Introduction

I. GOALS

The goal of this work is to create a comparative dictionary that fully supplants Dempwolff (1938) as the primary source of historical data on the entire Austronesian language family.

Scope. Although valuable contributions were made by van der Tuuk, Brandstetter and others during the late nineteenth and early twentieth centuries, the foundation-laying work in Austronesian comparative linguistics was provided by Otto Dempwolff (1934-1938), who reconstructed a complete (but inadequate) sound system, and over 2,200 lexical reconstructions in the form of an Austronesian comparative dictionary. Dempwolff's dictionary served as a fait accompli for over three decades before a renewed interest in Austronesian lexical comparison was expressed in Blust (1970). During this time Dempwolff (1938) was commonly treated as if it had identified all or nearly all possible comparisons and provided support for them. However, it is doubtful that Dempwolff would have shared this view, since his primary purpose was to provide a sound system that was adequate for the entire language family, now believed to contain over 1,200 languages (Ethnologue), and the only practical way to achieve this goal was by restricting his comparison to a small subset of languages that met the demanding requirement of representing all phonemic contrasts that were present in the earliest ancesteral stage. He did this by selecting just three languages, Tagalog, Toba Batak and Javanese, to reconstruct the phonology of 'Uraustronesisch' (Dempwolff 1934), and then testing the adequacy of this reconstruction with eight other languages scattered from Madagascar to Samoa (Dempwolff 1937). While this restriction enabled him to provide a phonological reconstruction within a reasonably restricted compass, it also excluded many possible lexical comparisons that were only inferrable from evidence in other languages, either by themselves, or in combination with one of the 11 languages used to posit and test the reconstructed phonology.

Following Blust (1970), several article-length publications (one of them 181 pages) over the next 19 years served to show that much more could be done with the newer descriptive sources that had become available since Dempwolff's death in 1938. These also showed that Dempwolff's reconstruction could not be called 'Proto-Austronesian', since it excluded the critically important aboriginal languages of Taiwan, which most researchers now assign to more than one primary branch of the language family. Moreover, it began to become clear that many -- perhaps as many as a third -- of Dempwolff's reconstructions were either late innovations in western Indonesia-Malaysia, or loan distributions from Malay.

In 1990 a three-year grant was obtained through the National Science Foundation of the United

States, and a full-scale attempt was made at beginning a comprehensive new comparative dictionary of the Austronesian languages. Although other commitments made progress slow at first, a two-year no cost extension to 1995 enabled the writer to compile fairly exhaustive comparative materials for etyma beginning with a vowel, *b, *h, *q, *s, *S, or *w, with preliminary computer assistance from David Stampe. However, other letter-groups were only partially explored at the conclusion of the grant, and remained in this inchoate condition until a chance meeting with Steve Trussel in February, 2010 set the stage for reviving work on the dictionary. Earlier work by Stampe, and subsequently by Richard Nivens was transformed by Steve into the present ongoing online product.

To give some idea of the structure of Dempwolff's classic dictionary, and how it maps onto the present effort, a breakdown of total reconstructions by letter group is given below. By matching the Dempwolff number to the ACD number in sections that have been more-or-less completed, as *b, *h, *q, *S, *w, or any of the vowels, it becomes possible to predict what a similarly intensive investigation of other groups is likely to yield once that work has been pressed to the limit permitted by the sources. For some letter groups a one-to-one matching is impossible, since Dempwolff's treatment of the 'laryngeals' was flawed, and was corrected by Dyen (1953). In other cases Dempwolff made distictions that are no longer generally accepted, at least in onset position, as with *D and *T; these are now combined with the lower-case equivalents. It should also be noted that while Dempwolff reconstructed at only one level (Uraustronesisch), many of his reconstructions are confined to languages of western Indonesia, where Malay loanwords are very common and sometimes go undetected (Wolff 2010). By contrast, the ACD contains reconstructions at nine distinct levels, each of which is explicitly marked:

1. **Proto-Austronesian**(PAN), requiring data from at least one Formosan language; reconstructions based exclusively on Formosan comparisons are treated in a separate file as a precaution against the possibility that some of these are post-PAN innovations that spread by borrowing at a stage that is too early to leave detectable traces,
2. **Proto-Malayo-Polynesian** (PMP), requiring data from at least one Western Malayo-Polynesian witness and one Central-Eastern Malayo-Polynesian witness,
3. **Proto-Wesern Malayo-Polynesian** (PWMP), generally accepted only if cognate sets include languages in both the Philippines and the Greater Sunda islands or Sulawesi south of the northernmost peninsula; this node may be equivalent to PMP, but the precise way to formulate this relationship remains unclear,
4. **Proto-Philippines** (PPH), based on cognate sets found in both the Bashiic and or Cordilleran languages, and the Greater Central Philippines languages or other members of the Philippine subgroup further south,
5. **Proto-Central-Eastern Malayo-Polynesian** (PCEMP); based on cognate sets found in both Central Malayo-Polynesian and Eastern Malayo-Polynesian languages,
6. Proto-Central Malayo-Polynesian (PCMP), based on cognate sets found in both the Lesser Sundas and the central Moluccas,

7. **Proto-Eastern Malayo-Polynesian** (PEMP), based on cognate sets found in both the South Halmahera-West New Guinea and Oceanic languages,
8. **Proto-South Halmahera-West New Guinea** (PSHWNG), based on cognate sets found in both South Halmahera and West New Guinea languages,
9. **Proto-Oceanic** (POC), following the general requirements set out in Ross, Pawley, and Osmond (1998, 2003, 2008, 2011, 2016).

