PROJECT MUSE®

The "Mystery Aspirates" in Philippine Languages

Robert Blust

Oceanic Linguistics, Volume 57, Number 1, June 2018, pp. 221-247 (Article)

Published by University of Hawai'i Press

DOI: <https://doi.org/10.1353/ol.2018.0008>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/695554>

The “Mystery Aspirates” in Philippine Languages

Robert Blust

UNIVERSITY OF HAWAII¹

The reconstruction of Proto-Malayo-Polynesian (PMP) *h as a syllable onset continuing Proto-Austronesian (PAN) *S has depended heavily, though not exclusively, on Central Philippine languages, where it is well supported. In coda position, however, the reconstruction of *h has been more difficult. Word-final *h was initially proposed solely on the basis of phonological alternations in Tagalog, and only later shown to be reflected as a coda by Itbayaten of the Batanes Islands, and less securely by Aklanon in the Bisayas. New evidence for PMP *-h continuing PAN *S is introduced here from Central Luzon languages, Bisayan languages, and Mamanwa of Mindanao. In addition, revised material on Aklanon now shows that it, too, preserves coda /h/. However, not all examples of *-h in Philippine languages can be traced to PAN *S. Rather, a surprising number of these segments correspond to zero in most previous reconstructions. More problematically, the cognate sets that contain *-h involve multiple sound correspondences, raising difficult questions about their origin.

1. INTRODUCTION.¹ Laryngeal consonants have a long and troubled history in Austronesian (AN) languages (Brandstetter 1916:86, 248, 273, 282–85; Dempwolff 1934–38; Dyen 1953a, 1962, 1965; Tsuchida 1976; Zorc 1982, 1996; Blust 2013:546–53, 567–74). Unlike other segments that have generated controversy in phonological reconstruction, such as *z (Dyen 1951), *R (Dyen 1953b), *b (Prentice 1974), or *d (Dahl 1976:58ff; Ross 1992:40ff), laryngeals in a small number of languages often correspond to zero in the majority of witnesses, raising questions about whether they have been added in a few languages, or lost in all others.

Dempwolff’s reconstruction of the *Uraustronesisch* laryngeals was methodologically flawed, and in correcting it Dyen (1953a) reconstructed Proto-Austronesian (PAN) *q and *h. Because the second of these consonants has sibilant reflexes in a number of Formosan languages, it was later replaced by *S, and then split into six subvarieties, *S₁–*S₆ (Dyen 1965; Tsuchida 1976). For more than a quarter of a century, these two segments

1. I am grateful to an anonymous referee, and especially to R. David Zorc for his exceptionally thorough review of an earlier version of this paper, which contributed substantially to its improvement. I also wish to thank Jason Lobel, who drew my attention to Mamanwa and the northern and eastern Samar dialects of Waray-Waray as additional witnesses for the “mystery aspirates.” Needless to say, any remaining errors of presentation or interpretation are mine alone.

were known as the PAN “laryngeals,” even though *q appears to have been a pharyngeal stop and *S an alveolar sibilant (Blust 2013:553ff).

In a paper notable more for its recognition of numerous new protophonemes on the basis of limited data than for its reconstruction of a plausible phonological system, Dyen (1965:301ff) proposed what he called “The *x-reconstructions” (*x₁, *x₂, *X), and “The H- and ?- correspondences,” with *H- assigned to three forms and *? to two. To a large extent, this approach, which posited new protophonemes without concern for possible phonetic correlates, was disfavored, and for nearly two decades most Austronesianists apart from Dyen’s own students (for example, Tsuchida 1976) left the matter alone.

As will be seen below, these proposals were developed on the basis of better evidence on the Formosan languages by Tsuchida (1976). In addition, Zorc (1982, 1996) raised the question whether PAN might have had a set of true laryngeals, noting that these are phonetically the weakest consonants, hence those that are most likely to disappear early in the history of a language family. Rather than follow the lead of Dyen (1965), he struck out in a new direction, proposing PAN *-? on the basis of a set of fairly richly attested sound correspondences different from that for which either *q or Dyen’s fragile *-? was reconstructed. In addition, he proposed PAN *h based on evidence different from that used to justify *S.

Zorc’s proposal for *-? was taken more seriously by scholars interested in the phonology of PAN because it was based on a conventional application of the comparative method, which uses converging lines of evidence as support for new distinctions, rather than basing these entirely on irregularities in a single language, and in some cases in a single form, as was the case with many of the reconstructions in Dyen (1965). However, despite its promising beginnings, the reconstruction of *-? has proven problematic (Blust 2013:567–74), and although Zorc also considered reflexes of Dyen’s *H as refined by Tsuchida (1976), the Philippine evidence he proposed for *-H rarely included languages with an unambiguous coda -h.

The sections that follow address two categories of word-final aspirates: -h from PAN *S, and -h in a handful of languages corresponding to zero in all others. Although the first of these categories involves a relatively straightforward application of the comparative method, the second presents a methodological conundrum, and these instances of -h will, for lack of a better term, be called the “mystery aspirates.” Our primary concern, then, will be with the appearance of word-final -h in certain languages of the Philippines where previously their Proto-Malayo-Polynesian (PMP) bases were thought to end with a vowel. But first, to avoid possible confusion, it will be best to review /h/ in Philippine languages as a reflex of PAN *S.²

2. PHILIPPINE REFLEXES OF *-S. The PAN sibilant *S is reconstructed in initial, medial, and final positions. In Proto-Malayo-Polynesian, the immediate ancestor of all AN languages outside the island of Taiwan, *S became *h, as shown in (1):

2. The phoneme /h/ in Philippine languages naturally has other sources as well, as PAN *s (Ayta Abellen, Ifugaw, etc.), *j (Casiguran Dumagat), or *l (sporadic in Tagalog). Since these are irrelevant to the present discussion, they will not receive further attention except in passing.

(1) PAN	*Sajek	PMP	*hajek	‘to sniff, smell’
Saisiyat	s<om>azek	Itbayaten	harek	
Pazeh	sa-sazek	Tagalog	halík	
Kavalan	sanek	Bikol	hadók	
Amis	sanek	Soboyo	hayo?	
PAN	*Sasaq	PMP	*hasaq	‘to whet, sharpen’
Bunun	sasaq	Tagalog	hása?	
Paiwan	t<m>ataq (<A)	Soboyo	hasa	
PAN	*CuSuR	PMP	*tuhuR	‘to string, as beads’
Kavalan	tusuR	Itbayaten	tohoy	
Bunun	ma-tusul	Tagalog	túhog	
Paiwan	tsusu	Hanunóo	túhug	
PAN	*iSiq	PMP	*ihiq	‘urine; to urinate’
Amis	isi?	Tagalog	íhi?	
Favorlang	isi	Bikol	íhi?	
Rukai (Maga)	isii	Hanunóo	íhi?	
Paiwan	isiq	W. Bukidnon		
		Manobo	íhi?	
PAN	*CiŋaS	PMP	*tiŋah	‘food stuck in teeth’
Saisiyat	ŋiŋaŋ	Itbayaten	tiñah	
Pazeh	siŋas	Tagalog	tiŋá	
Amis	tiŋas	Bikol	tiŋá	
Paiwan	tsiŋas	Cebuano	tiŋá	
PAN	*CumeS	PMP	*tumah	‘clothes louse’
Kavalan	tumes	Itbayaten	tomah	
Saisiyat	somæh	Tagalog	túma	
Amis	tumus	Bikol	túma	
		Hanunóo	túma	

At least since Dyen (1953a), the most reliable witnesses for PMP *h have been Tagalog and other Central Philippine languages, together with a few Philippine languages that Dyen did not consider, most notably Itbayaten (ITB), one of the four languages in the Batanes Islands between Taiwan and Luzon, Hanunóo and perhaps other languages of Mindoro, and the Manobo languages of Mindanao. In addition, Blust (1981) found that Soboyo, a Central Malayo-Polynesian language in the central Moluccas of eastern Indonesia, preserves PAN *S- as word-initial *h*-.

What stands out in data sets (1)–(6) is that even in languages that retain *S as *h* in onset position, it is lost word-finally in all languages but ITB. The only qualifications required in this statement are that *-h* appears in Tagalog alternations such as *abó* ‘ash’ : *abuh-án* ‘ash-pit’, and that the West Bisayan language Aklanon (Zorc 1969) sometimes reflects PAN *S- as *-h* in absolute final position. However, comparative evidence has shown that the thematic *-h* in Tagalog *abuh-án* and many other forms is historically secondary, thus negating its value as evidence for PMP *-h in cases like *tubó* ‘sugarcane’ : *tubuh-án* ‘sugarcane plantation’, where it *does* reflect PAN*S (cf. PAN *tebuS ‘sugarcane’).³

The situation is somewhat different in Aklanon as presented in Zorc (1969), where many examples of apparently nonetymological *-h* are given, as in PAN *apu ‘grandpar-

ent or grandchild (recipr.)' > *apó(h)* 'grandchild', *SadiRi > *haligi(h)* 'column, post, mainstay, support (of house)', *lima > *limá(h)* 'five', or *ta-telu 'three (of humans)' > *tátlo(h)* 'to raise or lower to three, make three' (cp. *tátlo* 'three'); and in still other cases PAN *-S is reflected as zero, as in *CiñaS > *tiñá* 'food particles caught between teeth', or *tebuS > *tubó* 'sugarcane', *ka-túbw-an* 'sugarcane plantation'.

Zorc (pers. comm., January 6, 2018) now says that his dictionary contains many errors in the representation of *-h*, and he has provided an extensive set of forms in a revised orthography which, in accordance with his suggestions, will be followed here. Unlike the case in Tagalog, where *-h* is commonly added before a suffix whether it is etymologically justified or not, Aklanon *-h* is said to appear both medially and word-finally as a syllable coda, where it is contrastive.

