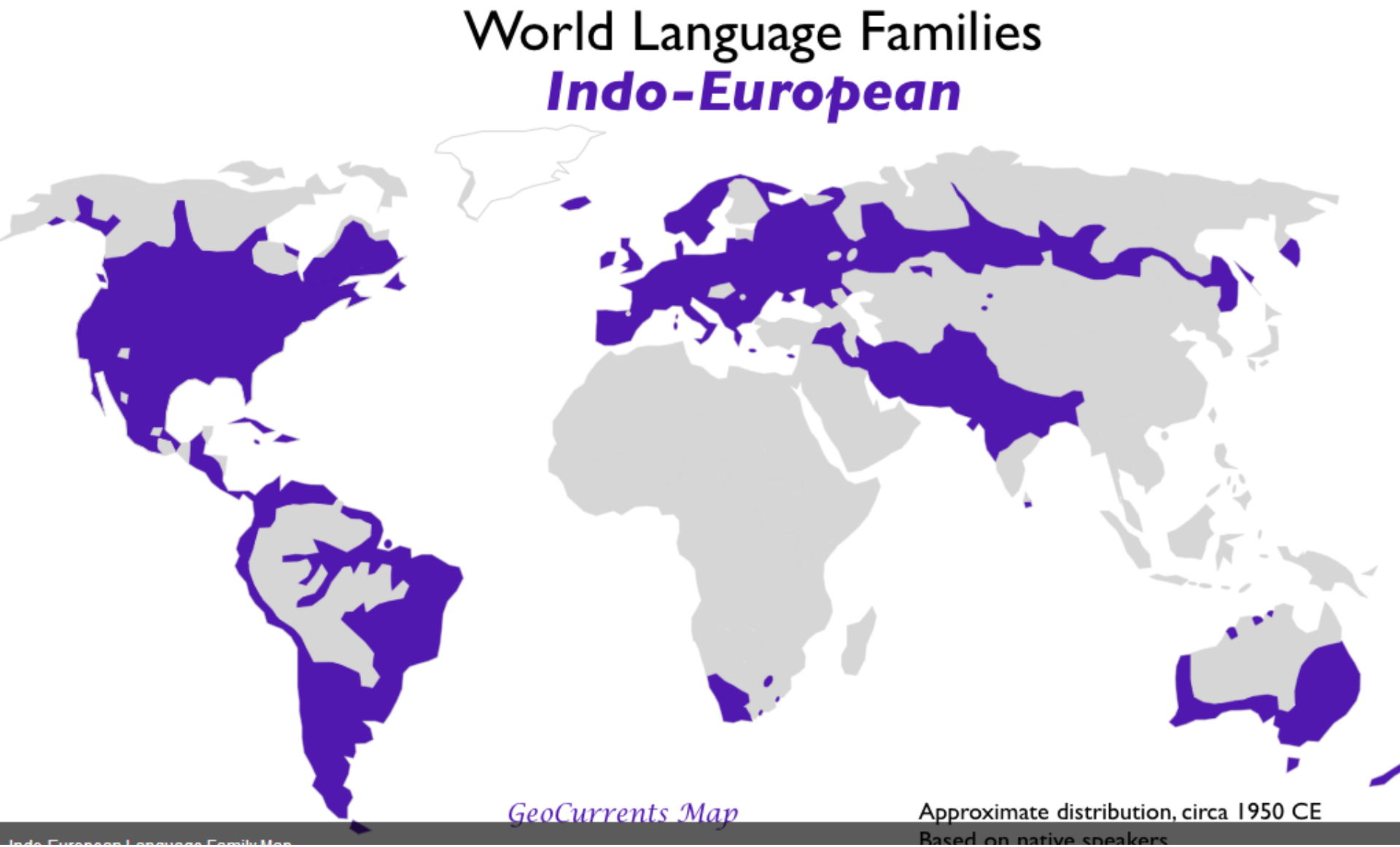
Austronesia: "Empire" of the Southern Islands

This is a map of the world published in 1707, at the dawn of European Pacific Discovery. (Note the ignorance of the Pacific and the supposition that California was an island!) This map was aware of about half of what we now call Austronesia, the Island Empire settled by people with a common linguistic ancestor. In this talk we will learn a little bit about the history of Austronesian settlement, and see what – if anything – lexical analysis can contribute to our understanding.