In addition, reconstructions include affixed forms as subentries to bases, providing a wealth of information about comparative morphology, and annotation to many entries. For all of these reasons, it is difficult to provide an exact matching of categories in Dempwolff (1938) and the ACD.

Keeping these caveats in mind, the mapping has the following form, where Dempwolff's orthography has been converted to that of Dyen and all subsequent researchers (cf. Blust 2013, Ch. 8):

Initial segment	Dempwolff	ACD	ACD/ Dempwolff proportion
*a	95	248	2.60
*b	280	1,033	3.69
*c	48	48	1.00
*d	68	192	2.82
*D	44	----	
*e	18	124	6.90
*g	106	143	1.35
*h	94	159	1.69

*i	60	207	3.45
*k	214	400	1.87
*l	175	372	2.12
*m	44	87	2.00
*n	17	62	3.65
*ñ	8	17	2.12
*ŋ	6	75	12.50
*p	208	430	2.07
*r	70	98	1.40
*R	24	90	3.75
*s	192	719	3.69
*t	286	384	1.34
*T	16	----	
*u	66	141	2.14
*w	11	71	6.45
*y	2	----	
*z	71	50	.70

*C	---	59	
*j	---	2	
*N	---	20	
*o	---	13	
*q	---	405	
*S	---	93	

Given a thoroughly explored letter group like *b-, then, where the ACD contains about 3.69 times as many base forms as Dempwolff (1938), we would expect a letter group like *k- to contain roughly the same multiple of Dempwolff's 214, hence around 790 entries. Of course, it could be a lower multiple, like 2.60 (*a), or a higher one, like 6.45 (*w), but for the most part I would compare obstruents with obstruents, sonorant consonants with sonorant consonants, and vowels with vowels in making such projections.

Interdisciplinary value. Since Dempwolff's work was published in the 1930s the growth of science has been explosive. This has led to two somewhat contradictory results. First, the potential value of work in historical linguistics to sister disciplines such as prehistoric archaeology or social/cultural anthropology, or of these disciplines to linguistics has increased immensely in level of detail and sophistication, so that interdisciplinary work can in principle take us much further than was possible in Dempwolff's day. Second, despite the rich materials that can now be compared with one another across disciplinary boundaries, most work today is so specialized that it is difficult to fully appreciate arguments in sister disciplines, and this has led in some ways to interdisciplinary barriers created by training, and reinforced by publication in journals that are seldom consulted on a regular basis by scholars outside the discipline they represent.

One of the goals of this project is to make a large body of material relating to material and non-material culture available to scholars in other academic fields to use as they see fit. If used by non-linguists it can thus serve as an interdisciplinary resource for issues regarding the prehistory of insular Southeast Asia and the Pacific. At the same time it will provide raw data for linguists outside the Austronesian field to test claims about phonological, morphological and lexical change.

II. SOURCES. Data for most languages is taken from a single source, but in some cases secondary sources have provided material not otherwise available. The primary sources for all languages cited in the ACD are given in parentheses in the 'languages' link. If material is drawn from some other source I have tried to note this in the citation of evidence itself.

III. STANDARDS OF EVIDENCE

A. I recognize five categories of comparisons in the ACD:

1. ***Canonical comparisons***: those with regular sound correspondences and close semantics. If there are additional forms that are strikingly similar but irregular, or that show strong semantic divergence, these are are added in a note. Every attempt is made to keep the comparison proper free from problems.
2. ***Near comparisons***: forms that are strikingly similar but irregular, and which cannot be included in a note to an established reconstruction. Stated differently, these are forms that appear to be historically related, but do not yet permit a reconstruction.
3. ***Loans***: forms that are related, but not directly inherited. This includes both *external loans* (words from non-Austronesian sources), and *internal loans* (Austronesian lexical items borrowed from one Austronesian language into another).
4. ***Shortfall***: canonical comparsions (like category 1), but innovated in a proto-language lower than any recognized here (in general these are simply ignored unless there is some question that they might be attributable to a higher proto-language once more information is collected).
5. ***Noise (look-alikes)***: Given the number of languages being compared and the number of forms in many of the sources, forms that resemble one another in shape and meaning by chance will not be uncommon, and the decision as to whether a comparison that appears good is a product of chance must be based on criteria such as
    a. how general the semantic category of the form is (e.g. phonologically corresponding forms meaning 'cut' are less diagnostic of relationship than phonologically corresponding forms for particular types of cutting),
    b. how richly attested the form is (if it is found in just two witnesses the likelihood that it is a product of chance is greatly increased),
    c. there is already a well-established reconstruction for the same meaning.