Given the situation that prevailed until recently, the reconstruction of PMP *-h based on coda /h/ had depended entirely on ITB, which alone of the four Batanic languages (Yami, Itbayaten, Ivatan, Ibatan) reflects PAN *-S in all positions, as shown in dataset (2) (Tsuchida, Yamada, and Moriguchi 1987):

(2)	PAN	*Sapuy	'fire'
	PMP	*hapuy	'fire'
	Yami (Imorod)	apoy	
	Itbayaten	hapoy	
	Ivatan (Isamorong)	apoy	
	Ibatan	apoy	
	PAN	*bukeS	'head hair'
	PMP	*buhek	'head hair'
	Yami (Imorod)	ovok	
	Itbayaten	vohok	
	Ivatan (Isamorong)	vo:ok	
	Ibatan	—	
	PAN	*CiñaS	'food particles in teeth'
	PMP	*tiñah	'food particles in teeth'
	Yami (Imorod)	ciña	
	Itbayaten	tiñah	
	Ivatan (Isamorong)	tiña	
	Ibatan	tiña	

While PAN *-S > PMP *-h is attested both initially and medially in various Philippine languages, and initially in Soboyo of the central Moluccas, then, it has appeared until now that a nonzero reflex of PAN *-S in Malayo-Polynesian (MP) languages is confined to a single reliable witness, namely Itbayaten of the Batanes Islands, with the small qualification that Cebuano Bisayan reflects coda *S as postconsonantal medial *h*, as in PAN *kuSkus 'scrape' (> *kuhkuh > *kuhku) > Cebuano *kukhú* 'scrape, scratch off something that sticks to a surface'. However, even in Cebuano, *-S is lost in absolute final position.

3. As will be seen, the PAN word for 'ashes' may well have had a final laryngeal of some type, but it did not contain *-S. Since morphophonemic *-h* appears regularly in Tagalog before a vowel-initial suffix to separate two unlike vowels, it has no diagnostic value for the reconstruction of a final laryngeal.

Given this rather longstanding situation, it comes as a surprise to discover that ITB is only one of several geographically scattered and genetically diverse Philippine languages that reflect PAN *-S as *-h*. A corollary of this discovery is that even where documentation is relatively dense, as with the Philippines, both minor languages and dialects of major languages may preserve archaic phonological features that have disappeared from all or nearly all of their better-described relatives. In the case of the Philippines, this turns out to be particularly true of languages spoken by the aboriginal Negrito population, a pattern that achieves a special irony in view of the fact that this population must have acquired ancestral forms of the languages they now speak by language shift (Reid 1987).

2.1 AYTA ABELLEN. Ayta Abellen (AYA) is a member of the Central Luzon subgroup of Philippine languages (Stone 2008; Himes 2012), a collection of still poorly described and generally small language communities that includes “Kapampangan (or Pampango), Sinauna (or Sinauna Tagalog), three dialects of Sambal (Bolinao, Tina, and Botolan), and a number of languages spoken by Ayta Negrito populations” (Himes 2012:490). By far the largest language in this group is Kapampangan, but because it plays no part in the following discussion, it will not be discussed further. Rather, the phonologically more conservative languages appear to be those of the Ayta Negrito groups, many of which were displaced after the catastrophic eruption of Mt. Pinatubo on June 15, 1991. Of these, Ayta Abellen, represented by a useful online dictionary (Stone 2007), stands out as particularly valuable for the purpose at hand.

The first clue that this is the case comes in seeing forms such as AYA *kokoh* ‘fingernail, toenail’ (PAN *kuSkuS, Tagalog *kukó*), or *toboh* ‘sugarcane’ (PAN *tebuS, Tagalog *tubó*). However, before reaching any conclusion, it is necessary to determine that this final aspirate (i) is distinctive, unlike the *-h* that was regularly added in many languages of northern Sarawak (Blust 1969:91), and in at least some dialects of Tboli and Tausug in the southern Philippines (Reid 1971), and (ii) is not a back-formation from a paradigm like Tagalog *tubó* ‘sugarcane’: *tubuh-án* ‘sugarcane plantation’. With only a few exceptions, to be noted below, words reconstructed with a final vowel in PAN have a final vowel in AYA: *lima > *lima* ‘five’, *maCa > *mata* ‘eye’, *Suaji > *ali* ‘younger sibling’, *ba-bahi > *babayi* ‘female’, *qaNiCu > *anito* ‘ghost, spirit of the dead’, *asu > *aso* ‘dog’, and so on. This strongly suggests that the *-h* in words like *kokoh* or *toboh* is a retention of an original final consonant.

The most complete demonstration that AYA reflects PAN *-S as *-h* is through a tally of all forms with PAN *-S that have reflexes in ITB, AYA, or both in the online open access *Austronesian comparative dictionary* (Blust and Trussel ongoing). This is shown in table 1.

The data in table 1 fall into two categories: forms that show *S metathesis, and those that do not. *S metathesis is a semi-regular (recurrent) sound change that distinguishes PMP from PAN.⁴ The essential condition for this innovation is a PAN sequence *-CVS, where C was a stop and *-CVS did not occur in a reduplicated monosyllable such as

4. Whether metathesis occurred before *S > *h* or vice versa remains an open question, and if the order was the reverse of that assumed here, we might prefer to call this *-h metathesis. However, preliminary data from Kavalan and Amis suggest that metathesis had begun before sibilant lenition, as seen in the Amis variants *toris/tosir* ‘a line’, and *tosor/toros* ‘knee’, or Kavalan *tusuz* ‘knee’, as compared with reflexes of PAN *tuduS ‘knee’ in Puyuma (*tuqu*), Siraya (*turux*; Adelaar 2011:389), and the Amis variant *toros*.

TABLE 1. ITBAYATEN AND AYTA ABELLEN REFLEXES OF PAN *S

	PAN	PMP	ITB	AYA	
1.	*baRiuS	*baRiuh	—	(bagyo)	typhoon
2.	*bukeS	*buhek	vohok	—	head hair
3.	*CebuS/tebuS	*tebuh	—	toboh	sugarcane
4.	*CiŋaS	*tiŋah	tiŋah	—	food in teeth
5.	*CumeS	*tumah	tomah	tomah	clothes louse
6.	*daSdaS	*dahdah	rahdah	—	chest (male)
7.	*kaSiw	*kahiW	kayoh (<M)	kayo (<M)	wood, tree
8.	*kaSkaS	*kakhah	—	kakhah	to scrape, rasp
9.	*kiSkiS	*kikihi	—	kikihi	to scrape off
10.	*kuSkuS	*kukuh	kokoh [†]	kokoh	claw, nail
11.	*kuSkuS	*kuhkuh	kohkoh	kohkoh	to scrape
12.	*liseqeS	*lisehaq	lisaha	—	nit, louse egg
13.	*piRaS	*piRah	piah	—	roe, fish eggs
14.	*qiluS	*qiluh?	—	ilo	wipe the anus
15.	*tapeS	*tahep	tahep	—	to winnow
16.	*tiR(e)peS	*tiR(e)pah	tipah	—	spittle
17.	*tuduS	*tuhud	tohod	—	knee
18.	*tuqaS	*tuhaq	toha	—	old, mature

[†] Despite their identical shape in PAN, items 10 and 11 show consistently different outcomes in PMP, the word for ‘fingernail, claw’ sporadically losing its preconsonantal *h.

*kuSkuS ‘claw, fingernail’ (Blust 1993:178–79). In table 1, this can be seen in nos. 2, 12, 15, 17, and 18, all of which show a change from PAN *-CVS to PMP *-hVC (the metathesis in *kaSiw > ITB *kayoh*, AYA *kayo* ‘wood, tree’ is sporadic, and unconnected with *S metathesis as described here). A sixth example, which lacks reflexes in either of these languages, is PAN *CaqiS > PMP *tahiQ ‘to sew’. *S metathesis is thus recurrent, although there are unexplained exceptions, as with PAN *CebuS/tebuS > PMP *tebuh (not **tehub) ‘sugarcane’, and PAN *tiR(e)peS > PMP *tipah (not **tihaP) ‘spittle’.

In metathesized forms, PAN *-S is reflected as medial *h* in many Philippine languages, but evidence for word-final *-h* from PAN *-S is much harder to find. What is striking about the forms in table 1 that have not undergone *S metathesis is that ITB and AYA agree in three of the four cases where both have relevant data: *tomah* ‘clothes louse’, *kokoh* ‘claw, fingernail’, and *kohkoh* ‘to scrape or scratch up’. Both the agreement with ITB and the more general correspondence with PAN *-S make it clear that Ayta Abellen is no less a witness for PMP *-h than is Itbayaten. Between these two languages, there are 12 nonborrowed reflexes of PAN forms with *-S that have not undergone *S metathesis, and 10 of them, or over 83 percent, have *-h*. The two exceptions are AYA *kayo* ‘wood’ and *ilo* ‘wipe the anus’. Although the absence of *-h* in these words is unexplained, the similar problem in the word for ‘typhoon’ is clearly a result of borrowing, since PAN *R regularly becomes *y* in all Sambalic languages. The question that this overwhelming agreement inevitably raises is whether there are other witnesses in the Philippines for word-final *h* as a reflex of PAN *S.

2.2 OTHER PHILIPPINE LANGUAGES REFLECTING PAN *-S AS *-h*.

So far, only Itbayaten and Ayta Abellen have provided clear support for PMP *-h as a continuation of PAN *-S. However, since these languages belong to different Philippine

microgroups (Blust 1991), it is likely that other Philippine languages preserve the contrast of -V vs. -Vh, and that this distinction has been underrepresented in the descriptive literature. A concerted effort to find such cases has, in fact, proved fruitful.

To date, discussions of the historical phonology of Central Luzon (CL) languages (Stone 2008; Himes 2012) have not recognized the retention of PAN *S as *h*.⁵ This oversight is surprising in view of the clear retention of *-S as a word-final aspirate in Ayta Abellen shown in table 1, but it cannot be judged too harshly in view of the limited number of examples and the absence of the contrast in the best-known languages of the group, namely Kapampangan and Botolan Sambal.

Other CL languages that have been investigated to date present a mixed picture: there is some evidence for retention of *-S as an aspirate, but it is limited in the number of forms available, in some cases as a result of phonotactic conditions. The cases uncovered so far are presented in the following subsections.

