PMP</u>	<u>PSumba</u>	
	hibiscus	*baRu	*βaru	(Ind <i>baru</i>)
	pluck, pull out	*Rabut	*raβut	(Ind <i>rabut</i>)
	lines in palm	*uRat	*urat	(Ind <i>urat</i>)
(c)	<u>Gloss</u>	<u>PMP</u>	<u>Hawu</u>	<u>Dhao</u>
	mangrove crab	*kaRakap	---	karaka
	ribs	*Rusuk	ruhu	---
				(Ind <i>rusuk</i>)

7.0 Conclusion

Evidence has been presented in this paper for three new phonemes (*f, *ʎ, and *g), as well as additional extra-Formosan evidence for *t and an expanded domain for *c within PMP. The evidence comes from two subgroups on Borneo, as well as three of the four corners of the Austronesian-speaking world: Nias in the southwest, PWOC in the northeast, and WCMP in the southeast, with the conservative Formosan languages of Taiwan in the extreme northwest completing the picture.

According to the methodology of historical linguistics, whatever is reconstructed for PMP that is not the result of a conditioned split can be projected to the level of Proto Austronesian. The Out-of-Taiwan ‘express train’ hypothesis predicts that phonemic mergers should have occurred as the Austronesian

⁷³ ‘stern’

⁷⁴ ‘young’

expansion proceeded in time and space; the number of inherited phonemes for any node would be equal to or less than the number of those in the node above, and any secondary splits increasing the phoneme inventory which occurred in a lower node would be localized within that node with the conditioning factors likely remaining transparent.

As shown below in Table 15, however, this is not the pattern that appears. The Formosan languages are still unique in directly preserving evidence for the palatal phonemes *ç and *ʎ. Formosan evidence for *ʈ, however, is now found in three other locations, and evidence for several other phonemes can be found in several other groups as well:

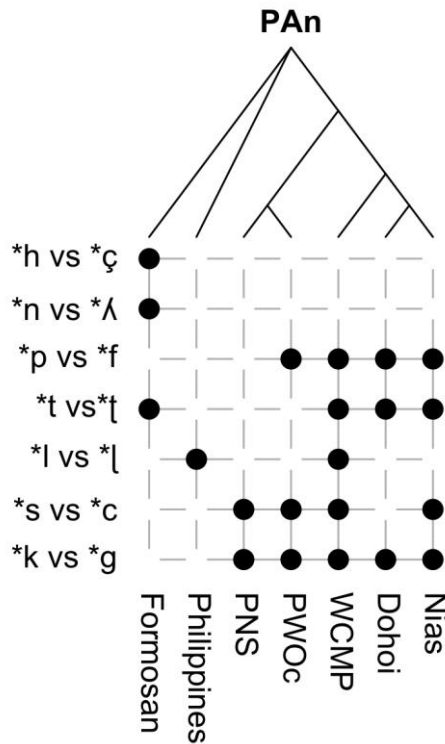


Figure 2: Archaic PAn phonemic distinctions

In terms of sheer number of distinctions preserved, WCMP is actually the most conservative group, followed by Nias. If one assumes CMP and EMP (SWHNG + Oceanic; See Fig. 1) to be the two lowest nodes of the Austronesian phylogenetic tree, then it is perplexing that they are more conservative than most WMP languages, the exceptions being the West Barito and North Sarawak groups on Borneo and the Barrier Islands group (to the extent that other languages of that region can be shown to subgroup with Nias⁷⁵). How to interpret these data?

As mentioned above, the most conservative groups lie either on the periphery of the Austronesian-speaking world or on Borneo. Our present working hypothesis is that these languages represent an older layer of Austronesian languages that have been located in their present positions for some time. The WMP languages (excluding the Barrier Island languages), on the other hand were more recent expansions by various groups out of Borneo, possibly triggered by climate stress or other cataclysmic factors. The hypothesis that the Malayo-Chamic languages originated on Borneo (see for example Collins & Sariyan 2006) is well known; the South Sulawesi languages are related to the Tamanic group on Borneo, and the Philippine languages may subgroup with Sabahan (although this is still conjectural (Blust 1998)). If these examples are any indication, then it may be shown eventually that other WMP languages and subgroups originated on Borneo (cf. Blust 2010), and that immigration out of Borneo and into the surrounding islands

⁷⁵ We were quite fortunate to come across Lase (2011) in which the interphrasal distinctions in Nias word onsets were recorded for (what we believe to be) the first time. It is unclear whether similar distinctions may exist in other Barrier Islands languages, and we consider this to be one important aspect of future research.

has been occurring for quite some time, including quite possibly even the Philippine languages from Sabah in northern Borneo.