THE WORK OF OTHER SCHOLARS

Needless to say the ACD has benefited substantially from the work of other scholars, most notably that of Dempwolff (1938). In addition to thoroughly searching all available materials from scratch, I have re-evaluated the data in these works before making a decision on whether to include them or not. The result has been the incorporation of most material from other sources, but the rejection of a fairly substantial number of comparisons. The following sample provides an idea of the problems that anyone sifting through the existing literature must

confront, and find solutions to.

1. Dempwolff (1938)

Problems with Dempwolff. Despite the brilliant pioneering effort of Dempwolff, which clearly placed Austronesian comparative linguistics on a firmer methodological footing than had previously been the case, there are problems with his work. The most important phonological problems were pointed out by Ogawa and Asai (1935) and later scholars who incorporated essential data from the aboriginal languages of Taiwan, in particular the distinctions between *t/C and *n/N, and the phonetic character of *S, and by Dyen (1953) in his important revision of Dempwolff's treatment of the 'laryngeals'.

However, there are also many problems with Dempwolff etymologies. About one third of his roughly 2,200 forms must be dismissed as not assignable even to PWMP for any of three reasons (or combinations of them):

1. the comparison is confined to languages now known to belong to a low-level subgroup, as with comparisons restricted to Ngaju Dayak and Malagasy (e.g. *sampaŋ 'side path, fork in the road'),
2. the comparison is found only in Malay and other languages of western Indonesia that have borrowed heavily from Malay for centuries, in particular Javanese; this also includes some comparisons that extend to Tagalog and occasionally other Philippine languages that show clear evidence of borrowing from Malay; although some cases are quite clear, it is often difficult to determine whether a form in a Philippine language is native or a loan, and as a result some reconstructions that are accepted here may be products of undetected borrowing,
3. a surprising number of comparisons must be considered false etymolgies. Many of these apparently resulted from Dempwolff's use of a select subset of eleven languages, and his need to compile a comparative dictionary based only on them. In many cases where the sound correspondences are recurrent there is a forced attempt to bring together forms that are semantically very dissimilar, and sometimes also phonologically irregular. Part of the contribution of the ACD, then, is to eliminate hundreds of poorly supported etymologies in Dempwolff (1938) while simultaneously adding many new ones arrived at by stricter criteria of subgrouping and control for borrowing.

To summarize, although most Dempwolff comparisons fall into category 1, 'canonical comparisons' a number of others fall into categories 3, 4 or 5 and have either been assigned to the sub-files on Loans or Noise, or have been excluded from the ACD.

Since the goal of the ACD is to compile as many etymologies as possible on nine different levels (PAN, PMP, PWMP, PPH, PCEMP, PCMP, PEMP, PSHWNG, POC) I have also had to

weigh the merits of the PPH reconstructions in Zorc (1971), the mostly PWMP reconstructions in Mills (1975, 1981), and the POC reconstructions in Milke (1961, 1968), Grace (1969), and those volumes of The lexicon of Proto Oceanic (hereafter LPOC) that have appeared to date (Ross, Pawley and Osmond 1998, 2003, 2008, 2011, 2016) before determining whether or not to incorporate them into the ACD. In doing this I have encountered a variety of problems.

2. Zorc (1971)

Zorc (1971) is a list of 3,773 reconstructions without supporting evidence, many of which are taken from Dempwolff (1938) with the apparent expectation that if a reflex is found in at least one Philippine language this justifies a PPH etymon. However, since as many as one third of the reconstructions in Dempwolff (1938) are problematic, the same must be said for those PPH forms that are inspired by Dempwolff proto-forms. In other cases it has proven difficult to find the supporting evidence upon which a reconstruction is based, or if evidence is found, it is confined to subgroups of shallower time-depth than Proto-Philippines. This is an inherent problem in that the reconstructions in this preliminary compilation are not keyed to an explicit subgrouping hypothesis.

To summarize, although many of the reconstructions in Zorc (1971) fall into category 1, none are provided with supporting evidence. When a search is undertaken for such evidence those comparisons that survive are often not assignable to Proto-Philippines as that term is used here (requiring reflexes in at least Cordilleran and GCP languages, or in any of these plus Bashiic, or in either of these plus Sangiric or Minahasan). In other cases problems with Dempwolff's material are simply copied onto proposed PPH forms.