2.2.1 Tina Sambal. Although the better-known Botolan Sambal invariably reflects *-S as zero, the Tina dialect has two known examples that support the view that PAN *-S became -h (Elgincolin, Goschnick, and Elgincolin 1988):

(3)	PAN *kuSkuS ‘claw, fingernail’	>	kokóh ‘hoof’
	PAN *tebuS	>	tobóh ‘sugarcane’

Since the final aspirate in these words contrasts with a final vowel in PAN *asu > Tina Sambal *aso* ‘dog’, *kuCu > *koto* ‘head louse’, or *siku > *hiko* ‘elbow’, it is clear that it is not a product of regular sound change, and we have little choice but to assume that PAN *-S was retained as -h in at least these two words, or that it was acquired by borrowing these forms from neighboring languages in which *-S was retained as -h. Needless to say, if we did not have the much richer data from Ayta Abellen, Tina Sambal would provide a very insecure foundation for the claim that PAN *-S was retained as -h in Proto-Central Luzon.

It is likely that this language has more examples of -h than are reported here, but the only dictionary currently available for it is highly provisional, and the lexical categories that it represents are based on English rather than the native system.

2.2.2 Ayta Mag-Antsi. This CL language also provides limited evidence for retention of PAN *-S as *h*, but although it is limited, it is conditioned in such a way as to leave little doubt that the supporting evidence is valid.

PAN *-S in nonreduplicated bases disappeared without a trace, as seen in the criterial cases in (4):

(4)	*tebuS	>	tubó ‘sugarcane’
	*CiŋaS ‘food particles caught between teeth’	>	tsiŋa ‘toothpick’

However, in reduplicated monosyllables that had the shape CVSCVS in PAN, the coda is retained as an aspirate in both syllables, as shown in (5):

5. Himes (2012:492), who is quite explicit about this, states that “at some time prior to the dispersal of the Central Luzon (LUZC) languages, certain of the PMP phonemes had undergone change.” Among these he lists *h > Ø, adding in a footnote that cites a personal communication from Lawrence Reid, that *h may have become ? in initial position, “but *h certainly was lost in all other environments.”

- (5) *kaSkaS > kakah 'to shave, scrape'
 *kiSkiS > kihkih 'to scrape'
 *kuSkuS > kuhkuh 'to scratch (as a cat scratching)'

Although the number of examples is small, then, there is complete consistency in showing that PAN *S in simplex bases disappeared word-finally in Ayta Mag-Antsi, but was retained in the reflexes of reduplicated monosyllables, whereas in Ayta Abellen PAN *S was retained in both environments.

The one case that might be considered questionable, but that turns out on closer inspection to further confirm this generalization, is the following:

- (6) *kuSkuS > koko 'claw, fingernail or toenail'

Since this word and *kuSkuS 'to scratch' were homophones in PAN, we would expect them to show the same development in MP languages, but as noted briefly in the footnote to table 1, they are consistently different in Philippine languages, as with Ayta Abellen *kokoh* 'claw, fingernail' vs. *kohkoh* 'to scrape or scratch up', or Cebuano *kukú* 'nail, claw' vs. *kukhú* ~ *kalukhú* 'scrape, scratch off something that sticks to a surface'.

2.2.3 Bolinao. Although few reflexes have been found so far, data in McFarland (1977) suggests that PAN *-S disappeared in Bolinao without a trace, as seen in (7):

- (7) *kuSkuS 'claw, nail' > kuko 'fingernail'
 *CiŋaS > tiŋa 'food caught in teeth'

Relevant material for other CL languages, apart from very limited data for Remontado (Sinauna), have not yet been accessed.⁶

2.2.4 Waray-Waray. For the northern and eastern Samar dialects of Waray-Waray in the central Philippines, Jason Lobel (pers. comm., August 14, 2017) has reported the occurrence of coda *h* in reduplicated monosyllables only. Most of the examples he has provided lack known PAN etymologies, but two comparisons show that *-h* reflects PAN *-S:

- (8) *kaSkaS > kakah 'to scratch an itch'
 *kiSkiS > kihkih 'to shave off'

These dialects of Waray-Waray, thus, show a pattern of retention for PAN *-S that is identical to that in Ayta Mag-Antsi, namely, preservation as *-h* in reduplicated monosyllables, but loss in simplex bases.

2.2.5 Mamanwa. Mamanwa, spoken by a Negrito population in northeast Mindanao, has not previously been reported as allowing *-h* (for example, Reid 1971). However, Jason Lobel (pers. comm., October 5, 2017) has recorded a clear contrast of final *-h* vs. final vowel in at least one dialect of the language, as shown in table 2. For *-V > -V, cf. PAN *maCa > *matá* 'eye', *lima > *limá* 'five', *ba-bahi > *babazi* 'woman', or *batu > *bató* 'stone'.

6. The nomenclature relating to this language has had a checkered history. While early reports described it as "Sinauna Tagalog," it eventually became clear that it is not a Tagalog dialect, but rather a Central Luzon language spoken by a Negrito population. Lobel (2013, map 1.17) described it as "Remontado Dumagat," but he now (pers. comm., February 7, 2018) states that the people do not call themselves or consider themselves "Dumagats," and that the only appropriate name for it is "Remontado."

TABLE 2. MAMANWA REFLEXES OF PAN *S

	PAN	PMP	Mamanwa	
1.	*CiŋaS	*tiŋah	tiŋáh	food stuck in teeth
2.	*CumeS	*tumah	tumáh	louse
3.	*tebuS	*tebuh	tubóh	sugarcane
4.	*kaSkaS	*kahkah	kahkah	to scratch up, as chicken does
5.	*kiSkiS	*kihkih	kihkih	to scratch an itch

2.2.6 Aklanon. Although his 1969 dictionary leaves the matter somewhat in limbo, as noted earlier, Zorc has now rechecked all examples of word-final *-h* in Aklanon, and finds that the forms given in table 3 occur as true word codas reflecting PAN *S. Of these, items 2, 7, 8, and 11 show *S-metathesis (with subsequent loss of *-h* in *eusá?*, for expected *eusahá?* ‘nit, louse egg’). The remaining seven words reportedly preserve word-final */h/* intact.

TABLE 3. AKLANON REFLEXES OF PAN *S

	PAN	PMP	Aklanon	
1.	*baRiuS	*baRiuh	bágyoh	typhoon
2.	*bukeS	*buhek	buhók	head hair
3.	*buReS	*buRah	bugáh ‘exhale’	to spew from mouth
4.	*kaSkaS	*kahkah	káhkah	to scratch at the ground
5.	*kiSkiS	*kihkih	kíkih	to scrape off; whisk away (?)
6.	*kuSkuS	*kukuh	kukúh	nail, claw
7.	*liseqeS	*lisehaq	eusá?	nit, louse egg
8.	*tapeS	*tahep	tahóp	to winnow
9.	*tebuS	*tebuh	tubóh	sugarcane
10.	*tiŋaS	*tiŋah	tiŋáh	food stuck in teeth
11.	*tuduS	*tuhud	túhod	knee

2.2.7 Summary. To summarize, with rare exceptions, PAN *S is reflected consistently in all Philippine languages that have been examined, although the nonzero reflexes fall into four types:

TYPE 1: Languages that reflect PAN *S as *h-*, *-h-*, *-ØC-*, *-Ø*:

- Central Philippines: Tagalog, Bikol, Hiligaynon, Tausug
- Mangyan: Hanunóo
- Manobo: Ata, Western Bukidnon Manobo, Tigwa Manobo, Binukid

TYPE 2: Languages that reflect PAN *S as *h-*, *-h-*, *-Ch-*, *Ø*:

- Central Philippines: Cebuano

TYPE 3: Languages that reflect PAN *S as *h-*, *-h-*, *-hC-*, *-h*, but the last two of these only in reduplicated monosyllables:

- Central Luzon: Ayta Mag-Antsi
- Central Philippines: Waray-Waray (northern and eastern Samar)

TYPE 4: Languages that reflect PAN *S as *h-*, *-h-*, *-hC-*, *-h* without qualification:

- Batanic: Itbayaten
- Central Luzon: Ayta Abellen, Tina Sambal
- Central Philippines: Mamanwa, Aklanon

3. THE “MYSTERY ASPIRATES”. So far, we have found a strikingly regular pattern in which PAN *-S—based on sibilant reflexes in various Formosan languages—predicts *-h* in Itbayaten and Ayta Abellen with nearly perfect regularity, with additional support from Tina Sambal, Ayta Mag-Antsi, Aklanon, Mamanwa, and some dialects of Waray-Waray. However, there is a puzzling asymmetry in this process: what is the result if we start with *-h* in a Malayo-Polynesian witness that corresponds to zero in most other languages, and try to trace its origin? Does this segment reliably predict a corresponding sibilant in Formosan languages that reflects PAN *-S? Surprisingly, the answer is ‘no’, but before discussing the correspondences that I will call the “mystery aspirates,” it is necessary to clear away some potentially obscuring features of the problem.

Yamada (2002) lists 294 Itbayaten words with *-h*. Most of these have no etymology, but those that do show that *-h* generally reflects PAN *-S, with some exceptions to which we turn shortly. By contrast, Ayta Abellen merged *S and *s as *h* in coda position (in onset position *s also became *h*, but *S disappeared). The lenition of *s is seen both in native words (PAN *beRas > *beyah* ‘husked rice’, *Sipes > *ipeh* ‘cockroach’, *Caŋis > *taŋih* ‘to cry’), and in Spanish loans, including many nouns that were borrowed in their plural forms in Philippine languages (*dioh* ‘God’, Spanish Dios; *bayawah* ‘guava’, Spanish *guayabas*; *hiboyah* ‘onion’, Spanish *cebollas*; *maih* ‘corn, maize’, Spanish *maiz*; *kamatih* ‘tomato’, Spanish *tomates*). Once these cases are excluded, we are left with a small number of native words ending with a word-final aspirate that cannot be derived from PAN *-s or *-S. These are shown without reconstructions in table 4, but with an ‘X’ indicating that a protoform is possible on the stated level with the available data. The problem, in effect, is to solve for ‘X’.

Each of these words is reflected in ITB, AYA, or both with a final *h* that is not explained by the established PMP reconstructions, all of which end with a vowel, and it is clear that some hypothesis must be formulated to account for this departure from expectation. We have essentially two choices: (i) consider the final aspirates innovations, or (ii) consider them reflexes of phonemes that have not previously been reconstructed. The first choice is quickly ruled out on the grounds that many words in both languages have a final vowel where this is expected based on established reconstructions, making the *-h* in table 4 unconditioned, and hence a violation of the regularity hypothesis. This leaves us with the second choice: the final aspirates are retentions of consonants in Proto-Philippines (PPH), PMP, and in some cases PAN.⁷ So, what do we reconstruct? The commonsense solution is **-h*, but this turns out to be more problematic than first impressions might suggest.

