

Supplementary materials for  
**“New methodologies for historical linguistics?  
Calibrating a lexicon-based methodology for diffusion vs. subgrouping”**

Mark Donohue, Tim Denham, Stephen Oppenheimer  
The Australian National University / La Trobe University / University of Oxford

### **1. Success at replication**

Gray et al. (2009) claim to replicate 26 of the 34 ‘major subgroups’ of Austronesian from Blust (2009).<sup>1</sup> We examined the subgrouping of languages in Gray et al.’s Supporting Online Materials, where a more detailed clustering is presented, and found a number of discrepancies.<sup>2</sup>

We discount the St. Matthias group from the comparison. Blust (2009) lists only two languages in this Oceanic group, Mussau and Tenis. Since no information is available on Tenis, the St. Matthias group is represented by a single language and so its uniqueness will automatically and trivially be replicated in any tree; any single language necessarily forms a ‘subgroup’ unto itself in any family or subgroup, and so its replication cannot be taken as a sign of success, just as the presence of Greek in Figure 3 cannot be counted for, or against, a claim for replicating the subgroups of Indo-European. A similar process was adopted for other ‘isolate’ languages that appear in either Blust’s or Gray et al.’s phylogenies, notably many of the languages of Taiwan that have not been subgrouped together. Following Blust (2009), this leaves 31 non-trivial subgroups outside Taiwan. We compared these to the tree presented in Gray et al.’s Supporting Online Materials (Table S1).

The numbered subgroups, following the list in Blust (2009) of ‘major subgroups’ in the Austronesian family, are listed on the left of Table 2. Successful replication, as reported by Gray et al. (2009), is shown in the column headed ‘Gray et al. A’; + represents a report of successful replication and – indicates that they do not report replication for this subgroup. The final column, ‘Gray et al. B’, indicates the results as reported in Gray et al.’s Supporting Online Materials, which are in some cases at variance with the results reported in Gray et al.’s main text. In this table \* indicates that a subgroup reported as successfully replicating Blust (2009) in Gray et al.’s print article contains languages which do not belong to this subgroup according to Gray et al.’s Supporting Online Materials. Groups that are discussed in these supplementary materials are indicated in bold. Carefully examining the material in Gray et al.’s Supporting Online Materials reduces the number of successful replications from 25 to 15 (excluding the St. Matthias group as discussed above). As mentioned above, in all cases geographic proximity or distance in social space explains the discrepancy.

---

<sup>1</sup> It is unclear what is meant by ‘major subgroup’, since the term does not refer to first-order subgroups of any sort. Rather than investigate this question, we simply adopt the same heuristic that was used by Gray et al. (2009).

<sup>2</sup> Parenthetically, we note that while Blust (2009) lists 34 subgroups in a table, a number of additional subgrouping hypotheses, not considered by Gray et al., are discussed in the text. We address some of the more topical of these subgroups in these supplementary materials.

**Table S1.** Comparing reported and actual replication of the Austronesian family.

	Linguistic subgrouping (Blust 2009)	Gray et al. A	Gray et al. B
1	<b>Polynesian</b>	+	+
2	Central Pacific	+	+
3	Nuclear Micronesian	+	+
4	South-East Solomonic	+	+
5	North-Central Vanuatu	+	+
6	South Vanuatu	+	+
7	New Caledonia/Loyalty islands	+	+
8	<b>North New Guinea</b>	+	*
9	Papuan Tip	+	+
10	<b>Meso-Melanesian</b>	+	*
11	Admiralties	+	+
12	<i>St. Matthias</i>	+	+
13	Western Oceanic	-	-
14	Oceanic	+	+
15	<b>South Halmahera/West New Guinea</b>	+	*
16	<b>Eastern Malayo-Polynesian</b>	+	*
17	Central Malayo-Polynesian	-	-
18	Central-Eastern Malayo-Polynesian	+	+
19	<b>Central Maluku</b>	+	*
20	<b>Yamdena-North Bomberai</b>	+	*
21	Celebic	+	+
22	<b>Greater South Sulawesi</b>	+	*
23	Greater Barito	-	-
24	Malayo-Chamic	-	-
25	Barrier Islands/North Sumatra	-	-
26	North Sarawak	+	+
27	<b>North Borneo</b>	+	*
28	<b>Philippines</b>	+	*
29	Greater Central Philippines	-	-
30	Western Malayo-Polynesian	-	-
31	Malayo-Polynesian	+	+
32	Western Plains	-	-
33	East Formosan	-	-
34	<b>Malayo-Sumbawan</b>	+	*
Subgroups replicated:		25/31	15/31

## 2. Inaccurate replications of subgroups

The exceptions to accurate replication that we identify in Table 1 are described one-by-one below, identified by subgroup name (following Blust 2009) and by number (corresponding to Table 1). As mentioned above, in all cases geographic proximity or distance in social space explains the discrepancy.

### 2.1 North New Guinea (8)

Gray et al. erroneously include Nakanai as a sister to the North New Guinea languages. Nakanai, a Meso-Melanesian language, is neither a member of the North New Guinea linkage, nor particularly closely related to it. Significantly for understanding this error, Nakanai is spoken in an area adjacent to that occupied by speakers of North New Guinea languages, and is geographically removed from speakers of most other Meso-Melanesian languages, suggesting that intimate contact with North New Guinea languages has affected the lexicon of Nakanai to the extent that it has been misclassified.