3. Mills (1975, 1981)

The first of these publications is a reconstruction of Proto-South Sulawesi, and provides a large number of proto-forms for that language. However, it also contains many brief references to languages of Indonesia outside the South Sulawesi group, and suggests various reconstructions that are labelled 'PIN' (Proto-Indonesian), without further explanation of what this label means. A number of these are valid and useful additions to the corpus of early Austronesian forms. However, as Mills himself notes, many others contain unexplained irregularities that raise questions about their cognation.

For this reason it has been necessary to independently check each of the etymologies in this large work before making a decision about whether or not to include them in the ACD. Mills (1981) is in many ways a refinement of these preliminary remarks that attempts to present just the most attractive of the comparisons mentioned in the earlier study. It includes 205 reconstructions, of which 33 are labeled 'Proto-South Sulawesi', 120 are labeled 'Proto-Indonesian', and 52 are labeled 'Proto-Austronesian'. Since Formosan evidence is not taken

into account, none of the latter meet the criteria of the ACD for this status; rather, all are at best PMP. However, even the latter status is often open to question. To choose just two examples, both of which would be assigned to 'Noise' in accordance with the uniformly applied standards of evidence adopted in the ACD, consider 'PAN' *bumbum 'cloudy or indistinct color', based on forms in South Sulawesi languages describing the color of chickens, next to Fijian vuvu 'muddy, troubled, of water', or 'PAN' *pudDr)us 'pull or gather together', based on forms in South Sulawesi languages that refer variously to stripping leaves from a branch or picking fruits, next to Fijian buru-ka 'nip between finger and thumb', Tongan mulu 'to strip, or grasp and run the hand along with a stripping movement'.

To summarize, Mills (1975, 1981) has contributed a number of previously overlooked comparisons to the literature. However, none of these justify the label 'PAN', and many of his reconstructions on levels above PSS contain unexplained irregularities in sound correspondences, and sometimes wide semantic divergence; in terms of the standards of evidence adopted here these would mostly be assigned to categories 4 and 5.

4. Milke (1961, 1968)

Proposals for POC reconstructions have proven especially problematic. Although Milke (1961, 1968) clearly added valuable new material to the POC lexicon, a number of forms cited in these publications show irregularities in sound correspondences, and/or widely divergent meanings. As a result some of the reconstructions that he proposed have been abandoned as insufficiently supported. To cite just two examples in illustration, Milke (1968) posited POC *gasub 'to spit' and *maliŋ 'bitter' based on the following comparisons:

(1) POC *gasub '*to spit*'

Kiriwina	kapul/a (metathesis)	'to spit'
Tokunu	kuruv/i	'to spit'
Misima	kuruv/i	'to expectorate'
Yabem	kasôp	'to spit'
Tami	ma/gidjub	'to spit'
Bulu, Bola, Xama	kalup/e	'to spit'

Gedaged	yasu	‘to spit’, yusu/ni ‘to spit on’
Fijian	kasiv/i	‘to spit’

(2) POC *maliŋ ‘*bitter*’

Yabem	máli’	‘poisonous, indigestible’
Tami	maɲiŋ	‘sour, acetous, stifling’
Gedaged	meleŋa/liŋ	‘sour, acetous’
Ali	miyi	‘sour, bitter’
Nggela	mali	‘bitter, salt’
Gilbertese	māi	‘bitter’

With regard to comparison (1) the only languages that support the vowels of *gasub are Kiriwina, Bulu, Bola, Xama, and Gedaged. However, /l/ in Bulu, Bola and presumably Xama (otherwise completely unknown) cannot reflect *s (Ross 1988:267), and Gedaged should reflect POC *k- as *k-* or zero, not as *y-* (Ross 1988:170), thus eliminating any comparative basis for the inference adopted by Milke. Many of these forms also rely on the assumption that they contain a fossilized transitive marker, which varies from *-a* to *-i* to *-e* to *-ni*. In short, whatever historical relationship these forms might ultimately prove to have cannot be clearly extracted from the material given here, and the reconstruction is basically ‘invented’ wholecloth rather than used to account for recurrent sound correspondences.

With regard to comparison (2), while Nggela *mali* can regularly reflect *maliŋ, none of the other forms can. For Yabem the final glottal stop is unexplained (Ross 1988:136), for Tami the medial velar nasal (Ross 1988:168), for Gedaged much of the form, for Ali the medial glide (Ross 1988:127), and for Gilbertese the absence of a medial *n* < *l. What we have here, then, is little more than a collection of forms that share varying degrees of phonetic similarity, but no recurrent sound correspondences apart, perhaps, from *ma-. No subsequent work in comparative Oceanic linguistics has strengthened this comparison with new data, and there appears to be little alternative to dropping it.

To summarize, Milke (1961, 1968) contributed valuable material to the reconstruction of the POC lexicon. However, he adopted comparative criteria that fall short of the standards of the ACD, with the result that a number of of his reconstructions must be dismissed as unjustified by the evidence presented for them.