Even before incorporating evidence from other languages in these comparisons, it can be seen that three of the four cognate sets in table 1 that are represented by both languages (that is, nos. 5, 10, and 11) agree in reflecting PAN *-S as *-h*, but the eight cognate sets in table 4 that are shared by both languages exhibit multiple sound correspondences (C1–C3), as shown in (9):

7. Not all scholars accept Proto-Philippines (Reid 1982; Ross 1995:73ff, 2005; Smith 2017). Nothing in the current argument depends upon the Proto-Philippines hypothesis, although the presence of over 1,300 Philippine-only etymologies in the online *Austronesian comparative dictionary* (Blust and Trussel ongoing) is difficult to explain without assuming the reality of a Philippine subgroup as formulated by Zorc (1986) and Blust (1991).

TABLE 4. ITBAYATEN AND AYTA ABELLEN *-h*
FROM UNKNOWN SOURCES

	PAN	PMP	PPH	ITB	AYA	
1.		X	X	—	apah	rice husk, chaff
2.		X	X	arah	—	to be, exist, have
3.	X	X	X	vayah ‘red’	bayah	ember, live coal
4.		X	X	voko	bokoh	joint, node, knuckle
5.		X	X	hadpa [†]	depah	fathom
6.	X	X	X	—	diwih	thorn
7.	X	X	X	d-inoh	no	where/when?
8.	X	X	X	itoh	—	being there
9.	X	X	X	—	iyoh	shark
10.		X	X	ninih	—	earthquake
11.			X	omah	oma	kiss
12.	X	X	X	sah	—	one
13.	X	X	X	—	tolih	ear wax
14.	X	X	X	koto	kotoh	head louse
15.		X	X	—	lejah	sesame
16.		X	X	—	nayah	narra tree [‡]
17.			X	—	panjah	jaw, jawbone
18.	X	X	X	pirah	—	how much/many?
19.		X	X	—	ponoh	bunch, cluster
20.		X	X	sasah	—	cut palm leaf roofing
21.	X	X	X	sichoh	hiko	elbow
22.		X	X	sisih	—	marine gastropod
23.	X	X	X	tova	tobah	derris root fish poison

[†] Cp. Yami (Imorod), Ivatan (Isamorong) *adpa* ‘fathom’, pointing to Itbayaten *hadpa* as a metathesis of earlier *adpah*. For Itbayaten reflexes of PMP *CeCV(C), cf. Blust (2017:494–95)

[‡] Also ITB *nara* ‘a tree: *Pterocarpus indicus*’, which is either a loan or a reflex of a doublet *nara (Blust and Trussel ongoing).

- (9)
	ITB	AYA	
C1	h	h	(3,5)
C2	Ø	h	(4, 14, 23)
C3	h	Ø	(7, 11, 21)

Since the final aspirates in table 4 do not continue PAN *-S, it is clear that there are at least four different sound correspondences in Philippine languages that appear to reflect some type of glottal spirant. By mechanically mapping sound correspondences onto proto-phonemes we might posit four reconstructed segments that are reflected as glottal fricatives or zero, as in (10):

- (10) *h₁: *tebuh₁ ‘sugarcane’, *tumah₁ ‘clothes louse’, *kih₁kih₁ ‘scrape off’, etc.
 *h₂: *baRah₂ ‘ember’, *depah₂ ‘fathom’
 *h₃: *bukuh₃ ‘joint, node, knuckle’, *kutuh₃ ‘head louse’, tubah₃ ‘derris root fish poison’⁸
 *h₄: *-nuh₄ ‘interrogative marker’, *umah₄ ‘kiss’, *siku₄ ‘elbow’

8. Cp. AYA *tobah* ‘a species of shrub from the seeds of which croton oil is extracted’, Ilokano *tuba* ‘fish poison, *Croton* sp. plant’, *tuba-en* ‘to poison fish in the water’, as against reflexes of PAN *tuba ‘fish poison: *Derris elliptica*’ in most languages.

This is the procedure that was followed by Dyen (1965) and Tsuchida (1976) for PAN, although it has not previously been proposed on the level of PMP, let alone PPH. Where a reflex is known in only one witness for PMP/PPH *-h, and diagnostic Formosan cognates are unknown, the reconstruction is ambiguous, as shown in (11):

- (11) *h_(2,3): *qapah_(2,3) 'empty husk (of grain)'
 *duRih_(2,3) 'thorn'
 *qiSuh_(2,3) 'shark'
 *Culih_(2,3) 'earwax'
 *leŋah_(2,3) 'sesame'
 *ñaRah_(2,3) 'the narra tree: *Pterocarpus indicus*'
 *paŋah_(2,3) 'jaw, jawbone'
 *puŋuh_(2,3) 'bunch, cluster'
 *h_(2,4): *sasah_(2,4) 'cut or collect palm leaves for roofing'

At this point, it may be worthwhile to consider the Formosan evidence for final aspirates, since Tsuchida (1976) posited just two, namely *-H₁ and *-H₂, which may have been *-h and *x, respectively, as will be shown below. If the Formosan and Philippine evidence is concordant, *h₁ in (10) would correspond to PAN *-S, *h₂ would correspond to PAN *-h, and *h₃ would correspond to PAN *-x, leaving *h₄ as the only unexplained exception. However, to determine whether this is the case, we must first integrate the Formosan evidence with the evidence for word-final aspirates in Itbayaten and Ayta Abellen, and conduct a further search for Philippine witnesses for word-final aspirates that do not reflect PAN *-S.

4. FINAL ASPIRATES IN FORMOSAN LANGUAGES. As noted earlier, both Dyen (1962, 1965) and Tsuchida (1976) found evidence in several Formosan languages for final aspirates where Dempwolff (1934–38) had reconstructed final vowels. Given the newly uncovered evidence for similar segments in Philippine languages, it is important to see whether the sound correspondences in these two areas are concordant. Table 5 summarizes the Formosan witnesses identified in Tsuchida (1976) with his reconstruction of PAN word-final aspirates, (where 1 = Atayal, 2 = Saisiyat, 3 = Pazeh, 4 = Amis, 5 = Saaroa, 6 = Bunun (Takituduh), and 7 = Seediq).

The cross-linguistic agreement between several Formosan languages (especially Saisiyat, Pazeh, Amis, and Takituduh/Northern Bunun) in distinguishing these correspondences from zero and from one another is consistent enough to warrant the conclusion that PAN had two word-final voiceless spirants in the back of the vocal tract. Only the Atayalic languages (Atayal, Seediq) provide a clue to the possible phonetic difference between them, namely *H₁ = *-h, and *H₂ = *x. This much is typologically reasonable.

But what picture emerges when we try to integrate the evidence for Tsuchida's *-H₁ and *-H₂ with the evidence for final aspirates in Philippine languages? Ideally, in accord with general assumptions about sound change, *h₂ and *h₃ in (10) should correlate with Tsuchida's *-H₁ and *-H₂, leaving just C3 in (9) unexplained. Table 6 alters Tsuchida's *-H₁ and *-H₂ to PAN *-h and *-x and marks the distinct sound correspondences with C1–C7. Because no Philippine language is known to reflect PAN *nunuh 'breast', *Calinax 'ear', or *wiRix 'left side' with a final consonant, these forms are not mentioned further.

TABLE 5. WITNESSES FOR PAN WORD-FINAL ASPIRATES†

	1	2	3	4	5	6	7	
PAN *baqeRuH ₁				-h	-ʔ	-h		new
PAN *baRaH ₁	-h	-h	-h	-h	-ʔ			ember
PAN *buɟuH ₁				-h	-ʔ	-h		skull
PAN *CapuH ₁	-h	-h	-h					to sweep
PAN *CuNuH ₁		-h		-h	-∅	-h		roast, burn
PAN *nunuH ₁			-h		-ʔ		-h	female breast
PAN *q ₂ umaH ₁	-h	-h	-h	-h	-∅	-h		swidden
PAN *tutuH ₁		-h	-h					strike, beat
PAN *amaH ₂		-h		-∅	-ʔ	-h		father
PAN *batuH ₂	-x				-ʔ	-h	-x	stone
PAN *CalijaH ₂			-∅	-∅	-∅	-h		ear
PAN *inaH ₂				-∅	-ʔ	-h		mother
PAN *iniH ₂			-∅		-ʔ			this
PAN *kaliH ₂				-∅	-∅	-h		excavate
PAN *kuCuH ₂		-∅	-∅	-∅	-ʔ	-h		head louse
PAN *pijaH ₂		-∅		-∅	-∅	-h		how many?
PAN *qabuH ₂		-∅	-∅	-∅	-ʔ	-h		ash
PAN *quluH ₂		-h					-x	head
PAN *taRaH ₂			-∅	-∅		-h		to wait
PAN *teluH ₂	-∅	-∅	-∅	-∅	-∅	-h		three
PAN *w ₁ aRiH ₂	-∅			-∅	-∅	-h		dry in the sun
PAN *w ₁ iRiH ₂		-∅	-∅	-∅		-h		left side

† After Tsuchida 1976. Tsuchida (1976:132) also included Aklanon as a witness, despite its problematic status at the time.

Some gaps in his sets have been filled to provide a more complete picture, and sets for ‘to sweep’, ‘head’, and ‘to wait’ have been added. Li (2004b:753–55) writes final vowels everywhere that Tsuchida has *-h* in Takituduh Bunun, but since he also nearly always writes a word-final glottal stop where Tsuchida reconstructs a final vowel, the contrast between *-Vh* : *-V* is maintained as *-V* : *-Vʔ*.

To sum up, PMP must have had **-h* from PAN **-S*, as shown in table 1. That much is uncontroversial, given the nearly perfect correlation of sibilant reflexes in Formosan languages with *-h* in either Itbayaten or Ayta Abellen, or in both. But what is the source of the other instances of *-h* in these two MP languages?⁹

As already noted, the possibility that these segments are innovations has been rejected on general methodological grounds. However, the possibility that they are retentions is hardly better, since to account for all correspondences with *-h*, and hence avoid the recognition of an unconditioned phonemic split, we must posit separate distinctions for each of the correspondences seen in (12):

9. Both PMP and PPH also reflect PAN nonfinal *h* in a small number of forms, as in PAN **bahi* > Western Bukidnon Manobo, Bukidnon *bahi* ‘female’, but considering these is beyond the scope of this paper.