The phylogenetic tree in Figure 2 supports this conclusion. The tree was derived via a binary distance matrix based on phonological mergers and neighbor-joining. It suggests that the closest relationships between the easternmost Austonesian groups (WCMP and PWOc) are not to each other, but rather to discrete groups on Borneo, WCMP joining with Dohoi and PWOc joining with Proto North Sarawak. Although this phylogeny must remain tentative for now, we note that it is geographically consistent with two eastern migrations out of Borneo – one from southeast Borneo into the Nusa Tenggara region, and one from northern Borneo to the Bird's Head region of New Guinea which then spread eastward.

With this in mind, we propose an alternative to the traditional Austronesian expansion hypothesis: that the Formosan languages do form a subgroup, and that this group represents the first migration away from the original Austronesian homeland (and therefore the first split in the Austronesian phylogenetic tree, similar to the place of Anatolian within Indo-European). This Formosan-Malayo-Polynesian sister-group hypothesis would predict that retentions and innovations would be found in both subgroups, and not necessarily be constrained to WMP, as the OoT hypothesis would imply.

If it can be shown convincingly that the Formosan languages form a discrete innovation-defined subgroup and that the Formosan group is effectively a sister of PMP, then the question of Formosan origins becomes open – did the early Formosans migrate from mainland southeast Asia, as is commonly supposed, or might they have migrated from somewhere further south, perhaps ultimately from Borneo itself? Figure 2 above confirms that the Formosan languages still maintain a unique position in the Austronesian family tree, as they are the only languages to preserve concrete evidence for the phonemes we interpret as *ç and *ʎ. The same cannot be said for Proto Philippines, which retains evidence for *ʎ with WCMP (and evidence for *ŋ -- if this hypothesis is valid -- with North Sarawak and Malayic). This question is ultimately outside the scope of this paper, but we hope to explore it in the future using both linguistic and non-linguistic evidence. The implications in regard to Southeast Asian prehistory and demographic processes are profound, and future work (especially interdisciplinary work) should bear this in mind as novel hypotheses are explored.

References

- Adelaar, K. Alexander & Nikolaus P. Himmelmann. *The Austronesian Languages of Asia and Madagascar*. London/New York: Routledge.
- Amster, Matthew H. 1995. *Kelabit/English, English/Kelabit Glossary: A Concise Guide to the Kelabit Language*. Kuching, Malaysia: Rurum Kelabit Sarawak.
- Blust, Robert. 1977. The Proto-Austronesian Pronouns and Austronesian Subgrouping: A Preliminary Report. *Working Papers in Linguistics* 9 (2): 1-15.
- _____. 1982. The Linguistic Value of the Wallace Line. *Bijdragen tot de taal-, land- en volkenkunde* 138: 231-50.
- _____. 1984/85. The Austronesian Homeland: A Linguistic Perspective. *Asian Perspectives* 26: 45-67.
- _____. 1988. *Austronesian Root Theory: An Essay on the Limits of Morphology*. Amsterdam: John Benjamins.
- _____. 1991. The Greater Central Philippines Hypothesis. *Oceanic Linguistics* 30 (2): 73-129.
- _____. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32 (2): 241-93.
- _____. 1995a. The Prehistory of the Austronesian-Speaking Peoples: A View from Language. *Journal of World Prehistory* 9 (4): 453-510.
- _____. 1995b. Notes on Berawan Consonant Gemination. *Oceanic Linguistics* 34 (1): 123-38.
- _____. 1995c. The Position of the Formosan Languages: Method and Theory in Austronesian Comparative Linguistics. In *Austronesian Studies Relation to Taiwan*, Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.). Taipei: Academia Sinica. 585-650.
- _____. 1995d. Sibilant Assimilation in Formosan Languages and the Proto-Austronesian Word for 'Nine': A Discourse on Method. *Oceanic Linguistics* 34 (2): 443-53.
- _____. 1995e. *Austronesian Comparative Dictionary*.
- _____. 1997a. Ablaut in Northwest Borneo. *Diachronica* 14 (1): 1-30.