5. Ross, Pawley and Osmond (1998, 2003, 2008, 2011)**. The Lexicon of Proto Oceanic (LPOC)**

The material in LPOC is semantically far more precise, particularly in those volumes concerned with flora and fauna, but again sound correspondences are frequently allowed to depart from regularity or even recurrence, and it has proven difficult or impossible in many cases to see how a proposed reconstruction can be justified.

As already noted, about one-third of the etymologies in Dempwolff (1938) have been excluded from the ACD, either because the distribution of supporting forms is limited to Malay and languages of western Indonesia that have borrowed heavily from Malay, or because the proposed semantic connections are unconvincing. Unlike many Dempwolff etymologies, there are few subgrouping or semantic problems with the reconstructions in LPOC. There is a clear reason for this difference: whereas Dempwolff searched for forms to which he could subsequently assign meanings, the compilers of LPOC have adopted a meaning-based approach, using the semantic categories of contemporary Oceanic-speaking societies to generate a list of meanings for which reconstructions are sought at various levels within the Oceanic group. As a result of this difference (form-based vs. meaning-based), Dempwolff was acutely aware of the regularity of sound correspondences and conscientiously noted any deviation from the expected form, even while his treatment of meaning was often rather liberal.

The material in the LPOC often confronts the critical reader with the opposite problem: while the forms compared usually have a firm semantic connection, the treatment of sound correspondences is sometimes quite loose. Many comparisons are well-established, and there can be no question as to their validity. However, a surprising number are proposed on the basis of material that would permit nothing stronger than a 'Near comparison', or even an attribution to 'Noise' in the ACD. This problem only became apparent by making a serious attempt to incorporate as much material from the LPOC as possible, which led to the discovery that many comparisons which were initially attractive unravelled under closer scrutiny.

To make it clear that these remarks are not unfair, I have chosen 16 POC reconstructions that begin with *s in Ross, Pawley and Osmond (1998, 2003, 2008, 2011), and subjected them to the same standards of evidence that are used in the ACD generally, as seen below. To ensure that it is representative this sample includes both comparisons in LPOC that have been rejected, and those that have been incorporated into the ACD, either wholesale or with qualifications (PT= Papuan Tip, SES = Southeast Solomonic, PCP = Proto-Central Pacific, Fij = Fiji, Pn = Polynesian, MM = Meso-Melanesian, NCV = North-Central Vanuatu, TM = Temotu/Santa Cruz, Mic = Micronesia, Adm: Admiralties, SV = southern Vanuatu, NCal. = New Caledonia).

(1) POC *(sabi-)sabi '*shell disc used as earring*'

PT:	Tawala	sapi-sapi	‘earring’
SES:	Arosi	tabi-tabi	‘ear ornaments’

cf. also

PT:	Muyuw	lap	‘red discs in string of *veigun* (shell wealth)’
PT:	Molima	sapi-sapi	‘large red shell discs sewn to belt’

	PCP	***sau**	‘*ear pendant*’
Fij:	Rotuman	sau	‘earring or ear ornament’
Fij:	Bauan	sau (ni daliŋa)	‘earring’ (daliŋa ‘ear’)
Pn:	Tongan	sau	‘earring, ear ornament; nose ring, nose ornament’
Pn:	E. Uvean	hau	‘ear pendant’
Pn:	E. Futunan	sau	‘ear pendant’

DECISION: Accepted based on Tawala, Arosi, Muyuw and Molima, together with additional evidence from Tubetube. The CP forms cannot be derived from this reconstruction, as neither loss of *p nor *-i > *-u* is a recurrent change.

---

(2) POC *saja(q) ‘*prepare thatching materials or begin to thatch a roof*’

SES:	Gela	sada	‘tie the thatch in beginning a roof’

DECISION: Accepted based on ‘inverted reconstruction’ from PMP *sasah (Blust 1980). The final *q, however, is rejected, since it is not reflected by Tagalog, Aklanon, Cebuano,

or Iban (the final glottal stop in the last language does not reflect *q).

---

(3) POC *sake '*embark, ride on a canoe*'

DECISION: Accepted, as PMP *sakay is widely reflected from the Philippines to Fiji and Micronesia.

---

(4) POC *sakup '*kind of cooking banana*', mentioned in Osmond (1998:127), but documented in Ross (2008:278), where the suggested gloss is '*banana cultivar with long fruit*' (?)