TABLE 6. A PRELIMINARY INTEGRATED SCHEMA FOR FORMOSAN AND PHILIPPINE *-h CORRESPONDENCES

	PAN	ITB	AYA	PAN	PMP	ITB	AYA	
C1	*-h	-Ø	-Ø	*baqeRuh	*baqeRu	vaʔyo	bayo	new
	*-h	—	-Ø	*buŋuh	*buŋu	—	boŋo	skull
C2	*-h	-h	-h	*baRah	*ʔ	vayah	bayah	ember
C3	*-x	-Ø	-h	*kuCux	*ʔ	koto	kotoh	head louse
C4	*-x	-Ø	-Ø	*amax	*ama	ama	ama	father
				*batux	*batu	vato	—	stone
				*kalix	*kali	kali	kali	excavate
				*qabux	*qabu	avo	abo	ash
				*qulux	*qulu	oxo	ō	head
				*telux	*telu	atlo	ta-tlo	three
		-Ø	—	*inax	*ina	ina	—	mother
C5	*-x	-h	—	*pijax	*ʔ	pirah	—	how much?
C6	*-Ø	-Ø	-h	*tuba (?)†	*ʔ	tuva	tobah	fish poison
	*-Ø	—	-h	*duRi	*ʔ	—	diwih	thorn
	*-Ø?	—	-h	*Culi (?)	*ʔ	—	tolih	ear wax
C7	*-Ø	-h	-Ø	*siku	*ʔ	sichoh	hiko	elbow

† Where a PAN reconstruction is possible, but data from Formosan witnesses for a final aspirate is lacking, the reconstruction is followed by a question mark to distinguish it from cases in which diagnostic Formosan witnesses point unambiguously to a final vowel.

- (12) C1 h : Ø : Ø (or its incomplete equivalent) = 3 examples
 C2 h : h : h = 1 example
 C3 x : Ø : h = 1 example
 C4 x : Ø : Ø (or its incomplete equivalent) = 8 examples
 C5 x : h : — = 1 example
 C6 Ø : Ø : h = 3 examples
 C7 Ø : h : Ø = 1 example

Even if we ignore unique exemplifications, we still are left with C1, C4, and C6. Add to this already overburdened reconstruction the need to assume PMP *-h from PAN *-S, and the challenge to the comparative method becomes painfully obvious. It would be pointless to dignify the treatment of these correspondences with eight subscripted varieties of *h, or even four, as none of the attested languages has more than two fricatives in the glottal or velar region, and most have only one (Atayalic languages have -x and -h, described by Li [2004a:628] as “voiceless velar and pharyngeal fricatives respectively”). But the primary data remain, and the question of how to integrate them into a comprehensive and typologically realistic protolanguage will not go away.

There is, moreover, something about the mystery aspirates in MP languages that distinguishes them from the proposals of Dyen (1965) or Tsuchida (1976), who based their claims on data from Formosan languages that most linguists now assign to separate primary branches of the AN family, with the result that the proliferation of protophonemes in their reconstructive strategies was at the remotest possible level. By contrast, the word-final aspirates that would be required to account for the data in Philippine languages following the same approach will be much closer to the present, since at least C2, C3, C5, C6, and C7 need to be distinguished not only in PMP, but also in PPH.

This result already appears to offer no way out: either we posit PMP *-h₁ ... -h₈ (where *-h₁ reflects PAN *-S, and the other aspirates do not), or we recognize widespread irregular change in Philippine languages where -h was added to a final vowel in some lexical items but not others. Both of these options will be considered below, but first we must ask whether ITB and AYA are the only Philippine languages that contain mystery aspirates, since if they are not, the number of sound correspondences that contain -h that is not a reflex of PAN *-S may be even larger.

As noted in an earlier section, the Central Luzon languages Tina Sambal and Ayta Mag-Antsi, as well as Mamanwa, the northern and eastern Samar dialects of Waray-Waray, and Aklanon (Central Philippine languages) all reflect PAN *-S as -h. Given this atypical development, these languages would appear to be plausible candidates as further witnesses for the mystery aspirates. However, as seen in 2.2.7, Ayta Mag-Antsi and the northern and eastern Samar dialects of Waray-Waray are Type 3 languages, which reflect *-S as -h in reduplicated monosyllables, but as zero in simplex bases. In both languages, word-final /h/ is evidently found only in words of the shape C₁V₂hC₁V₂h, and since the mystery aspirates have been attested so far only in simplex bases (table 4), the instances of -h in Type 3 languages can have no bearing on the question at hand. This leaves Tina Sambal, Mamanwa, and Aklanon as prospective sources of -h that does not reflect PAN *-S, and each of these will be examined in turn.

5. ARE THERE OTHER WITNESSES FOR THE “MYSTERY ASPIRATES”? Only Itbayaten and Ayta Abellen have been shown to provide fairly robust support for the reconstruction of PMP *-h < PAN *-S, and each of these languages also contains examples of -h that have more obscure histories. In addition, Tina Sambal reflects PAN *-S as -h both in reduplicated monosyllables and in simplex bases. Given this pattern of development for PAN *-S, it is reasonable to expect that Tina Sambal will also provide additional examples of the mystery aspirates.

The four Batanic languages are sufficiently well described that we can be sure only Itbayaten has a word-final glottal spirant that corresponds to zero in most MP witnesses. However, the situation is descriptively quite different for Central Luzon (CL) languages. In approximate order of descending population size these are:

- Kapampangan
- Tina Sambal
- Bolinao
- Botolan Sambal
- Ayta Abellen
- Ayta Mag-Antsi
- Ayta Mag-Indi
- Ayta Ambala
- Ayta Bataan
- Remontado

The last six languages (“Ayta” varieties plus Remontado) are spoken by Negrito populations that number less than 10,000 individuals, and considerably fewer in most cases. As a result, they have received little attention from linguists as compared

to languages that have larger speaker numbers. Moreover, even the largest CL language, Kapampangan, lacks a dictionary comparable to those of most Philippine major languages. It is clear from the available lexical resources that Kapampangan and Botolan Sambal lack *-h* in cognates of AYA words with the mystery aspirates, but given the data in table 4 there is a pressing need to search for examples of the same type in other languages of this understudied group.

As already noted, Ayta Mag-Antsi can be disposed of quickly. Of the words in table 4 that have a final aspirate in either Itbayaten or Ayta Abellen (or both), Ayta Mag-Antsi has a final vowel in each of the seven forms that have a cognate morpheme: *apa* ‘empty rice husk’, *baya* ‘ember’, *boko* ‘node, joint’, *dapa* ‘armspan, fathom’, *divi* ‘thorn’, *koto* ‘head louse’, *hiko* ‘elbow’. There is one partial exception. As will be seen, Tina Sambal reflects Tsuchida’s PAN *qabuH₂ as *abóh* ‘ash’, and the cognate form in Ayta Mag-Antsi is given as *abo* ~ *aboh* ‘ashes, embers’, suggesting that *-h* in this language is undergoing loss that is complete in some forms, but only partial in others. However, Ayta Mag-Antsi appears to offer no further information regarding the mystery aspirates, and apart from a modified Swadesh 200-word list for what is called “Sinauna” in Greenhill, Blust, and Gray (2003–18) and a small set of functors for Remontado (its proper name) in Lobel (2013), little information is publicly available for other CL languages.

5.1 TINA SAMBAL. The situation in Tina Sambal (TS) is quite different. Like Itbayaten and Ayta Abellen, TS reflects PAN *-S as *-h*, as in *CebuS/tebuS > *tobóh* ‘sugar-cane’, *kuSkuS > *kokóh* ‘claw, fingernail’, and *buReS ‘spray water from the mouth’ > *i-bogáh* ‘to spew’. The last word is particularly interesting, since it is unknown in Itbayaten or AYA, and shows an irregular change *R > g, yet has a final /h/, unlike any other known language in the Philippines. If it is accepted as evidence, it will be the thirteenth case in which *-S > *-h* in a nonmetathesizing MP witness is regular, while at the same time complicating the claim that AYA *bagyo*, which does not show the expected *-h*, is a loan.

However, what is theoretically most troubling about the limited data for TS is the evidence it provides for *additional* sound correspondences that involve the mystery aspirates, specifically in cases where TS has *-h* but AYA has a final vowel. These are summarized in (13):

(13)	PAN	PMP	ITB	AYA	TS	
	*Culi (?)	*?	tilo (<M)	tolih	tolíh	‘ear wax’
	*qabux	*?	avo	abo	abóh	‘ash’
	*qapeju	*?	apdo	aplo	aplóh	‘gall’
	*qiSu (?)	*?	–	iyoh	iyóh	‘shark’

The sound correspondence x : -Ø : -h can be assigned to C4 in table 6, but this invites additional problems, since other TS words that fit this correspondence set, as *bató* ‘stone’, or *oló* ‘head’, show no final consonant, thus splitting C4 into two subtypes, as shown in (14):

(14)	PAN	PMP	ITB	AYA	TS	
	C4.1.	*-x	-Ø	-Ø	-h	(ash)
	C4.2.	*-x	-Ø	-Ø	-Ø	(stone; head)

The correspondence exemplified by ‘gall’ is unclear. Saisiyat *pæʔzoʔ*, Pazeh *apuzu*, Bunun *paqav* ‘bile, gall’ may form a single cognate set with metathesis of the first two consonants either in Saisiyat and Bunun, or in Pazeh, or alternatively they may reflect doublets, *paqejʉ (for Saisiyat and Bunun) and *qapeju (for Pazeh), with a possible final laryngeal. Both Saisiyat and Pazeh could reflect a form with a final vowel, or PAN *-x. Bunun *paqav* (< *paqejʉ with *j > -Ø, *e > a, and *-au > -av) can reflect a form that ended with either PAN *-x or a vowel, unless *paqav* is from the Takituduh dialect, in which case the PAN form based on Formosan evidence would end in a vowel. Bunun *paqav* is cited from Tsuchida (1976:224), who gives it as Southern Bunun (Ishbukun), making it ambiguous for final vowel or *-x. As a result of these uncertainties, the correspondence for ‘gall’ could be either $x : \emptyset : h$ or $\emptyset : \emptyset : h$. If we choose the former, it goes with ‘ash’; but if the latter, it forms an entirely new correspondence in which only TS distinguishes it from zero.