- _____. 1997b. Nasals and Nasalization in Borneo. *Oceanic Linguistics* 36 (1): 149-79.
- _____. 1998. The Position of the Languages of Sabah. In *Pagtanaĭw: Essays on Language in Honor of Teodoro A. Llamzon*, Ma. Lourdes S. Bautista (ed.). 29-52. Manila: Linguistic Society of the Philippines.
- _____. 1999. Subgrouping, Circularity and Extinction: Some Issues in Austronesian Comparative Linguistics. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, Elizabeth Zeitoun & Paul Jen-kuei Li (eds.). Taipei: Academia Sinica.
- _____. 2000. Low-Vowel Fronting in Northern Sarawak. *Oceanic Linguistics* 39 (2): 285-319.
- _____. 2001a. Language, Dialect, and Riotous Sound Change: The Case of Sa'ban. In *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society 1999*, Graham Thurgood (ed.). 249-359. Arizona State University Program for Southeast Asian Studies, Monograph Series Press, Tempe: Arizona State University.
- _____. 2001b. Malayo-Polynesian: New Stones in the Wall. *Oceanic Linguistics* 40 (1): 151-55.
- _____. 2002a. The History of Faunal Terms in Austronesian Languages. *Oceanic Linguistics* 41 (1): 89-139.
- _____. 2002b. Kiput Historical Phonology. *Oceanic Linguistics* 41 (2): 384-438.
- _____. 2005. The Linguistic Macrohistory of the Philippines: Some Speculations. In *Current Issues in Philippine Linguistics and Anthropology Parangal kay Lawrence A. Reid*, Hsiu-chuan Liao & Carl R. Galvez Rubino (eds.). 31-68. Manila: The Linguistic Society of the Philippines and The Summer Institute of Linguistics, Philippines.
- _____. 2006a. Anomalous Liquid : Sibilant Correspondences in Western Austronesian. *Oceanic Linguistics* 45 (1): 210-16.
- _____. 2006b. The Origin of the Kelabit Voiced Aspirates: A Historical Hypothesis Revisited. *Oceanic Linguistics* 45 (2): 311-38.
- _____. 2007a. OĀma LoĪngh Historical Phonology. *Oceanic Linguistics* 46 (1): 1-53.
- _____. 2007b. The Linguistic Position of Sama-Bajaw. *Studies in Philippine Languages and Cultures* 15: 73-114.
- _____. 2008. Is there a Bima-Sumba Subgroup? *Oceanic Linguistics* 47 (1): 45-113.
- _____. 2009a. The Position of the Languages of Eastern Indonesia: A Reply to Donohue and Grimes. *Oceanic Linguistics* 48 (1): 36-77.
- _____. 2009b. *The Austronesian Languages*. Canberra : Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- _____. 2010. The Greater North Borneo Hypothesis. *Oceanic Linguistics* 49 (1): 44-118.
- _____. 2012. Hawu Vowel Metathesis. *Oceanic Linguistics* 51 (1): 207-233.
- Collins, James T. & Awang Sariyan (eds.). 2006. *Borneo and the Homeland of the Malays*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Dahl, Otto. 1976. *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series, No. 15. London: Curzon Press.
- Donohue, Mark & Charles Grimes. Yet More on the Position of the Languages of Eastern Indonesia and East Timor. *Oceanic Linguistics* 47 (1): 114-58.
- Dyen, Isidore. 1963. The Position of the Malayopolynesian Languages of Formosa. *Asian Perspectives* 7 (1-2): 261-71.
- _____. 1995. Borrowing and Inheritance in Austronesianistics. In *Austronesian Studies Relation to Taiwan*, Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.). Taipei: Academia Sinica. 455-519.
- Greenberg, Joseph H. 1970. Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics* 36 (2): 123-145.
- Greenhill, Simon, Robert Blust & Russell Gray. 2003-2012. Austronesian Basic Vocabulary Database.
- Grimes, Charles. 2008. *Lii Hawu (Sabu) Online Dictionary*. Kupang: UBB-GMIT.
- _____. 2010. Hawu and Dhao in Eastern Indonesia: Revisiting Their Relationship. In *East Nusantara: Typological and Areal Analyses*, Michael Ewing & Marian Klamer (eds.). Canberra: Pacific Linguistics.
- Grimes, Charles, Tom Therik, Barbara Dix Grimes & Max Jacob. 1997. *A Guide to the People and Languages of Nusa Tenggara*. Artha Wacana Press. Kupang.