### 2.2 Meso-Melanesian (10)

Gray et al. incorrectly include Mussau, the sole attested representative of the St. Matthias family, in the Meso-Melanesian subgroup. For accurate replication, Meso-Melanesian should include Vitu, which is found more separated from the languages in the Meso-Melanesian group than is Mussau in Gray et al.'s clustering, and Nakanai (see 2.1) while excluding Mussau. Either Mussau is included with Vitu, and the replication suffers from imprecision, or Mussau and Vitu are both excluded, and a poorer recall of the necessary languages is achieved.

The problems in the replication of North New Guinea and Meso-Melanesian can be accounted for by lexical borrowing due to social interaction. Mussau is spoken in the St. Matthias area, with all trade and contacts running south to New Ireland, where Meso-Melanesian languages are spoken (Lynch, Ross & Crowley 2002). These problems are acknowledged by Gray et al. (2009b: 6):

The Willaumez subgroup of Meso-Melanesian [containing Vitu – MD, TD & SO] is not included here, but is placed as a sister group to the North New Guinea languages (1.00). These Willaumez dialects are spoken on New Britain and are close neighbors to the North New Guinea languages there, and this placing may therefore reflect contact-induced change in these languages.

This is an explicit acknowledgement that the method used is susceptible to local influences, rather than ancient inheritances.

### 2.3 South-Halmahera/West New Guinea (15)

Although the inclusion of Irarutu (the general name applied by Lewis (2009), equivalent to Gray et al.'s (2009b) 'Kasira') in the South-Halmahera/West New Guinea group is firmly supported

by numerous regular sound correspondences (Ross 1995), it is not replicated in Gray et al.'s classification. Irarutu is spoken in a bay facing the west of New Guinea, socially more proximal to the languages of Maluku which are not part of the South-Halmahera/West New Guinea group. Gray et al.'s classification of this language outside the South-Halmahera/West New Guinea subgroup reflects the effects of contact and lexical diffusion with languages of Maluku.

#### **2.4 Eastern Malayo-Polynesian (16)**

See 2.3. This subgroup should also include Irarutu, since South-Halmahera/West New Guinea is a subgroup within Eastern Malayo-Polynesian.

#### **2.5 Central Maluku (19)**

The languages of Aru (an archipelago in south-eastern Maluku), such as Ujir and Ngaibor, should not be included in this grouping (as classified by Gray et al.). The social prestige of societies in the Central Maluku area may have been responsible for lexical diffusion into the Aru Islands where Ujir and Ngaibor are spoken, and where a Central Maluku presence is documented historically. However, there are no indications that Central Maluku and Aru show any shared innovations, and no linguists have suggested linking them together to the exclusion of other languages of the region.

#### **2.6 Yamdena-North Bomberai (20)**

Gray et al.'s composition of this subgroup is problematic and seems to reflect geographical proximity and lexical borrowing. Blust (2009) does not include Koiwai or Kei in this subgroup, though these languages are subgrouped closely within Yamdena-North Bomberai by Gray et al. By way of an explanation we note that Koiwai is close geographically and historically to Sekar, which is accurately placed with Yamdena, and that Kei is found midway between Yamdena and the North Bomberai region. Diffusion between the northern Bomberai peninsula and the southern languages of Tanimbar, around Yamdena, accounts for the inclusion of geographically intermediate groups such as Kei. Again, geography explains the clustering of the languages.

#### **2.7 Greater South Sulawesi (22)**

According to Blust (2009), this group should comprise the languages of South Sulawesi and the Tamanic languages of central Borneo, specifically Maloh in Gray et al.'s dataset. Maloh has, however, been treated by Gray et al. as an isolate, not closely linked to the languages of South Sulawesi in Gray et al.'s clustering. Maloh is located inland in Borneo, far from the otherwise geographically contiguous South Sulawesi languages, and it is apparent that this disjunct distribution has resulted in Gray et al.'s failure to detect the subgroup. Gray et al.'s result is significant in that the only difference between the low-level subgroup 'South Sulawesi' and the Greater South Sulawesi group is the inclusion of the Tamanic languages of Borneo with the South Sulawesi languages of South Sulawesi. By failing to include Maloh, Gray et al.'s replication is of the geographically contiguous South Sulawesi group, not of the more significant, and geographically disparate, Greater South Sulawesi group.

## **2.8 North Borneo (27)**

According to Blust (2009), this group should not include Barito languages, such as Ngaju Dayak and Katingan, whereas these languages are included within North Borneo by Gray et al. These Barito languages are geographically contiguous with the other languages in the north of Borneo, including the languages of the North Borneo subgroup, and lexical borrowing is likely to be the cause of the incorrect association by Gray et al.

## **2.9 Philippines (28)**

According to Blust (2009), this group should include the Sangiric languages of northern Sulawesi. In Gray et al.'s classification the Sangiric languages are treated as their own group nested closely with the other Sulawesi-area languages, their southern neighbours with which they are more socially engaged. Lexical borrowings from Sulawesi have resulted in the incorrect subgrouping of Sangiric languages as an outlier of a subgroup that contains the other Sulawesi groupings.