In other comparisons, TS has a final vowel where AYA has -h, as in AYA *bayah*, TS *baya* ‘ember’, AYA *kotoh*, TS *koto* ‘head louse’, AYA *tobah*, TS *toba* ‘derris root fish poison’, or AYA *diwih*, TS *dowi* ‘thorn’. However, the addition of TS data in these comparisons does not produce clear evidence of new sound correspondences, since reference to (12) shows that ‘ember’ can be assigned to C2, ‘head louse’ can be assigned to C3, and the other two terms to C6. Finally, TS has a glottal stop in two comparisons that involve final aspirates in PAN or PMP, as shown in (15):

(15)	$h : - : \emptyset : ʔ : -$	$x : \emptyset : \emptyset : ʔ : -$
	PAN *buŋuh ‘skull’	PAN *amax ‘father’
	PMP *buŋu	PMP *ama
	ITB —	ITB ama
	AYA boŋo	AyA ama
	TS boŋóʔ	TS amáʔ

Even without better data for TS, then, it is clear that there are at least nine distinct sound correspondences involving a word-final aspirate in Itbayaten, Ayta Abellen, or Tina Sambal, together with Formosan evidence, namely reflexes of PAN *-S, C1–C7 in table 6, and the further split of C4 shown in (14).

5.2 MAMANWA. Thanks to the efforts of Jason Lobel, it has become apparent that Mamanwa, a Central Philippine language spoken by a Negrito population in northeast Mindanao, is also a witness for the mystery aspirates. It has already been shown that Mamanwa reflects PAN *-S as -h, both in reduplicated monosyllables and in simplex bases (table 2). However, a number of examples of Mamanwa -h do not reflect *-S. These include the final consonant in *pirah* ‘how much/how many?’, *depah* ‘fathom’, *upah* ‘rice husk’, *liŋah* ‘sesame’, *sikuh* ‘elbow’, and *ubih* ‘yam’.

Mamanwa data are relatively limited, but they do provide evidence of -h in some words that were not previously known to end in a consonant, as with *ubih* ‘yam’, and they agree with the appearance of -h in some other witnesses, as with *sikoh* next to ITB *sichoh* ‘elbow’ for what has traditionally been reconstructed as *siku.

5.3 AKLANON. As noted earlier, Zorc (1969) reportedly was compiled before the author was a linguist. Partly for this reason, and partly because his Aklanon informants objected to writing *-h* in forms where it is phonetically present (but does not occur in the more prestigious national language), final aspirates were written in parentheses, leaving it unclear exactly what was intended. To add to the uncertainty, Spanish loanwords have added *-h* where it is historically secondary, as in *abúno(h)* ‘fertilizer’ (Spanish *abono*), *bála(h)* ‘bullet’ (Spanish *bala*), *dibúho(h)* ‘drawing, sketch’ (Spanish *dibujo*), or *giya(h)* ‘to guide, lead’ (Spanish *giya*). Moreover, in many cases, words with final /h/ exist next to vowel-final bases that represent a different part of speech, as seen in Aklanon *abúno* ‘to fertilize’, *bála* ‘to load, put bullets into a gun’, *dibúho* ‘to draw, sketch’, or *giya* ‘a guide, leader’. To make the final aspirate in Aklanon even harder to interpret, some native forms show a similar pattern of alternation between coda *-h* and zero, as with *tátlo* ‘three’ next to *tátlo(h)* ‘to raise or lower to three, make three’, or *apó(h)* ‘grandchild’, *in-ápo* ‘grandchildren’, *ka-apó-apóh-an* ‘future generations (of grandchildren)’. Add to this the absence of a final aspirate in *CiñáS > *tiñá* ‘food particles caught between teeth’ or *tebuS > *tubó* ‘sugarcane’, *ka-túbw-an* ‘sugarcane plantation’ (now regarded as errors for correct *tiñáh* and *tubóh*), and the reader can readily understand the reluctance of some linguists to accept Aklanon as a witness for *-h* based on the published sources. However, Zorc has rechecked all of his data relating to Aklanon *-h*, and affirms that there is a contrast between words ending with a vowel, and words ending with an aspirate.

6. PUTTING TOGETHER ALL THE EVIDENCE. The impact of Tina Sambal, Mamanwa, and Aklanon data on the correspondences that appear in (12) is summarized below. All correspondences represent equivalent sets, whether they are attested in full or partial form. However, some sets are so minimally attested that they are assignable to multiple correspondence classes, and are consequently so ambiguous as to have little value for establishing contrast.

Because the data that must be considered cannot easily be displayed in a tabular format, they are presented here in individual cognate sets. This is shown in table 7, with Tsuchida’s *H₁ and *H₂ replaced by *h and *x (and where MMN = Mamanwa and AKL = Aklanon). Where Tsuchida lacks a relevant PAN reconstruction, one has been supplied from data in the *Austronesian comparative dictionary* (Blust and Trussel ongoing), as with *duRi ‘thorn’. If Formosan cognates are available but do not include diagnostic witnesses for PAN *-h or *-x, the PAN form is followed by (?), meaning that evidence is inconclusive as to whether it ended in a vowel, or had a word-final consonant that was lost in all surviving forms.

Table 8 attempts to identify the number of distinct sound correspondences in table 7 that are known to have a “mystery aspirate” in Philippine languages. Where available, cognates in Formosan languages are cited in column 1. However, Formosan forms with *-h or *-x are omitted if all known Philippine cognates end in a vowel.

The key to compiling table 8 is to avoid conflicting signals among cognate sets that are assigned to the same correspondence class. Reference to table 7 shows that it is not difficult to find other combinations of cognate sets than those chosen here. For example, sets 6, 7, 9, 23, and 29 are mutually compatible under the assumption that they all derive

TABLE 7. FINAL *-h*: CORRESPONDENCE SETS

1. new h : Ø : Ø : - : h PAN *baqeRuh ITB vaʔyo AYA bayo TS baʔyo MMN – AKL bágʔoh	2. ember h : h : h : Ø : - : h PAN *baRah ITB vayah ‘red’ AYA bayah TS baya MMN – AKL bágah	3. skull h : - : Ø : ʔ : - : ʔ PAN *bujuh ITB – AYA boŋo TS boŋóʔ MMN – AKL bujóʔ
4. swidden h : - : - : - : - : h PAN *qumah ITB – AYA – TS – MMN – AKL umáh	5. father x : Ø : Ø : ʔ : - : h PAN *amax ITB ama AYA ama TS amáʔ MMN – AKL amáh	6. stone x : Ø : - : Ø : Ø : h PAN *batux ITB vato AYA – TS bató MMN bató AKL batóh
7. mother x : Ø : - : - : - : h PAN *inax ITB ina AYA – TS – MMN – AKL ináh	8. excavate x : Ø : Ø : Ø : - : h PAN *kalix ITB kali AYA kali TS maŋali MMN – AKL kálìh	9. head louse x : Ø : h : Ø : - : h PAN *kuCux ITB koto AYA kotoh TS koto MMN – AKL kútoh
10. how many? x : h : - : - : h : h PAN *pijax ITB pirah AYA – TS – MMN piráh AKL piláh	11. ash x : Ø : Ø : h : - : h PAN *qabux ITB avo AYA abo TS abóh MMN – AKL abóh	12. head x : Ø : Ø : Ø : - : h PAN *qulux ITB oxo AYA ō TS olo MMN – AKL úeoh
13. three x : Ø : Ø : - : - : h PAN *telux ITB atlo AYA tatlo TS – MMN – AKL tá-tloh	14. fathom - : h : h : - : h : h PAN – ITB hadpa AYA depah TS – MMN depáh AKL dupáh	15. thorn Ø : - : h : Ø : - : h PAN *duRi ITB – AYA diwih dowi MMN – AKL dúgih
16. rice husk - : - : h : - : h : h PAN – ITB – AYA apah TS – MMN upáh AKL opáh	17. sesame - : - : h : h : h : h PAN – ITB – AYA leŋah TS leŋah MMN liŋah AKL euŋáh	18. cut thatch - : h : - : - : - : h PAN – ITB sasah AYA – TS – MMN – AKL sasáh

TABLE 7. FINAL *-h*: CORRESPONDENCE SETS (CONTINUED)

19. jaw, jawbone - : - : h : h : - : -	20. bunch, cluster - : - : h : h : - : h	21. shark ? : - : h : h : - : -
PAN -	PAN -	PAN *qiSu (?)
ITB -	ITB -	ITB -
AYA paŋah	AYA poŋoh	AYA iyoh
TS paŋah	TS poŋoh	TS iyóh
MMN -	MMN -	MMN -
AKL -	AKL puŋóh	AKL -
22. elbow x : h : Ø : Ø : h : h	23. fish poison ? : Ø : h : Ø : - : h	24. node, joint - : Ø : h : h : - : h
PAN *sikux	PAN *tuba (?)	PAN -
ITB sichoh	ITB tuva	ITB voko
AYA hiko	AYA tobah	AYA bokoh
TS hiko	TS toba	TS bokoh
MMN sikoh	MMN -	MMN ----
AKL sikoh	AKL túbah	AKL bukóh
25. ear wax ? : Ø : h : h : - : h	26. yam - : Ø : - : - : h : h	27. gall, bile ? : Ø : Ø : h : - : h
PAN Culi (?)	PAN -	PAN *qapeju (?)
ITB tilo (<M)	ITB ovi	ITB apdo
AYA tolih	AYA -	AYA aplo
TS tolih	TS -	TS aploh
MMN -	MMN ubih	MMN -
AKL tulih	AKL úbih	AKL ápdoh
28. rice straw ? : - : Ø : Ø : - : h	29. narra tree - : - : h : - : - : h	30. fork of a branch - : Ø : Ø : Ø : - : h
PAN *zaRami (?)	PAN -	PAN -
ITB -	ITB -	ITB saŋa
AYA dayami	AYA nayah	AYA haŋa
TS dayami	TS -	TS haŋá
MMN -	MMN -	MMN -
AKL dagámih	AKL náгах	AKL saŋáh
31. crumbs of rice - : - : Ø : - : - : h	32. where/when? Ø : h : Ø : Ø : - : -	33. kiss - : h : Ø : - : - : -
PAN -	PAN *-nu	PAN -
ITB -	ITB di-noh 'where?'	ITB omah
AYA momo	AYA no 'when?'	AYA oma
TS -	TS maka-no	TS -
MMN -	MMN -	MMN -
AKL múhmuh	AKL -	AKL -

from a gapless set of the form $x : \emptyset : h : \emptyset : \emptyset : h$. However, sets 6 and 7 are also compatible with C5, which potentially derives from a gapless set of the form $x : \emptyset : \emptyset : \emptyset : \emptyset : h$, but sets 9, 23, and 29 clash with this, since AYA has /h/ in each of these forms. To make this clearer, (16) shows the idealized gapless sets for C1–C13. Only C3, C8, C12, and C13 cannot be generalized in this way, the first because of the unexplained appearance of glottal stop in both Tina Sambal and Aklanon where other languages have zero or /h/, making this correspondence unique; the second because no Mamanwa data are available for any of the three cognate sets used to justify this correspondence; the third because of the absence of unambiguous evidence for a final aspirate in Formosan languages; and the