- Haudricourt, André-Georges. 1950. Les consonnes préglottalisées en Indochine. *Bulletin de la Société de Linguistique* (46): 172-182.
- Heine, Bernd. 1978. The Sam languages: a history of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6.
- Hudson, Alfred B. 1967. *The Barito Isolects of Borneo: A Classification Based on Comparative Reconstruction and Lexicostatistics*. Data paper: number 68, Southeast Asian Program, Department of Asian Studies, Cornell University, Ithica, New York.
- Inagaki, Kazuya. 2005. Phonemic Sketch of Dohoi/Kadorih. *Kyoto University Linguistic Research* (24): 15-43.
- _____. 2006. A sketch on the Morphosyntax of Kadorih (Dohoi: Austronesian). *Kyoto University Linguistic Research* (25): 41-65
- _____. 2011. Word duplication in Kadorih. *Asian and African Languages and Linguistics* (6): 27-44.
- Klamer, Marian. 1998. *A Grammar of Kambara*. Berlin/New York: Mouton de Gruyter.
- _____. 2002. Typical Features of Austronesian Languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41 (2): 363-83.
- _____. 2004. East Nusantara: Genetic, Areal, and Typological Approaches. *Oceanic Linguistics* 43 (1): 240-44.
- _____. 2009. The Use of Language Data in Comparative Research: A Note on Blust (2008) and Onvlee (1984). *Oceanic Linguistics* 48 (1): 250-63.
- Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.
- Lase, Apolonius. 2011. *Kamus Li Niha: Nias-Indonesia*. PT Kompas Media Nusantara. Jakarta.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
- Mansyur, Ismail, Azis Muhidin, Yakub Saleh, H. Taufik & Usman Kasim. 1985. *Kamus Bima-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Norquest, Peter & Sean Downey. Forthcoming. Austronesian implosives.
- Onvlee, Louis. 1984. *Kambaraas (Oost-Soembaas)-Nederlands Woordenboek*. Dordrecht: Foris.
- Pawley, Andrew & Malcolm Ross. 1993. *Annual Review of Anthropology* 22:425-59.
- Paz, Consuelo J. 1981. *A Reconstruction of Proto-Philippine Phonemes and Morphemes*. Dilliman, Quezon City: Cecilio Lopez Archives of the Philippine Languages and the Philippine Linguistic Circle, University of the Philippines.
- Ross, Malcolm. 1988. Proto Oceanic and the Austronesian Languages of Western Melanesia. Canberra: Pacific Linguistics.
- _____. 1992. The Sound of Proto-Austronesian: An Outsider's View of the Formosan Evidence. *Oceanic Linguistics* 31 (1): 23-64.
- _____. 2009. Proto Austronesian verbal morphology: a reappraisal. In *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*. Alexander Adelaar & Andrew Pawley (eds.). Pacific Linguistics: Canberra. 295-326.
- Smoll, Laetitia. 2011. Towards a Typology of Ideophones.
http://www.academia.edu/2649993/Towards_a_Typology_of_Ideophones.
- Wolff, John U. 2010. *Proto-Austronesian Phonology with Glossary* (2 volumes). New York: Southeast Asia Program Publications, Cornell University.
- Yunus, H. Anwar. 1981. *Kamus Bahasa Bima-Indonesia, Indonesia-Bima (Nggahi Mbojo)*. Jakarta: Yayasan "Muhammad Salahuddin".

Appendix

The following list of PMP forms have been amended to include the evidence presented in this paper for the proposed phonemes *f, *t, *l, *c, and *g. Additional distinctions not discussed in this appendix have been made between plain and implosive *b/*ɓ, *d/*ɗ and *j/*ʝ which will be the topic of another paper (Norquest & Downey, in preparation). Traditional PMP *e and *y have been rewritten as *ə and *j respectively, in conformance with the IPA. Forms with unexpected correspondences, or in which otherwise crucial distinctions are neutralized due to phonotactic environment, are placed in parentheses; cognates that are suspected to be Indonesian loans are placed in curly brackets. We have only cited forms that include at least two cognates from two branches. We have also reconstructed Proto Hawu-Dhao (PHD) on the basis of forms in both languages as well as single forms in each, as the two languages are phonologically quite close to each other.