PT:	Gumawana	yagowa	'a long non-sweet banana'
PT:	Taupota	hakova	'banana'
PT:	Taboro	daua	'k.o. banana: white flesh'
PT:	Motu	dau	'k.o. banana: very long'
MM:	Roviana	hakua	'banana'
MM:	Maringe	cau	'banana'
SES:	Kwara'ae	sa-sao	'k.o. banana with upright bunches and large fruit'
SES:	'Āre'āre	sao-sao	'k.o. wild banana'
NCV:	NE Ambae	haka	'banana'
NCV:	Larevat	(nəv)saɣ	'banana'
NCV:	Tape	(ni)saɣ	'banana'
NCV:	Paamese	sou-sou	'k.o. banana'

DECISION: It is difficult to see how this comparison can be made to work, at least in its present form. Neither POC *u > *o* nor *u > *a* are known sound changes in any Oceanic language, raising questions about the Gumawana, Taupota, Kwara'ae, 'Āre'āre and NE

Ambae forms; little is known about Taboro (which Ross 1988:205 reports as a dialect of Sinagoro), and what information we do have about the phonological history of this language is inconsistent with *daua* reflecting *sakup (Ross 1988:205-206); Roviana *hakua* cannot regularly reflect *sakup, as the implied reflexes for *all three consonants* of the proto-form are irregular; from what is known of the phonological history of Maringe (Ross 1988:219-222) *cau* cannot reflect *sakup; the Larevat, Tape and Paamese forms appear to be connected with Motu *dau*, allowing us to salvage a POC *saku, but this remains only weakly supported.

---

(5) POC *saba(l) '*petrel or albatross*'

TM:	Buma	saba	Wandering albatross: *Diomedea exulans*
Mic:	Puluwat	hapal	petrel
	Namoluk	sapal	sea bird, dark colored, blunt winged, size of noddy, never comes on land

DECISION: Proposed by Clark (2011:352). However, the Micronesian forms can only reflect a trisyllable, and since very little is known about Buma it is unclear whether *saba* could regularly derive from *sabalV. On the basis of the evidence given this comparison probably is best attributed to chance.

---

(6) POC *(s,j)abiŋ *'Acanthurus spp., incl. A. guttatus, white-spotted surgeonfish*'

Adm:	Loniu	capaŋ	possible tang or surgeonfish
Mic:	Kiribati	riba	*Acanthurus*, surgeonfish generic (vowel metathesis)
	PCP	***(s,ð)abiŋ**	*Acanthurus guttatus*, spotted surgeonfish (Geraghty)
Fij:	Lau	sabi	*A. guttatus*

	PPn	***hapi**	*Acanthurus guttatus* (Hooper)
Pn:	Tongan	hapi	k.o. fish
Pn:	Niue	hapi	*A. guttatus*, surf surgeonfish
Pn:	Niuatoputapu	hapi	*A. guttatus*
Pn:	Rennellese	api	some species of surgeonfishes
Pn:	Tokelauan	api	surgeonfish
Pn:	Tikopia	api	surgeonfish

DECISION: Proposed by Osmond (2011:104). The Central Pacific part of this comparison is straightforward, but Loniu is doubly irregular, pointing to a trisyllable with penultimate *a. Kiribati riba may be cognate, but it could also be a Polynesian loan. If accepted, it is not entirely clear whether the shape of the reconstruction would be *(s,j)abiŋ or *(s,j)ibaŋ, so until better evidence becomes available this comparison is best limited to PCP.

---

(7) POC *sapulu ‘*goatfish*’

PT:	Motu	dahuru	k.o. fish (-r- for *l)
Mic:	Pohnpeian	epil	goatfish: *Mulloidichthys vanicolensis*
Mic:	Mokilese	ɔpil	goatfish
Fij:	Fijian (Lau)	yavulu	*Mulloidichthys vanicolensis*
	PPn	***hafulu**	growth stage of goatfish
Pn:	Niuean	hafulu	goatfish
Pn:	Samoan	afulu	*M. vanicolensis*, juvenile

Pn:	Tuvaluan	afulu	yellow-banded goatfish
Pn:	Tikopia	afuru	goatfish, larger stage of *vete*
Pn:	Tahiaian	ahuru	goatfish spp.
Pn:	Hawaiian	ʔāhulu-hulu	*Upeneus porphyreus*, juvenile

DECISION: This comparison was first proposed by Osmond (2011:86). However, as she noted, Motu dahuru is phonologically irregular (expected **dahui),), and is glossed simply as 'k.o. fish', both Pohnpeian and Mokilese regularly reflect *s as d- (Bender et al. 2003), and so far as I have been able to determine no dialect of Fijian regularly loses *s, which then leaves an initial low vowel open to palatal glide epenthesis (Geraghty 1983). This leaves only PPN *hafulu as a reliable reconstruction, although the Hawaiian form should perhaps be excluded on the grounds of the unexplained initial consonant and long vowel.

---

(8) POC *s(ai)waRa *Clupeidae, sardine or herring*

Adm:	Loniu	caway	kind of sardine or anchovy
PT:	Dobuan	siwala	gold-spot herring
PT:	Molima	siwala	sardine
PT:	Kilivila	lawiya	k.o. fish (vowel metathesis)

DECISION: This comparison was proposed by Osmond (2011:40, 86). However, the first vowel in Loniu caway is unexplained, and the referent of Kilivila lawiya is not identified. Moreover, the notation on vowel metathesis implies that **liwaya would be regular, but *s > l and *R > y are otherwise unknown (*R normally was lost, and apparent examples of *R > y are products of other changes, as with the excrescent palatal glide in *Rabia > yabia 'sago', or *piRaq > viya 'taro variety'). This leaves just two closely related languages, Dobuan and Molima, as evidence for this form, providing no basis for a POC reconstruction.