The Sama/Bajau group is incorrectly subgrouped with the Philippine languages; the proximity of most of the socially dominated Sama/Bajau languages to languages of the southern Philippines has resulted in numerous lexical borrowings that have skewed the results away from a true representation of their position in the Austronesian tree. Indeed, Gray et al. (2009b: 8) acknowledge the role of social and geographic factors in this false subgrouping:

Sama-Bajaw languages are placed at the base of the Greater Central Philippines subgroup. This probably reflects the large amounts of contact-induced change between the Sama-Bajaw and the speakers of Philippine languages (Blust 1991).

## **2.10 Malayo-Sumbawan (34)**

According to Blust (2009), this group should not include Javanese, or languages from Sumatra such as Lampung, Gayo or Batak. While the internal structure of Malayo-Sumbawan is not completely resolved, there is a consensus (Adelaar 2005, Blust 2009) that Javanese and most of the Sumatran languages should not be included. The prestige and historical dominance of the Malayo-Polynesian isolate Javanese in western Indonesia has resulted in extensive lexical borrowing by neighboring languages, affecting its classification by Gray et al. as part of a subgroup with its geographic neighbours, and reinforcing the dangers of comparison based purely on unscreened lexical materials. All of the false positives identified by Gray et al. in this group are geographically surrounded by members of the (true) Malayo-Sumbawan subgroup, isolated from other non-Malayo-Sumbawan Austronesian languages, and all show extensive contact-induced lexical change (e.g., Blust 1981, Adelaar 2005).

## **2.11 Malayo-Chamic (24)**

In Gray et al. (2009: Supporting Online Materials), the Chamic languages are listed as a first order sister to Malayo-Sumbawan, and are not closely linked with Malayic, contradicting all linguistic opinion (eg., Blust 1981, Thurgood 1999). The Chamic languages are maximally distant, socially and geographically, from all other languages of the Malayo-Sumbawan group and have been documented as having undergone extensive lexical borrowing from non-

Austronesian languages (eg., Thurgood 1999, Grant 2005), making the lexical signal picked up by Gray et al. contradict the evidence provided by sound change and lexical innovation.

In addition to these failures to accurately replicate accepted linguistic subgroups, further discrepancies can be found in the clustering that Gray et al. produce for Austronesian languages.

### **2.12 Malayo-Polynesian (31)**

The node is accurately replicated, but Paiwan is incorrectly assigned by Gray et al. (2009b) to a subgroup containing Malayo-Polynesian to the exclusion of other mainland Taiwanese languages. Such a subgrouping is not advocated anywhere in the literature on Austronesian languages, and is not supported by any work on the languages of Taiwan or on Austronesian historical linguistics (e.g., Blust 2009, Ross 2009). Paiwan, spoken at the extreme south of mainland Taiwan, is the closest Taiwanese language, geographically, to the Malayo-Polynesian languages spoken on various islands to the south and is known to have been lexically affected by recent (in historical times) contact with northern Philippine languages. Gray et al.'s deviation from accepted Austronesian language phylogenies can be attributable to geographical proximity facilitating erratic social contact.

### **2.13 'Bima-Sumba'**

Gray et al. (2009b: 8) report “a strongly supported clustering (1.00) of the controversial Bima-Sumba subgroup”, which contains the languages” of Sumba, Sawu and Bima. However, following a reassessment of historical relationships between these languages, Blust (2008: 45) concluded on basis of the linguistic evidence from Sumba that “there is no support for a more inclusive subgroup that incorporates Bimanese”. Bima is immediately north of, and very close to, the languages of Sumba, and has been the home of a highly influential regional sultanate in historical times. This has probably resulted in extensive lexical diffusion, a factor that can most likely account for Gray et al.'s erroneous inclusion of Bima together with Sawu and the languages of Sumba.

## **3. Polynesian replication quantified**

Table S2 evaluates the success that Gray et al. have in replicating the different subgroups of Polynesian, as defined by work by Marck (2000) and others. The table shows the number of true positives, false positives and false negatives (tp, fp and fn, respectively), and the precision and recall values calculated from them (following van Rijsbergen 1979), followed by the harmonic mean (F1) of these two scores. The result is the same 83% replication rate that we saw as the average of the values in Table 3 in the main article.

**Table S2.** Replication rates for Polynesian subgroups.

	tp	fp	fn	Precision	Recall	F1
Tongic	1	1	1	0.5	0.5	0.5
Nuclear Polynesian	28	1	1	0.96	0.97	0.97
Ellicean Outliers	5	1	2	0.83	0.71	0.77
Eastern Polynesian	10	0	0	1	1	1
Marquesic	2	1	0	0.67	1	0.8
Tahitic	6	1	0	0.86	1	0.92
<b>Average</b>						<b>0.83</b>

Table S3 is the twin of Table S2, but approaches the question of replication from the other perspective: how well does the Marck (2000) tree for Polynesian replicate the clusterings described in Gray et al. (Supporting Online Materials).