PMP initial *p

Gloss	RMP	Nias	Bima	PHD	PWOc
causative	*pa-	fa-	pa-	*pa-	---
warm/hot	*panas	---	pana	*pana	---
fruit bat	*panigi	---	panihi	*niʔi	---
broken	*paʔaq	(a-fatə)	mpada	*paɗa	---
squeeze	*pəRəs	---	pua	---	*poRos
how many	*[p/f]iɗa	---	pila	*pəri	(*βiza)
choose	*piliq	fili	---	*pili	---
seven	*piʔu	fitu	pidu	*piɗu	---
dragnet	*pukət	---	puka	*pəku	---
white	*puʔiq	(a-fusi)	---	*puɗi	---

PMP medial *p

Gloss	RMP	Dohoi	S-N	Bima	PHD	PWOc
fathom	*dəpa	---	---	ndupa	*rəpa	---
fold	*ləpət	---	---	lipi	*ləpa	---
pinch	*qapit	---	---	api	*ŋ-api	*kapi
four	*ə[p/f]ac	(ohpat)	pa	upa	*əpa	(*βati)
sharp pain	*ha[p/f]əɗəs	([β]əxə)	---	pili	*pəɗa	---
mat	*təpir	---	---	dipi	*ɗəpi	---

PMP initial *f

Gloss	RMP	Nias	Bima	PHD	PWOc
paddy	*faɗaj	[β]axe	fare	*are	---
to shoot	*fanaq	---	fana	---	*βana
leucoderma	*fanaw	---	fano	*ano	---
pandanus	*faŋdan	[β]andra	fanda	---	---
thigh	*faqa	[β]aha	---	---	*βaqa
chisel	*[p/f]aqət	[β]ahə	paa	*paa	---
hoarse	*faraw	(a-fau)	fao	---	---
stingray	*farih	---	fai	---	*βari
turtle	*fənu	[β]ənu	fonu	*ənu	*βənu
ten	*ca-ŋa-fu[uq]	[β]ulu	mpuru	*ŋ-uru	---
navel	*fusəɗ	[β]usə	---	*əsu	---

PMP medial *f

<u>Gloss</u>	<u>RMP</u>	<u>Dohoi</u>	<u>S-N</u>	<u>Bima</u>	<u>PHD</u>	<u>PWOc</u>
thick	*ma-kafal	kahpan	---	---	*ma-aa	---
tooth	*nifən	nihpo	ihpe	---	---	---
dream	*h-in-ifi	nuhpi	---	nifi	*nii	*m-niβi
thin	*ma-nifis	mihpi	mihpi	{nipi}	*ma-nii	---
centipede	*qa u-hifan	(joripan)	---	rifa	---	*qaliβan
gall	*qafədu	---	---	folu	*ədu	---
lime (for betel)	*qafur	---	---	afu	*ao	---
fire	*hafuy	ahpui	ahpe	afi	*ai	*(j)aβi
reward	*ufaq	uhpa	---	ufa	---	---

PMP initial *t

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PHD</u>
natural cockspur	*[t]ara	([ndr]ara)	tara	*[t/t]ara
head covering	*təduŋ	---	todu	*tədu
three	*təlu	təlu	tolu	*təlu
pierce	*tuβak	---	tuβa	*təβu
knock, peck	*tuktuk	---	tutu	*tutu
pole for boat	*tuku	---	tuku	*tuku
roast, burn	*tunu	tunu	---	*tunu
old (people)	*tuqah	tua	---	*tua
knee	*tuqud	---	ta-tuʔu	*tuu

PMP medial *t

<u>Gloss</u>	<u>RMP</u>	<u>Dohoi</u>	<u>S-N</u>	<u>Bima</u>	<u>PHD</u>
calf	*b[ə/i]tiəs	botih	---	wisi	---
cut, shop	*hə[t/t]ək	(n-ohtok)	---	---	*əta