TABLE 8. A COMPREHENSIVE AND INTEGRATED SCHEMA FOR FORMOSAN AND PHILIPPINE *-h CORRESPONDENCES

	PAN	ITB	AYA	TS	MMN	AKL	set
C1:	h	Ø	Ø	Ø	—	h	: 1
	h	—	—	—	—	h	: 4
	—	Ø	—	—	h	h	: 26
	?	—	Ø	Ø	—	h	: 28
	—	Ø	Ø	Ø	—	h	: 30
	—	—	Ø	—	—	h	: 31
C2:	h	h	h	Ø	—	h	: 2
	h	—	—	—	—	h	: 4
	—	h	h	—	h	h	: 14
	—	—	h	—	h	h	: 16
	—	h	—	—	—	h	: 18
	—	—	h	—	—	h	: 29
C3:	h	—	Ø	?	—	?	: 3
	—	h	Ø	—	—	—	: 33
C4:	x	Ø	Ø	?	—	h	: 5
	x	Ø	—	—	—	h	: 7
	x	Ø	Ø	—	—	h	: 13
	—	Ø	—	—	h	h	: 26
	—	—	Ø	—	—	h	: 31
C5:	x	Ø	—	Ø	Ø	h	: 6
	x	Ø	—	—	—	h	: 7
	x	Ø	Ø	Ø	—	h	: 8,12
	—	—	Ø	—	—	h	: 31
C6:	x	Ø	—	Ø	Ø	h	: 6
	x	Ø	—	—	—	h	: 7
	x	Ø	h	Ø	—	h	: 9
	?	Ø	h	Ø	—	h	: 23
	—	—	h	—	—	h	: 29
C7:	x	h	—	—	h	h	: 10
	—	h	h	—	h	h	: 14
	—	—	h	—	h	h	: 16
	—	—	h	h	h	h	: 17
	—	h	—	—	—	h	: 18
	—	—	h	h	—	—	: 19
	—	—	h	h	—	h	: 20
	?	—	h	h	—	—	: 21
C8:	—	—	h	—	—	h	: 29
	x	Ø	Ø	h	—	h	: 11
	x	Ø	Ø	—	—	h	: 13
	?	Ø	Ø	h	—	h	: 27
C9:	—	—	Ø	—	—	h	: 31
	—	h	h	—	h	h	: 14
	Ø	—	h	Ø	—	h	: 15
	—	—	h	—	h	h	: 16
	—	h	—	—	—	h	: 18
C10:	—	—	h	—	—	h	: 29
	x	h	—	—	h	h	: 10
	—	h	—	—	h	—	: 18
	x	h	Ø	Ø	h	h	: 22

TABLE 8. A COMPREHENSIVE AND INTEGRATED SCHEMA FOR FORMOSAN AND PHILIPPINE *-h CORRESPONDENCES (CONTINUED)

	PAN	ITB	AYA	TS	MMN	AKL	set
	?	—	Ø	Ø	—	h	: 28
	—	—	Ø	—	—	h	: 31
	—	h	Ø	—	—	—	: 33
C11:	—	—	h	—	h	h	: 16
	—	—	h	h	—	—	: 19
	—	—	h	h	—	h	: 20
	?	—	h	h	—	—	: 21
	—	Ø	h	h	—	h	: 24
	?	Ø	h	h	—	h	: 25
	—	Ø	—	—	h	h	: 26
	—	—	h	—	—	h	: 29
C12:	—	Ø	—	—	h	h	: 26
	Ø	Ø	Ø	h	—	h	: 27
	—	—	Ø	—	—	h	: 31
C13:	Ø	h	Ø	Ø	—	—	: 32
	—	h	Ø	—	—	—	: 33

fourth because of gaps in both Mamanwa and Aklanon. Finally, only C10 is represented by a gapless comparison (set 22: ‘elbow’).

(16)	PAN	ITB	AYA	TS	MMN	AKL
C1	h	Ø	Ø	Ø	h	h
C2	h	h	h	Ø	h	h
C3	h	—	Ø	?	—	?
C4	x	Ø	Ø	?	h	h
C5	x	Ø	Ø	Ø	Ø	h
C6	x	Ø	h	Ø	Ø	h
C7	x	h	h	h	h	h
C8	x	Ø	Ø	h	—	h
C9	Ø	h	h	Ø	h	h
C10	x	h	Ø	Ø	h	h
C11	?	Ø	h	h	h	h
C12	Ø	Ø	Ø	h	h	h
C13	Ø	h	Ø	Ø	—	—

All attempts to reduce the number of distinct correspondence classes by reassignment of cognate sets has proven fruitless, and we are left with an apparently irreducible set of thirteen. Although C3 is justified by only one cognate set (set 3: ‘skull’), the number of cognate sets that represent other correspondence classes will vary with the assignment of ambiguous sets. To illustrate, set 26 (‘yam’: — : Ø : — : — : h : h) is compatible with C1, C4, C8, C11, and C12. If set 26 is arbitrarily assigned to C1, that correspondence class will gain support, and the other four correspondence classes will lose support; and so on with any arbitrary reassignment. For this reason, based on the data currently available, it is impossible to say with certainty how much support a given correspondence class has. However, it can be said with certainty that 33 cognate sets are represented by C1–C13, for an average of about 2.5 forms each, although the assumption of an equal distribution

across correspondence classes is arbitrary and perhaps untrue. Add to these the more numerous instances of PAN *-S > -h, and it is hard to escape the conclusion that Philippine languages have at least 12 distinct sound correspondences (all but C3) that involve -h in some languages in correspondence with zero in most others.

All ambiguities in the final consonant of the 33 items in (16) are indicated explicitly in table 9.

TABLE 9. AMBIGUOUS ASSIGNMENTS OF PROTOFORMS

	PAN	ITB	AYA	TS	MMN	AKL	
C1,2:	h	—	—	—	—	—	: 4
C5,6:	x	Ø	—	Ø	Ø	h	: 6
C4,5,6:	x	Ø	—	—	—	h	: 7
C7,10:	x	h	—	—	h	h	: 10
C5,8:	x	—	—	Ø	—	—	: 13
C2,7,9:	—	h	h	—	h	h	: 14
C2,7,9:	—	—	h	—	h	h	: 16
C7,11:	—	—	h	h	h	h	: 17
	—	—	h	h	—	—	: 19
	—	—	h	h	—	h	: 20
	?	—	h	h	—	—	: 21
C2,7,9,10:	—	h	—	—	h	—	: 18
C1,4,11,12:	—	Ø	—	—	h	h	: 26
C8,12:	?	Ø	Ø	h	—	h	: 27
C1,5,10:	?	—	Ø	Ø	—	h	: 28
C2,6,7,9,11:	—	—	h	—	—	h	: 29
C1,5:	—	Ø	Ø	Ø	—	h	: 30
C1,4,5,8,10,12:	—	—	Ø	—	—	h	: 31
C3,10,13:	—	h	Ø	—	—	—	: 33

It is hard to say how much longer this list might become if more witnesses for the mystery aspirates were discovered, but the data considered so far are already sufficient to show that there is a major comparative puzzle in the historical phonology of Philippine languages waiting to be solved.

7. SEEKING EXPLANATIONS. What historical events could possibly have produced the range of sound correspondences shown in Table 8? Two possibilities come to mind:

1. The mystery aspirates were added in the separate history of several languages in the Philippines.
2. The mystery aspirates are retentions from PPH, PMP, and PAN.

Possibility 1 is an *a priori* violation of the Regularity Hypothesis. To circumvent this problem, we might argue that in many Philippine languages -V alternates with -Vh before a vowel-initial suffix. This may have been an old pattern that led to analogical wrong-division in particular cases, producing word-final -h where before there had been only intervocalic aspirates as a regular phonological feature of suffixation. However, this interpretation cannot easily explain agreements between widely separated languages. For example, if the word for ‘elbow’ had been *siku in a language ancestral to the modern

languages of the Philippines as is suggested by Ayta Abellen, Tina Sambal *hiko*, there is no obvious explanation why *-h* would have been added independently in Itbayaten *sichoh*, and Mamanwa, Aklanon *sikuh*, but not in words such as Itbayaten *avo* ‘ashes’, when Tina Sambal and Aklanon both show *-h* in the cognate form.

To avoid the unpredictability of *-h* in this approach, we might, therefore, posit a final consonant in each of the protoforms ancestral to the comparisons in table 7. However, the problem then becomes “What do we reconstruct?” The normal rule of thumb following the comparative method is to posit a distinct protophoneme for each correspondence class that is not in complementary distribution with another. In this case, we would be forced to posit thirteen aspirates $*h_1 - *h_{13}$ word-finally (*-h* from PAN $*-S$, and one of each $*h$ for each correspondence class except C3 in table 8). The typological implausibility of this approach is no less objectionable than the unpredictability of approach 1. We are left, then, at an impasse, which is why these aspirates in Philippine languages are “mysterious”: they fail to agree with theoretical expectations, and so leave the analyst with no good explanatory alternative.