---

(9) POC *sara(Ra) *sardine-like fish, possibly Atherinidae*

NNG	Yabem	(i)sala	a small slippery fish
MM	Halia	sela	sardine
	PCP	***sarā**	k.o. small schooling fish (Geraghty)
Fij	Rotuman	sarā	k.o. fish
Fij	Bauan	sarā	small fish like daniva, but with round white body
Fij	Wayan	sarā	*Atherinidae* sp., small silvery fish in coastal waters
	PPn	***sarā**	small schooling fish (Hooper)
Pn	Tongan	hā	very small schooling fish, like whitebait
Pn	Nukuoro	salā	flying fish
Pn	Luangiua	salā	small blue fish
Pn	Sikaiana	salā	k.o. fish
Pn	Takuu	sarā	k.o. small fish

DECISION: Proposed by Osmond (2011:60). However, the long vowel of the Central Pacific forms implies *saraRa, which should have become Halia **salala, and perhaps Yabem **salal. Until a firmer basis for the reconstruction is found it is best to put this comparison aside, as it could well be a product of chance.

---

(10) POC *sasaRi *midrib of coconut frond*

NNG:	Kove	sasali	midrib

NNG:	Bariai	sasal	midrib
NNG:	Mangap-Mbula	sasar	midrib of a coconut leaf
SES:	Longgu	sali-sali	rip a leaf along its midrib

DECISION: This comparison was proposed by Ross and Evans (2008:383-384), but it contains only four languages, of which the first three are confined to a single recognized subgroup of Oceanic. Given this distribution the POC status of the reconstruction depends crucially on Longgu, but here both the form and the meaning diverge from the NNG forms sufficiently to raise questions about cognation. Although *sasaRi may be valid, confidence in its validity will be shaky until a firmer foundation is established.

---

(11) POC *sele(kai) *tern*

Adm:	Loniu	cɛlɛhɛy	small white bird, possible a tern
MM:	Marovo	celekae	*Sterna* spp., esp. *albifrons*
MM:	Nduke	helekai	*Sterna* spp.
MM:	Roviana	helekae	white sea bird, often seen in flocks over a shoal of fish
MM:	Simbo	elekai	white sea bird sp.
MM:	Kia	helekai	seagull
SES:	Nggela	sele	seagull

DECISION: This comparison was proposed by Clark (2011:366). However, Loniu h reflects POC *p, not *k (Blust 1978:35-36), and there is no known basis for a morphological analysis of this form, strongly suggesting that the similarity of the Loniu word to the others cited here is due to chance. This leaves just the five forms in a single subgroup of Oceanic and Nggela sele as a basis for this comparison. While this more limited comparison may ultimately prove to be valid, the difficulty of demonstrating that the longer forms are bimorphemic, and the limitation of the distribution of these forms to the Solomons chain, where significant borrowing is known to have taken place across a major subgroup

boundary, favors caution until a stronger comparison becomes available.

---

(12) POC *seRa *Ficus sp., perhaps F. adenosperma*

Adm:	Mussau	si	*Ficus* sp.
MM:	Patpatar	sera	*Ficus adenosperma*
MM:	Tolai	ere	*Ficus adenosperma*
MM:	Nehan	her	*Ficus* sp.

DECISION: This comparison was proposed by Ross (2008:309). Since three of the member languages belong to a single putative subgroup the assignability of this form to POC depends crucially on Mussau si. However, the expected Mussau reflex of *seRa would be **sea, not si (Blust 1984), and this reconstruction is thus unjustified without further supporting evidence.

---

(13) POC *sipa *Hemiramphus spp.*

PT:	Kilivila	seva(leya)	garfish
MM:	Tolai	ive	k.o. fish
	PPn	***sipa**	young flying fish (Hooper)
Pn:	Niuean	hipa	young small flying fish
Pn:	Samoan	sipa	young flying fish
Pn:	Tokelau	hipa	young flying fish
Pn:	Tikopia	sipa	young flying fish

DECISION: First proposed by Osmond (2011:50). While the Polynesian forms are clearly related, the connection of these to the Kilivila and Tolai words is far less convincing, given the apparently arbitrary word segmentation in Kilivila seva(leya), which Senft (1986:363) writes as a single morpheme, the uncertain gloss of Tolai ive, and the irregular vowel correspondences of both Kilivila and Tolai forms to PPN *sipa.