PMP initial *t

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PHD</u>
fear	*tagut	(a-taʔu)	dahu	*ma-ɖaʔu
rope	*taljh	{tali}	dari	*ɖari
bury	*tanəm	[d]anə	---	*pa-ɖana
earth	*tanəq	[d]anə	dana	---
feces	*taqi	[d]ai	(taʔi)	*ɖei
indigo	*tarum	---	dau	*ɖao
sea	*tasik	---	dasi	*ɖasi
person	*tau	---	dou	*ɖau
sugarcane	*təβuh	---	doβu	*ɖəβu
mat	*təpɪr	---	dipi	*ɖəpi
elder person	*tua	[d]ua	dua	---
fish poison	*tuba	[d]uβa	duwa	---
freshwater eel	*tuna	---	duna	*ɖəno

PMP medial *t

<u>Gloss</u>	<u>RMP</u>	<u>Dohoi</u>	<u>S-N</u>	<u>Bima</u>	<u>PHD</u>
stone	*baʔu	bahtu	wahto	wadu	*βaɖu
blind	*βuʔa	---	---	mbuda	*βəɖu
send	*(pa-)gaʔu	---	---	ŋgaɖu	*pa-ʔaɖu
louse	*guʔu	kuhtu	uhtu	hudu	*uɖu
see	*(g)iʔa	---	---	eda	*-əɖi
we (excl)	*giʔa	ihto	---	(ndai-ta)	*əɖi

raw	*ma-qəʔaq	---	---	mada	*maɖa
eye	*maʔa	mahta	---	mada	*maɖa
die	*m-aʔaj	mahtoi	mahta	made	*maɖe
seven	*piʔu	pihtu	---	pidu	*piɖu
white	*puʔiq	---	---	pudi	*puɖi
liver	*qaʔaj	ahtoi	---	ade	*aɖe
cucumber	*qaʔimun	---	---	dimu	*ɖimu
egg	*qaʔəlur	(kotoruh)	---	dolu	*ɖəlu
porcupine	*ʔaru[t/ʔ]uŋ	(tohotuŋ)	---	dudu	*ɖuɖu
vomit	*(um)-uʔaq	ŋ-uhta	---	---	*məɖu

PMP *ɪ

<u>Gloss</u>	<u>RMP</u>	<u>PPh</u>	<u>Bima</u>	<u>PHD</u>
return	*ɖaliw	---	ɖali	*ka-ɖali
buy	*bəli	*bəlɪʔ	weli	*βəli
seaward	*lahud	---	ka-lau	*ləu
sail	*ləjaR	---	lodʒa	*ləi
fold	*ləpət	---	lipi	*ləpa
five	*lima	*limáʔ	lima	*ləmi
day	*qaləɖaw	---	liro	*ləɖo
egg	*qaʔəlur	*ʔiklúg	dolu	*ɖəlu
three	*təlu	(*ta-tʔúʔ)	tolu	*təlu
ear(wax), deaf	*ʔuli, *ʔilu	*tulɪʔ	---	*ɖilu

PMP *ɹ

<u>Gloss</u>	<u>RMP</u>	<u>PPh</u>	<u>Bima</u>	<u>PHD</u>
reverse	*balɪk	---	wari	*βari
moon	*buʔan	*búʔan	wura	*βəru
ear of grain	*bu[l/ʔ]iR	(*búlig)	wuri	*βuri
body hair	*buʔu	*buʔbúʔ	wuru	*βuru
in(side)	*ɖaʔəm	*ɖáʔəm	---	*ɖara
road	*jaʔan	*daʔan	---	*fara
skin	*guʔit	---	huri	*ka-ʔuri
rat	*ʔabaw	---	ka-rawo	*ma-raβo
man, husband	*ʔagi	*la-ʔáki	rahi	(*laʔi)
ginger	*[l/ʔ]aɖia	(*laʔújaʔ)	rea	---
run away, flee	*ʔariw	---	rai	*rai
ten	*ca-ŋa-fúʔuɖ	*púʔuʔ	mpuru	*ŋ-uru
pestle	*qahəʔu	*háklʔuŋ	aru	*aru
head(waters)	*quʔu	---	uru	*uru
torch	*suʔuɖ	*suʔúʔ	---	*suru
forest	*haʔas	*háʔas	ara	*ara
snake	*huʔaR	*ʔuʔóg	---	*əru
rope	*ʔaʔih	---	dari	*ɖari
eight	*waʔu	*waʔúʔ	waru	*aru