**Table S3.** Rates of reproduction of accepted Polynesian subgroups by the hierarchical organisation of Polynesian in Gray et al. (2009).

	tp	fp	fn	Precision	Recall	F1
Tongan-Samoan	1	1	0	0.5	1	0.67
Anuta-plus	28	1	1	0.97	0.97	0.97
Rennellese-plus	27	1	2	0.96	0.93	0.95
Niue-E.Uvea-E.Futuna	1	2	1	0.33	0.5	0.4
Niue-E.Uvea	1	1	1	0.5	0.5	0.5
Emae-W.Futuna	1	5	2	0.17	0.33	0.22
Tikopia-Luangiua	7	1	2	0.88	0.78	0.82
Nukuoro-Luangiua	5	0	2	1	0.71	0.83
Tuvalu-Tokelau	2	0	5	1	0.29	0.44
Eastern Polynesian	10	0	0	1	1	1
Rapa Nui-Marquesan	2	1	1	0.67	0.67	0.67
Mangareva-Marquesan	2	0	1	1	0.67	0.8
Tuamotuan-Rurutu	6	1	0	0.86	1	0.92
Rarotonga-Rurutu	5	1	1	0.83	0.83	0.83
Penrhyn-Rurutu	4	1	2	0.8	0.67	0.73
Hawaiian-Rurutu	3	1	3	0.75	0.5	0.6
Tahitian-Rurutu	2	0	0	1	1	1
<b>Average:</b>						<b>0.73</b>

#### 4. Associated geography

If we assume, following Gray et al. (2009: 482), that linguistic phylogenies can “resolve questions about human prehistory”, specifically to “trace the history of human populations” (2009: 479), then we must assume that the demic diffusion of languages links linguistic splits with population movements.

We have seen that Gray et al.'s Bayesian phylogenetic trees obtained from cognate comparison of basic vocabulary are not, *contra* Gray et al. (2009: 479), “congruent with the traditional subgroups identified by phonological and morphological evidence”. Here we examine the implications that the traditional tree (following Blust 2009) and the tree presented by Gray et al. have for theories of demic diffusion. Following the subgrouping in Figure S1A, which represents Blust (2009), we have an associated geographic range as shown S1C, and discussed in various publications by Blust, including Blust (2009). The Gray et al. tree, shown in Figure S1B, has an associated geography as shown in Figure S1D. From these, proceeding from the root of the tree to its terminal nodes, we can construct the phylogeographies shown as Figure S1E, following Blust's tree, and Figure S1F, following Gray et al. (2009). The figures were constructed by drawing a line of best fit between the centroids of each parent node, and assuming that the location of each parent node can be approximated by averaging the centroid locations of each of its first-order daughter subgroups (in practise, then, the figure is constructed from the bottom of the tree up). In very few respects are there any congruences between the two beyond the pegged ‘start’ in Taiwan and the ‘finish’ in the Pacific.