As a compromise, we might assume that PMP and PPH had two word-final aspirates, $*-h$ and $*-x$, parallel to the reconstruction of PAN $*-H1$ and $*-H2$ in Tsuchida (1976). The first of these reflected PAN $*-S$, and is consistent across all languages in all comparisons. The second reflected PAN $*-h$ and $*-x$, and is far less consistent across languages. However, even here we encounter what appear to be insoluble problems. Laryngeals may be inherently unstable, and subject to sporadic loss, but why would this be true of PMP $*-x$ from PAN $*-h$ and $*-x$, but not of PMP $*-h$ from PAN $*-S$? Or should we assume that PAN $*S$ became PMP $*x$, and, hence, was more resistant to lenition than PMP $*-h$? And what of the subsequent history of PMP $*-h$: why was it retained in some forms in some languages but lost in others in a pattern that resists any simple analysis?

In short, there seems to be no plausible basis for treating the mystery aspirates as either innovations or retentions, leaving us with a methodological dilemma. Treating them as innovations requires the recognition of widespread sporadic change that inexplicably targeted the same morphemes in widely separated languages, while treating them as retentions would appear to require the reconstruction of as many as thirteen types of $*h$, violating typological plausibility. It is true that many of the sound correspondences summarized in (1) are represented by single forms, but as already noted, these numbers in some cases are arbitrary, given the multiple possibilities of compatibility for some correspondences with critical gaps in attestation. Moreover, comparisons such as those for ‘ember’ or ‘elbow’ cannot easily be explained as innovations, since they agree in widely separated languages in having an unexpected final aspirate.

This raises an important issue of a more general kind. In any branch of science, research does not end in a collection of data. Rather it ends in a set of generalizations about those data—a set of principles that show why the observable data behave the way they do. Given this imperative, it is clear that theory and data may appear irreconcilable, as in the present case. When that happens, there are no generally accepted guidelines that show how the data should be reported. Repeated reactions from referees to manuscripts of my own have shown that there is a substantial faction among linguists that opposes the publication of anomalous data without an accompanying theory, as unexplained data

establish no generalization, and, hence, fail to meet the signal requirement of all sciences, that the analysis of data should lead to a formulation of general principles. Unfortunately, this valid concern is easily twisted into an implicit claim that when theory and data conflict, theory must predominate, even to the point of excluding data from consideration. I believe this attitude is seriously mistaken.

In perhaps the earliest generalization made in linguistics, Jacob Grimm in 1822 formulated the sound correspondence that he called the first Germanic consonant shift, a change that subsequently came to be known as “Grimm’s Law.” What Grimm found was a pattern of correspondence between letters of the alphabet (*Buchstaben*) representing the stops of other Indo-European languages and their corresponding alphabetic symbols in Germanic, a pattern that was repeated so often that it clearly required a hypothesis of a common origin of the Indo-European languages, and of the unity of the Germanic subgroup.

However, not every word fit this pattern, and rather than sweep nonconforming cases under the proverbial rug, Grimm listed them conscientiously so that four decades later, Hermann Grassmann, and a decade after him, Karl Verner were able to show that these “exceptions” to Grimm’s laws are in fact subregularities. It is possible, and, in fact, likely, that the subregularities formulated as Grassmann’s Law and Verner’s Law would have been discovered even without Grimm’s list of exceptions, but his openness in listing nonconforming cases probably accelerated the discoveries that ultimately made Grimm’s Law more complete than it was in its original formulation.

In my view, the same principle that governed this classic case should govern others: when theory and data disagree, theory does not trump data, leading to its exclusion from consideration until some undefined future time. Rather, irregularities should be highlighted, not treated as an embarrassment to science, since cases that challenge existing assumptions are the ones most likely to lead to advances in understanding. The anomalous “mystery aspirates” that are reported here are offered with no current explanation, in the hope that in time an explanation will be found, and it is in this spirit that I have brought them to the attention of other linguists who may in time provide a better understanding for them than is currently available.

REFERENCES

- Adelaar, Alexander. 2011. *Siraya: Retrieving the phonology, grammar and lexicon of a dormant Formosan language*. Trends in Linguistics Documentation 30. Berlin: de Gruyter Mouton.
- Blust, Robert. 1969. Some new Proto-Austronesian trisyllables. *Oceanic Linguistics* 8:85–104.
- . 1981. The Soboyo reflexes of Proto-Austronesian *S. In *Historical Linguistics in Indonesia*, ed. by Robert A. Blust, 21–30. *NUSA: Linguistic studies in Indonesian and languages in Indonesia*, vol. 10.
- . 1991. The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30:73–129.
- . 1993. *S metathesis and the Formosan/Malayo-Polynesian language boundary. In *Language—a doorway between human cultures: Tributes to Dr. Otto Chr. Dahl on his ninetieth birthday*, ed. by Øyvind Dahl, 178–83. Oslo: Novus.

- . 2013 [2009]. *The Austronesian languages*. Rev. ed. Asia-Pacific Linguistics Open Access Monographs. Canberra: Research School of Pacific and Asian Studies, The Australian National University.
- . 2017. Regular metathesis in Batanic (Northern Philippines)? *Oceanic Linguistics* 56:491–504.
- Blust, Robert, and Stephen Trussel. Ongoing. *Austronesian comparative dictionary*. Open access online resource available at www.trussel2.com/acd.
- Brandstetter, Renward. 1916. *An introduction to Indonesian linguistics: Being four essays by Renward Brandstetter*. Trans. by C. O. Blagden. London: Royal Asiatic Society Monographs XV.
- Dahl, Otto Chr. 1976 [1973]. *Proto-Austronesian*. 2nd, rev. edition. Scandinavian Institute of Asian Studies Monograph Series, No. 15. London: Curzon Press.
- Dempwolff, Otto. 1934–38. *Vergleichende Lautlehre des austronesischen Wortschatzes*. *Zeitschrift für Eingeborenen-Sprachen*, Supplement 1. *Induktiver Aufbau einer indonesischen Ursprache* (1934); Supplement 2. *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen* (1937); Supplement 3. *Austronesisches Wörterverzeichnis* (1938). Berlin: Reimer.
- Dyen, Isidore. 1951. Proto-Malayo-Polynesian *Z. *Language* 27:534–40.
- . 1953a. *The Proto-Malayo-Polynesian laryngeals*. William Dwight Whitney Linguistic Series. Baltimore: Linguistic Society of America.
- . 1953b. Dempwolff's *R. *Language* 29:359–66.
- . 1962. Some new Proto-Malayopolynesian initial phonemes. *Journal of the American Oriental Society* 82:214–15.
- . 1965. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285–305.
- Elgincolin, Sotero B., Hella E. Goschnick, and Priscilla R. Elgincolin. 1988. *Diksiyonaryon English–Sambali Tinã–Pilipino*. Manila: Summer Institute of Linguistics, Philippine Branch.
- Greenhill, Simon J., Robert Blust, and Russell Gray. 2003–18. *The Austronesian basic vocabulary database*. Online resource: <https://abvd.shh.mpg.de/austronesian/>.
- Himes, Ronald S. 2012. The Central Luzon group of languages. *Oceanic Linguistics* 51:490–537.
- Li, Paul Jen-kuei. 2004a. Reconstruction of Proto-Atayalic phonology. In *Selected papers on Formosan languages*, ed. by Paul Jen-kuei Li, 625–92. Language and Linguistics Monograph Series No. C3. Taipei: Institute of Linguistics, Academia Sinica.
- . 2004b. A comparative study of Bunun dialects. In *Selected papers on Formosan languages*, ed. by Paul Jen-kuei Li, 743–66. Language and Linguistics Monograph Series No. C3. Taipei: Institute of Linguistics, Academia Sinica.
- Lobel, Jason William. 2013. Philippine and North Bornean languages: Issues in description, subgrouping, and reconstruction. PhD diss., University of Hawai'i.
- McFarland, Curtis D. 1977. *Northern Philippine linguistic geography*. Studies of the Languages and Cultures of Asia and Oceania monograph series, No. 9. Tokyo: Institute for the Study of Languages and Cultures of Asia and Oceania.
- Prentice, D. J. 1974. Yet another PAN phoneme? *Oceanic Linguistics* 13:33–75.
- Reid, Lawrence A. 1982. The demise of Proto-Philippines. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, ed. by Amran Halim, Lois Carrington, and S. A. Wurm, 201–16. Canberra: Pacific Linguistics.
- . 1987. The early switch hypothesis: Linguistic evidence for contact between Negritos and Austronesians. *Man and culture in Oceania* 3:41–59.

- Reid, Lawrence A., ed. 1971. *Philippine minor languages: Word lists and phonologies*. Oceanic Linguistics Special Publication No. 8. Honolulu: University of Hawai'i Press.
- Ross, Malcolm. 1992. The sound of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics* 31:23–64.
- . 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary: An introduction to Austronesian studies*, Part 1, Fascicle 1, ed. by Darrell T. Tryon, 45–120. Berlin: Mouton de Gruyter.
- . 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2):1–24.
- Smith, Alexander D. 2017. The Western Malayo-Polynesian problem. *Oceanic Linguistics* 56:435–90.
- Stone, Roger. 2007. *Ayta Abellen dictionary and texts Introduction – Work in Progress*. Online resource available through SIL, Philippines. http://www.philippines.sil.org/resources/works_in_progress/abp.
- . 2008. The Sambalic languages of central Luzon. *Studies in Philippine languages and cultures* 19:158–83.
- Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. SLCAA Monograph Series 5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Tsuchida, Shigeru, Yukihiro Yamada, and Tsunekazu Moriguchi. 1987. *Lists of selected words of Batanic languages*. Tokyo: Department of Linguistics, Faculty of Letters, University of Tokyo.
- Yamada, Yukihiro. 1976. A preliminary dictionary of Itbayaten. Typescript.
- Zorc, R. David. 1969. *A study of the Aklanon dialect*, vol. 2: *Dictionary*. Kalibo, Aklan: Public Domain.
- . 1982. Where, o where, have the laryngeals gone? Austronesian laryngeals reexamined. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, ed. by Amran Halim, Lois Carrington, and S. A. Wurm, 111–44. Canberra: Pacific Linguistics.
- . 1986. The genetic relationships of Philippine languages. In *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, ed. by Paul Geraghty, Lois Carrington, and S.A. Wurm, 147–73. Canberra: Pacific Linguistics.
- . 1996. The reconstruction and status of Austronesian glottal stop—Chimera or chameleon. In *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, ed. by Bernd Nothofer, 41–72. Abera Network, Asia-Pacific 3. Hamburg: Abera Verlag Meyer & Co.