---

(14) POC *simuk *mosquito, small biting fly*

NNG:	Tuam	sum	mosquito
NNG:	Mato	simak	sandfly
NNG:	Labu	sumu(si)	mosquito
PT:	Wedau	imo(kini)	mosquito
		kini	'to sting'
PT:	Tawala	himo(kini)	mosquito
		himo-himo(kini)	sandfly
PT:	Dawawa	simo(kin)	mosquito
MM:	Mono-Alu	simuʔu	midge
MM:	Varisi	simu-simu	midge
MM:	Avasö	simuku	mosquito
MM:	Maringe	si-simi	housefly
SES:	Lau	simi	sandfly
SES:	Baegu	si-simi	midge
SES:	Longgu	simi	mosquito
SES:	Kwai	simi(sakʷalo)	mosquito

SES:	Kwaio	simi	fly, sandfly
		simi(lak^{w}alo)	mosquito
SES:	Dori'o	simi(lak^{w}alo)	midge
		simi(ni-ōne)	mosquito; sandfly
SES:	ʻĀreʻāre	sime	mosquito
SES:	Sa'a	sime	mosquito

DECISION: This reconstruction was proposed by Osmond (2011:382). However, it is hard to see how it follows from the supporting evidence given, since the only forms that can be said to reflect *simuk without one or more irregularities are those in Mono-Alu, Varisi and Avasö, all of which belong to what is generally considered a single subgroup of Oceanic. All other forms differ etymologically in one or both vowels, and allow the reconstruction of well-supported forms only in lower-order proto-languages. While it is tempting to try to combine this diverse material under a single ancestral form, all that can be done securely is to posit *sVmV, qualifying this as a 'near comparison' in terms of the standards of evidence adopted in the ACD.

---

(15) PMP *siRik *dorsal fin?*

WMP:	Molbog	sirik	dorsal fin
	Minangkabau	siriʔ	dorsal fin
	Buginese	siriʔ	dorsal fin
	Palauan	sirik	dorsal fin
	POC	***siRiko**	fish fin

MM:	Lakalai	siliko-liko	fins
NCV:	Uripiv	siki	fin
Fij:	Wayan	siko-silo	fin

DECISION: Proposed by Osmond (2011:134). The PMP form reportedly is based on material drawn from Tryon (1995:2:338-339). However, the sound correspondences for Palauan are seriously irregular (*siRik should yield **tisk or **tisek), and in any case no such form is found in McManus and Josephs (1977), leading to the conclusion that it is simply an error in Tryon (1995). Although a 4-page sketch of Molbog is given by Zorc and Thiessen (1995:1.1:359-362), it provides no wordlist, and the source of the Molbog word given here is obscure, since it does not appear in the only published source of lexical data on this language (Thiessen 1981), or the most extensive set of lexical data on Molbog available to me (Lobel n.d.). Wilkinson (1959:1115) gives Malay variants sirik, sirip and sirit for 'fin (of fish)', but Buginese siri? (which could reflect any of these) does not appear in any primary sources I have been able to check, and Minangkabau is too closely related to Malay to provide evidence for a reconstruction of any significant time-depth, forcing us to conclude that PMP *siRik 'dorsal fin' is a fiction.

The evidence for POC *siRiko is equally problematic. Little is known of Uripiv, but what data we have suggests that *R usually yields r (Tryon 1976:30, where two examples of irregular loss are reported), and no parallels to the vowel change *o > i are known, making this a very weak comparison. Similarly, although the first half of the Wayan form given by Osmond could reflect *siRiko the second half would be unexplained, and in any event this form does not appear in Pawley and Sayaba (2003), where iriiri is given for 'fin'. This etymology clearly illustrates the dangers of using questionable secondary sources, and of substituting what amounts to little more than guesswork based on phonetic similarity for serious application of the comparative method.

---

(16) POC *sua-sua goatfish

NNG:	Manam	sua-sua	goatfish, catfish
MM:	Marovo	sua(ra)	generic for goatfish
NCV:	S. Efate	sus	goatfish

SV:	Kwamera	(ie)su	goatfish
NCal:	Iaai	(wa)si	*M. flavolineatus*
Fij:	Wayan	ðūðū	*Parapeneus* spp. goatfish

NOTE: This comparison was proposed by Osmond (2011:85). However, apart from Manam sua-sua none of these forms can be derived from the proposed reconstruction by regular changes. The Marovo form suggests *suaR (with regular echo vowel rather than some unrelated morphological attachment). It is unclear whether the South Efate, Kwamera or Iaai forms are related, and Wayan ðūðū is irregular both in failing to reflect a final vowel, and in the length of the vowel it has (cp. POC *matuqa > matua 'mature, fully developed', *puaq > vua 'fruit', *rua > rua 'two', etc.). Tentatively, it is perhaps possible to posit POC *suaR 'goatfish', although this is still not strongly supported.

R.A.B. 2/22/2016

printable (pdf) version of this introduction

---

see also:
The Austronesian Comparative Dictionary: A Work in Progress
by Robert Blust and Stephen Trussel
(*Oceanic Linguistics*, Vol. 52. no. 2, December 2013)