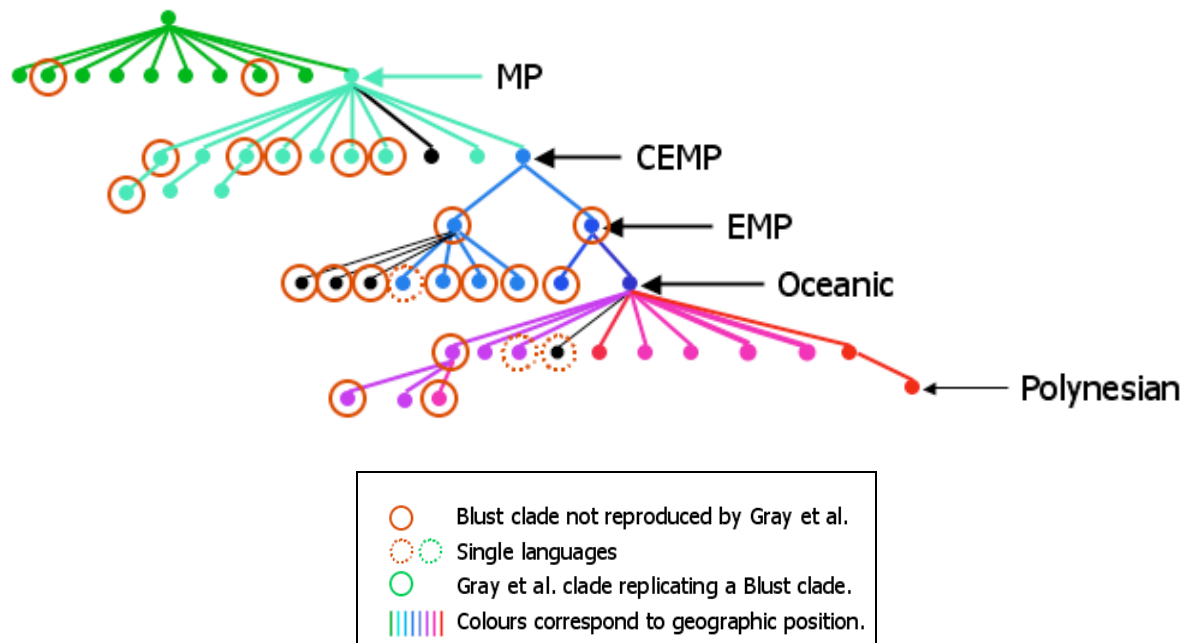
### Supplementary References

- Adelaar, Alexander K. 1994. “Malay and Javanese Loanwords in Malagasy, Tagalog and Siraya (Taiwan)”. *Bijdragen tot de Taal-, Land en Volkenkunde* 150. 50–65.
- Lewis, M. Paul, ed. 2009. *Ethnologue: Languages of the world*. Dallas, Texas: SIL International.
- Ross, M. D. 1995. “Some Current Issues in Austronesian Linguistics”. *Comparative Austronesian Dictionary: An introduction to Austronesian studies* ed. by Darrell T. Tryon, Malcom Ross, Adrian Clynes, Charles Grimes & Alexander Adelaar, Vol. 1, Pt. 1, 45–120. Trends in Linguistics, Documentation 10. Berlin: Mouton de Gruyter.
- Ross, Malcolm D. 2009. “Proto Austronesian Verbal Morphology: A reappraisal”. *Austronesian Historical Linguistics and Culture History: A festschrift for Robert Blust* ed. by Alexander Adelaar & Andrew Pawley, 295–326. Canberra: Pacific Linguistics.
- Thurgood, Graham. 1999. *From Ancient Chamic to Modern Dialects: Two thousand years of language contact and change* (= Oceanic Linguistics Special Publications, 28). Honolulu: University of Hawai'i Press.
- van Rijsbergen, Keith. 1979. *Information Retrieval*. London: Butterworth.

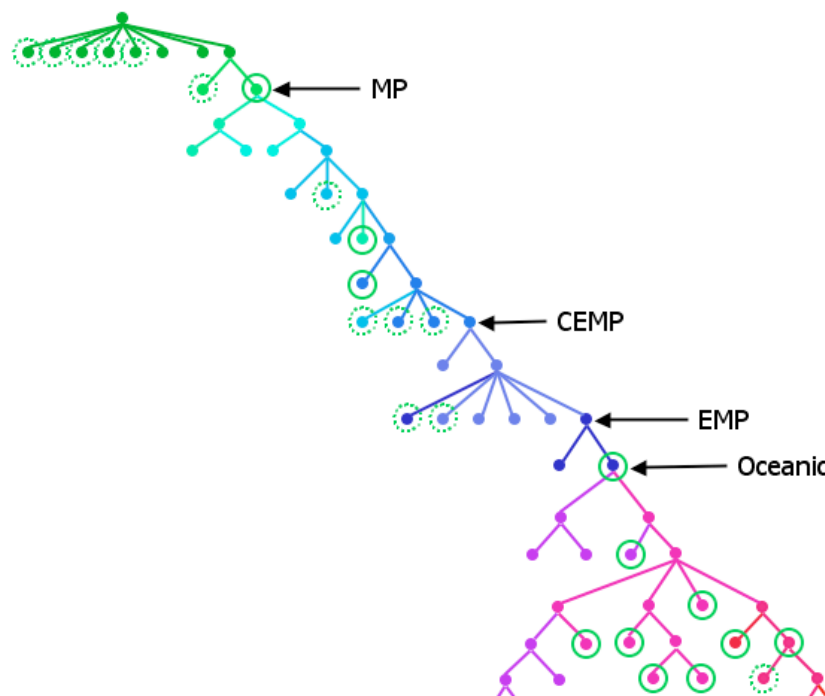


**Fig. S1.** Linguistic discrepancies between Blust (2009) and Gray et al. (2009): S1A. Subgrouping of Austronesian languages by Blust; S1B. Clustering of Austronesian languages by Gray et al.; S1C. Geography following Blust (2009); S1D. Geography following Gray et al. (2009); S1E. Schematic phylogeography following Blust (2009); S1F. Schematic phylogeography following Gray et al. (2009). The coloured geocoding of languages is consistently applied for Blust’s (2009) subgroups (Figs. S1A, S1C and S1E) and for Gray et al.’s clusters (Figs. S1B, S1D and S1F).