PMP initial*s

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PDayic</u>	<u>PSumba</u>	<u>PHD</u>	<u>PWOc</u>
error	*sa aɣ	(sala)	sara	salaʔ (K)	*sala	(*sala)	---
bamboo floor	*sa əɾ	[z]alo	sari	---	*ka-[s]al[e]	---	---
python	*sawa	[z]awa	sawa	---	---	---	---
douse fire	*sə[ɓ/b]u	---	suwu	---	---	*səɓu	---
ant	*səɖəm	[z]ixə	sa-sili	---	---	---	---
elbow	*[c/s]igu	[z]iʔu	(cihu)	(*ijuh)	*siyu	*siʔu	---
torch	*suluɣ	[z]ulu	---	---	*[s]ulu	*suru	---
breast, milk	*susu	---	susu	tusu (K)	*susu	---	*susu

PMP medial*s

<u>Gloss</u>	<u>RMP</u>	<u>Bima</u>	<u>PDayic</u>	<u>PSumba</u>	<u>PHD</u>	<u>PWOc</u>
iron	*bəsi	bəsi	---	*bəçi	*bəsi	---
paddle	*bəɾsaj	wese	bəsay (K)	*bōse	---	*boze
contents	*isi	isi	---	*içi	---	---
heart	*pusuɣ	---	*pusuʔ	(*puçu)	*usu	---
pierce, stab	*rusuk	rusu	n-usuk (K)	---	---	---
breast, milk	*susu	susu	tusu (K)	*susu	---	*zuzu
ripe	*[a-əsak	ndasa	---	---	*ma-ɖasa	---
sea	*[a-sik	dasi	---	*tasik	*ɖasi	---

PMP initial *c

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PDayic</u>	<u>PSumba</u>	<u>PHD</u>	<u>PWOc</u>
climb	*cagaɣ	---	---	---	*[s]ayi	*caʔe	*saye
bifurcation	*caŋa	---	caŋa	---	*[s]aŋa	---	*saŋa
comb	*caɾu	---	cau	---	*[s]ai	---	---
nine	*ciwa	siwa	ciwi	*iwaʔ	*çiwa	*ceo	---
comb	*cuɖud	suxu	---	udud (K)	---	---	---
spoon, ladle	*cuduk	---	ciru	---	*[s]uru	*curu	---

PMP medial *c

<u>Gloss</u>	<u>RMP</u>	<u>Bima</u>	<u>PDayic</u>	<u>PSumba</u>	<u>PHD</u>	<u>PWOc</u>
wet	*bacaɣ	beca	*bahaʔ	*mba[s]a	{*basa}	---
one	*ica	ica	(*əɖəh)	*iça	*əci	---
who	*i-cai	---	*iih	---	*cee	*[s/z]ai
sated	*ma-bəcɾ	mcocu	*baur	*ka-mbəçu	*bəcu	---
salty	*ma-qacin	maci	m-ain (K)	*maçi	{*masi}	---
mercy, pity	*ma-qaciq	meci	*m-aheʔ	---	---	---
mortar	*əcuŋ	nocu	iuŋ (K)	*ŋoçu	---	---
navel	*fu[c/s]əɖ	---	*puəɖ	*pə[s]u	(*əsu)	---
smoke	*(ma)-qacu	---	---	*maçu	---	*qa[s/z]u

PMP initial *k

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PSumba</u>	<u>PHD</u>
bite	*kaɾaɖ-i	---	---	*kati	*kaɖi
scratch	*kaɾaw	---	kao	*kaʔu	*kao
stand	*kəɖəŋ	---	kidɪ	---	*kəɖi
belly	*kəmpuŋ	---	---	*kambu	*kambu
pinch	*kəpit	---	kapi	*kəpit	---
lightning	*kilat	[mb]a-kila	kila	---	---