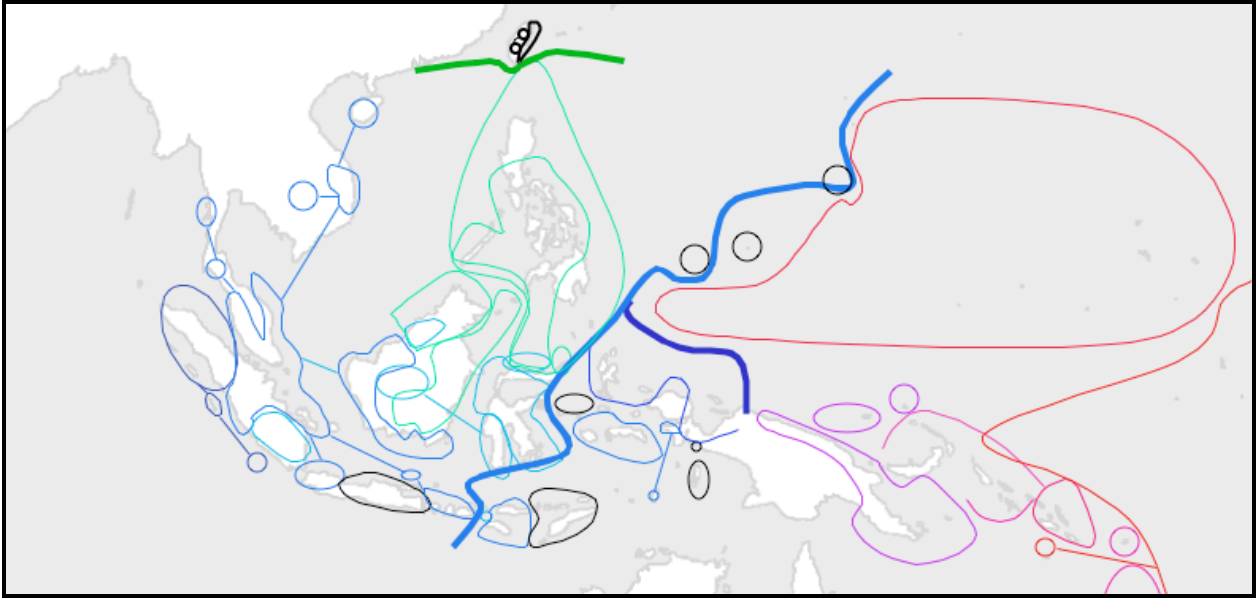
**S1A**



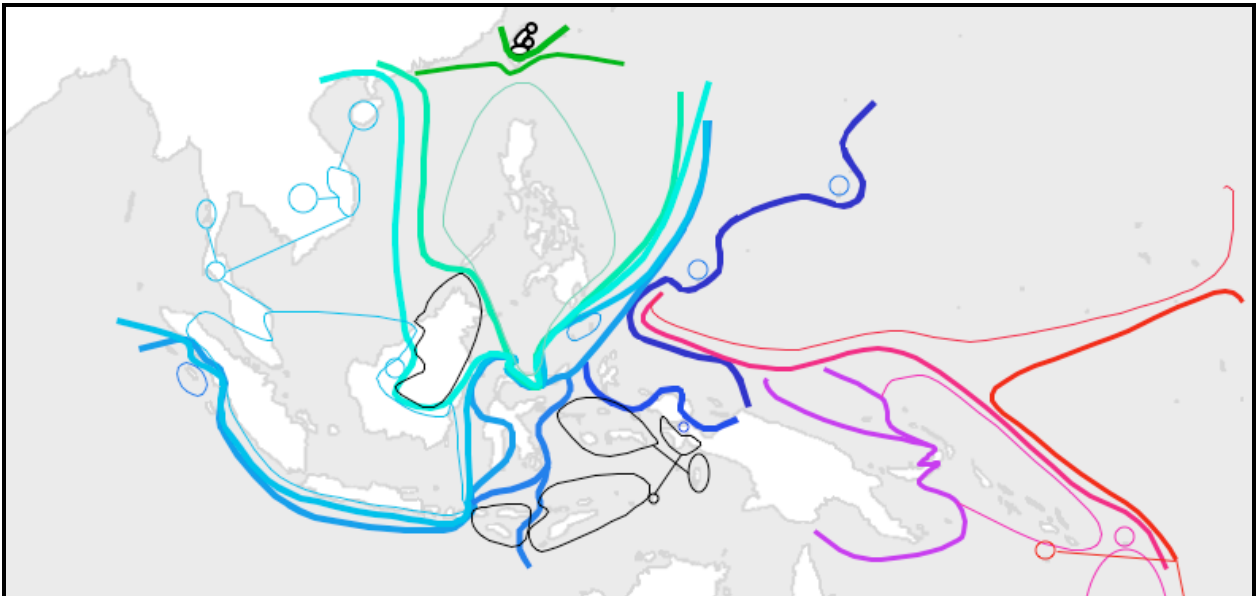
**S1B**



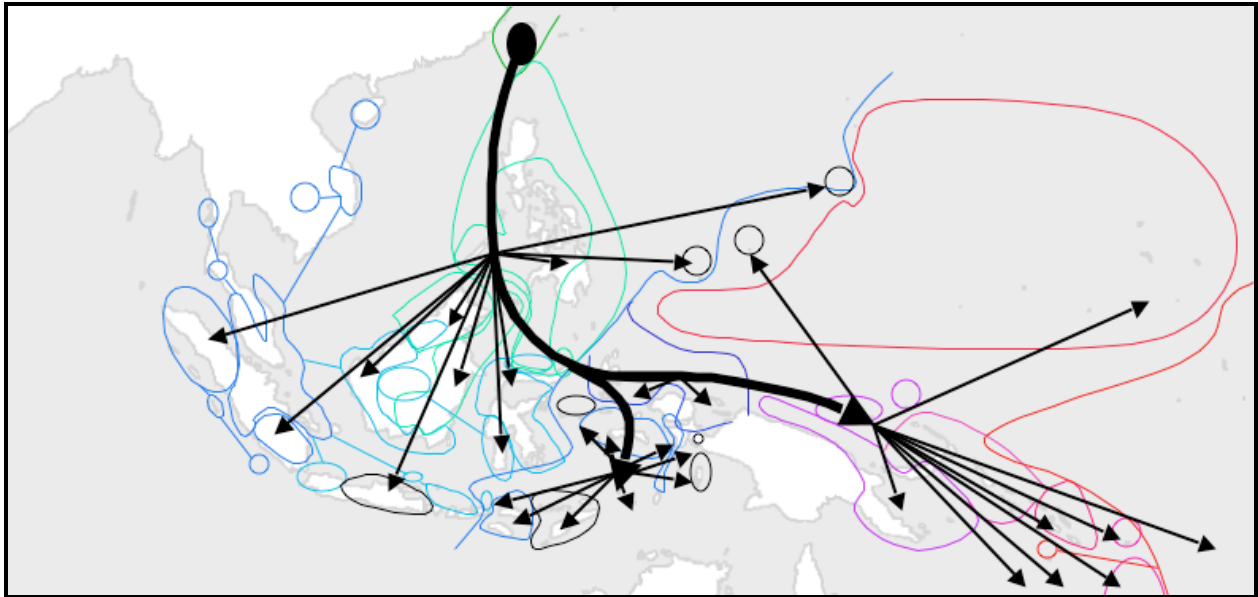
S1C



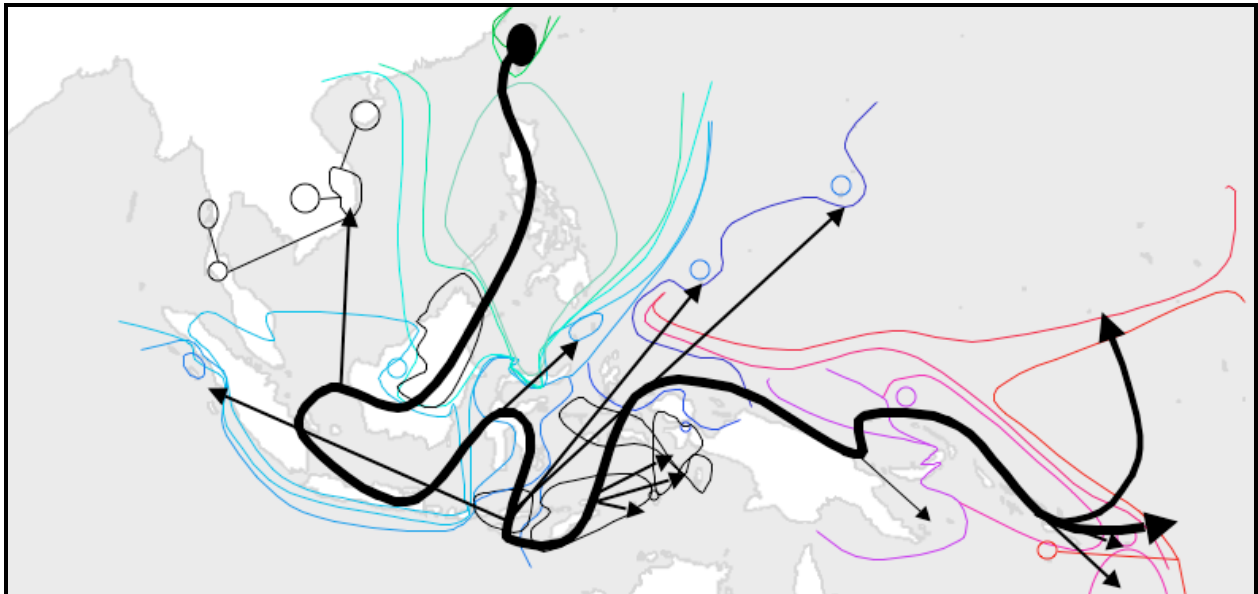
S1D



**S1E**



**S1F**



**Swadesh 200-item list**

The Swadesh list we used to calculate loans in modern English is shown below, with the provenance of different lexemes indicated.

	<b>Item</b>	<b>Provenance</b>
1	All	Proto-Germanic
2	And	Proto-Germanic

3	Animal	<b>French</b>
4	Ashes	Proto-Germanic
5	At	Proto Indo-European
6	Back	Proto-Germanic
7	Bad	Old English?
8	bark (of a tree)	<b>Norse</b>
9	Because	<b>French</b>
10	Belly	Proto-Germanic
11	Big	<b>Norse</b>
12	Bird	Old English
13	to bite	Proto Indo-European
14	Black	Proto-Germanic
15	Blood	Proto-Germanic
16	to blow (wind)	Proto Indo-European
17	Bone	Proto-Germanic
18	to breathe	Proto-Germanic
19	to burn (intrans)	Proto-Germanic
20	child (young)	Old English
21	Cloud	Proto-Germanic
22	cold (weather)	Proto Indo-European
23	to come	Proto Indo-European
24	to count	<b>French</b>
25	to cut	<b>Norse</b>
26	day (not night)	Proto-Germanic
27	to die	<b>Norse</b>
28	to dig	<b>French</b>
29	Dirty	<b>Norse</b>
30	Dog	Old English
31	to drink	Proto-Germanic
32	dry (substance)	Proto-Germanic
33	dull (knife)	Proto Western Germanic
34	Dust	Proto-Germanic
35	Ear	Proto Indo-European
36	earth (soil)	Proto-Germanic
37	to eat	Proto Indo-European
38	Egg	<b>Norse</b>
39	Eye	Proto Indo-European
40	to fall (drop)	Proto Indo-European
41	Far	Proto Indo-European
42	fat (substance)	Proto-Germanic
43	Father	Proto Indo-European

44	to fear	Proto-Germanic
45	feather (large)	Proto-Germanic
46	Few	Proto Indo-European
47	to fight	Proto Indo-European
48	Fire	Proto Indo-European
49	Fish	Proto Indo-European
50	Five	Proto Indo-European
51	to float	Proto-Germanic
52	to flow	Proto-Germanic
53	Flower	<b>French</b>
54	to fly	Proto Indo-European
55	Fog	ModernEnglish
56	Foot	Proto Indo-European
57	Four	Proto Indo-European
58	to freeze	Proto Indo-European
59	Fruit	<b>French</b>
60	to give	<b>Norse</b>
61	Good	Proto-Germanic
62	Grass	Proto-Germanic
63	Green	Proto-Germanic
64	Guts	Proto-Germanic
65	Hair	Proto-Germanic
66	Hand	Proto Indo-European
67	He	Proto Indo-European
68	Head	Proto Indo-European
69	to hear	Proto Indo-European
70	Heart	Proto Indo-European
71	Heavy	Proto-Germanic
72	Here	Proto-Germanic
73	to hit	<b>Norse</b>
74	hold (in hand)	Proto-Germanic
75	How	Proto Indo-European
76	to hunt (game)	Proto-Germanic
77	Husband	<b>Norse</b>
78	I	Proto Indo-European
79	Ice	Proto-Germanic
80	If	Proto-Germanic
81	In	Proto Indo-European
82	to kill	Old English
83	know (facts)	Proto Indo-European
84	Lake	<b>French</b>

85	to laugh	Proto-Germanic
86	Leaf	Proto-Germanic
87	left (hand)	Proto Western Germanic
88	Leg	<b>Norse</b>
89	to lie (on side)	Proto Indo-European
90	to live	Proto-Germanic
91	Liver	Proto-Germanic
92	Long	Proto Indo-European
93	Louse	Proto Indo-European
94	man (male)	Proto Indo-European
95	Many	Proto Indo-European
96	meat (flesh)	Proto-Germanic
96b	Moon	Proto Indo-European
97	Mother	Proto Indo-European
98	Mountain	<b>French</b>
99	Mouth	Proto Indo-European
100	Name	Proto Indo-European
101	Narrow	Proto-Germanic
102	Near	Proto-Germanic
103	Neck	Proto-Germanic
104	New	Proto Indo-European
105	Night	Proto Indo-European
106	Nose	Proto Indo-European
107	Not	Proto Indo-European
108	Old	Proto-Germanic
109	One	Proto Indo-European
110	Other	Proto Indo-European
111	Person	<b>French</b>
112	to play	Proto Western Germanic
113	to pull	<b>Low German</b>
114	to push	<b>French</b>
115	to rain	Proto-Germanic
116	Red	Proto Indo-European
117	right (correct)	Proto Indo-European
118	right (hand)	Proto Indo-European
119	River	<b>French</b>
120	Road	Proto-Germanic
121	Root	<b>Norse</b>
122	Rope	Proto-Germanic
123	rotten (log)	<b>Norse</b>
124	Rub	<b>Low German</b>

125	Salt	Proto Indo-European
126	Sand	Proto Indo-European
127	to say	Proto Indo-European
128	scratch (itch)	MiddleEnglish
129	sea (ocean)	Proto-Germanic
130	to see	Proto Indo-European
131	Seed	Proto Indo-European
132	to sew	Proto Indo-European
133	sharp (knife)	Proto-Germanic
134	Short	Proto Western Germanic
135	to sing	Proto-Germanic
136	to sit	Proto Indo-European
137	skin (of person)	<b>Norse</b>
138	Sky	<b>Norse</b>
139	to sleep	Proto Indo-European
140	Small	Proto Indo-European
141	to smell (perceive odor)	Old English
142	Smoke	Proto-Germanic
143	Smooth	Proto Indo-European
144	Snake	Proto-Germanic
145	Snow	Proto Indo-European
146	Some	Proto Indo-European
147	to spit	Proto-Germanic
148	to split	<b>Dutch</b>
149	to squeeze	ModernEnglish
150	to stab (or stick)	MiddleEnglish
151	to stand	Proto Indo-European
152	Star	Proto Indo-European
153	stick (of wood)	Proto-Germanic/ <b>Norse</b>
154	Stone	Proto-Germanic
155	Straight	Proto-Germanic
156	to suck	Proto Indo-European?
157	Sun	Proto Indo-European
158	to swell	Proto-Germanic
159	to swim	Proto-Germanic
160	Tail	Proto-Germanic
161	That	Proto Indo-European
162	There	Proto-Germanic
163	They	<b>Norse</b>
164	Thick	Proto Indo-European
165	Thin	Proto Indo-European

166	to think	Proto-Germanic
167	This	Proto Indo-European
168	thou/you	Proto Indo-European
169	Three	Proto Indo-European
170	to throw	Proto Indo-European
171	to tie	Proto-Germanic
172	Tongue	Proto Indo-European
173	tooth (front)	Proto Indo-European
174	Tree	Proto-Germanic
175	to turn (veer)	<b>French</b>
176	Two	Proto Indo-European
177	to vomit	<b>French</b>
178	to walk	Proto Western Germanic
179	warm (weather)	Proto Indo-European
180	to wash	Proto-Germanic
181	Water	Proto Indo-European
182	We	Proto Indo-European
183	Wet	Proto Indo-European?
184	What	Proto Indo-European
185	When	Proto Indo-European
186	Where	Proto Indo-European
187	White	Proto Indo-European
188	Who	Proto Indo-European
189	Wide	Proto-Germanic
190	Wife	Proto-Germanic
191	wind (breeze)	Proto Indo-European
192	Wing	<b>Norse</b>
193	Wipe	Proto-Germanic
194	with (accompanying)	Proto Indo-European
195	Woman	Old English
196	Woods	Proto Indo-European
197	Worm	Proto Indo-European
198	Ye	Proto Indo-European
199	Year	Proto Indo-European
200	Yellow	Proto Indo-European