PMP medial *k

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>PDayic</u>	<u>Kiput</u>	<u>Dohoi</u>	<u>Bima</u>	<u>PSumba</u>	<u>PHD</u>
split	*b̥akaq	---	---	---	---	---	*b̥aka	*b̥aka
open	*buka	buko	bukaʔ(K)	buka:ʔ ---	---	---	*bukas	*mboka
node	*buku	[mb]uku	*bukuh	bukaw	---	---	*ka-buku	---
bent	*ma-bukuq	---	---	---	---	mbuku	*boko	*mbuku
adhere	*d̥akit	---	*d̥əkət	d̥əkət	---	d̥iki	*d̥akit	---
dragnet	*pukət	---	pukət(K)	---	---	puka	*pukat	*p̥aku
arrive	*t̥aka	---	---	---	---	---	*t̥aka	*d̥aka

PMP initial *g

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>Bima</u>	<u>PSumba</u>	<u>PHD</u>	<u>PWOc</u>
eat	*gaŋ	a	ŋa-ha	*ŋa-yaŋ	*ŋa-ʔa	*yaŋi
elder sibling	*gaga	gaʔa	---	*yaŋa	*aʔa	---
tree	*gaju	[g]eu	hadʒu	*yaŋu	*afu	*yaŋu
thick	*gafal	---	---	---	*ma-aa	---
send	*(pa)-gaʔu	---	ŋgadu	---	*pa-ʔaʔu	---
see	*(g)iʔa	---	eda	*ita	*-əʔi	---
1pl (incl)	*giʔa	---	ndai-ta	*yita	*əʔi	*yita
cooking pot	*gudən	---	---	*urəŋ	*əru	---
skin	*gulit	[g]uli	huri	*ka-lit	*ka-ʔuri	---
louse	*guʔu	[g]utu	hudu	*utu	*uʔu	---

PMP

medial *g

<u>Gloss</u>	<u>RMP</u>	<u>Nias</u>	<u>PDayic</u>	<u>Kiput</u>	<u>Dohoi</u>	<u>Bima</u>	<u>PSumba</u>	<u>PHD</u>	<u>PWOc</u>
monkey	*bagaj	[mb]aʔe	---	---	bahkai	---	---	---	---
intestine	*bituga	[mb]etuʔa	*b̥ətuəh	tufih	---	---	---	---	---
grime, dirt	*dagi	[ndr]aʔi	---	---	---	---	*raʔi	*raʔi	---
grass	*jugut	[ndr]uʔu	---	---	---	---	*rut	*fuʔu	---
fish	*higan	---	---	---	---	---	*iyaŋ	*iʔa	*iyaŋ
1sg	*i-agu	---	---	(kaw)	ahku	n-ahu	*jauwa	*jaʔa	*[i]au
1pl (excl)	*i-gami	jaʔami	---	---	ihkai	n-ami	*jamə	---	*ya[m]i
2pl	*i-gamu	---	*gamih	(kamaj)	ihkam	ndai-mu	*yəmi	*miu	*ya[m]u
2sg	*i-gahu	---	---	---	ihko	---	*yaʔu	*au	*yo[e]
tail	*igur	[g]iʔo	*iur	cəuʔ	ihku	---	(*ka-iku)	---	---
(finger)nail	*[k/g]ugu	---	---	---	---	uhu	*[k/ɣ]uyu	*kuʔu	*ku[k/ɣ]u
man	*[agi	---	*d̥ə- laʔih	la:ʒ	---	rahi	*laʔi	*laʔi	---
back	*ligud	---	---	cut	rihkut	---	---	---	---
steal	*nagaw	---	mə-no (K)	---	nahkou	---	---	*naʔo	*pa[i]nayo
wound	*nuga	---	---	---	---	---	*nuya	*nuʔa	---
fern	*fagu	---	*paʔuh	pa:w	pahku	fahu	*payu	---	---
flying fox	*panigi	---	---	---	---	pahihi	*paniyi	*niʔi	---
climb	*cagaj	---	---	---	---	---	*[s]aye	*caʔe	*[s/z]aye
elbow	*sigu	[z]iʔu	*ijuh	ticəw	(siku)	cihu	*siyu	*siʔu	---
fear	*ma-ʒagut	a-taʔu	*taʔut	---	---	dahu	*ma-ndaut	*ma-ʒaʔu	*ma-tayut
root	*wagar	waʔa	---	---	---	---	---	---	*wa[k/ɣ]ar