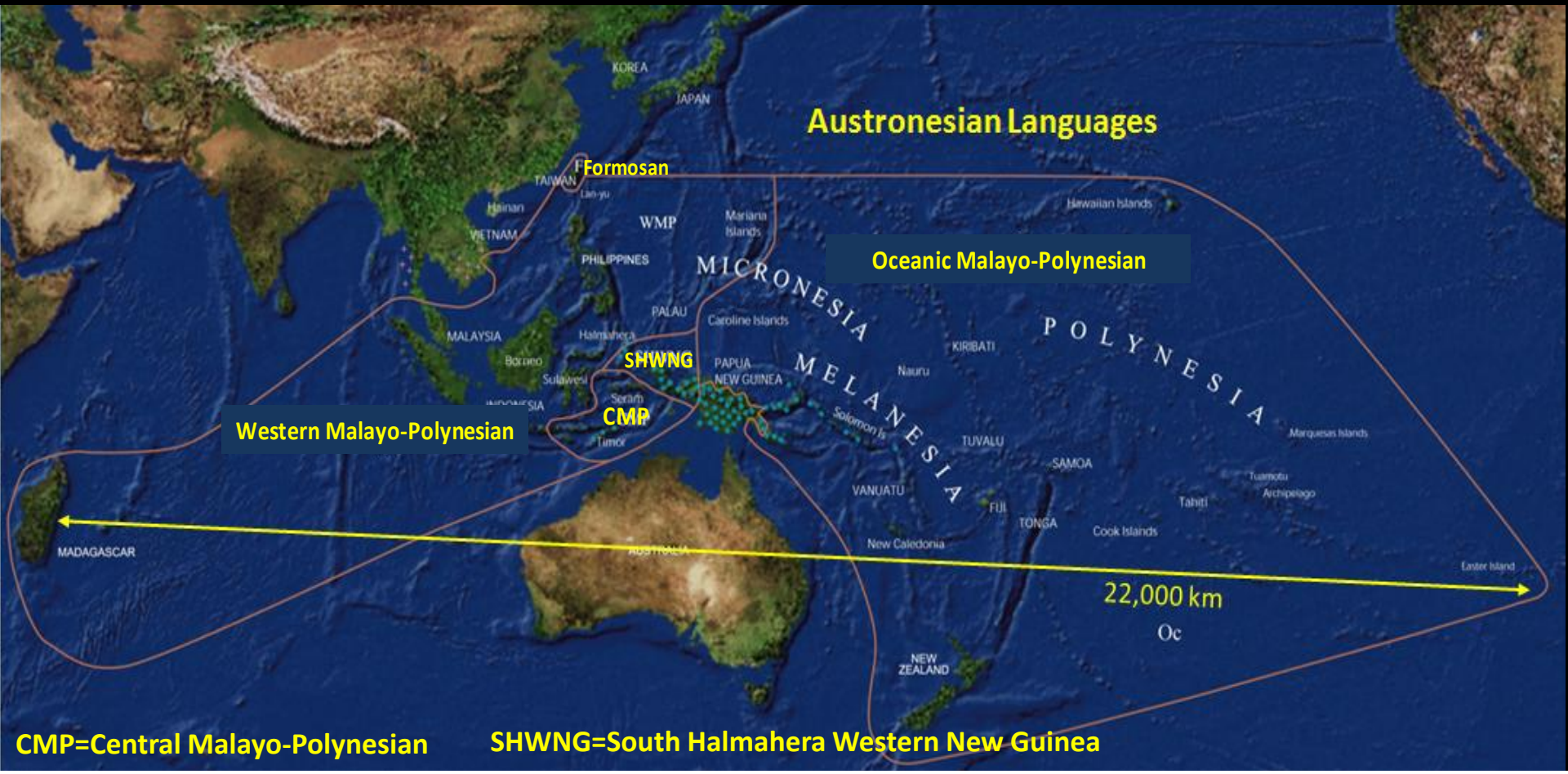


Indo-European languages are #1 in the world today, with over 3 billion native speakers, or about 45% of the world population.

Pre-Colonial Extent of Indo-European Languages

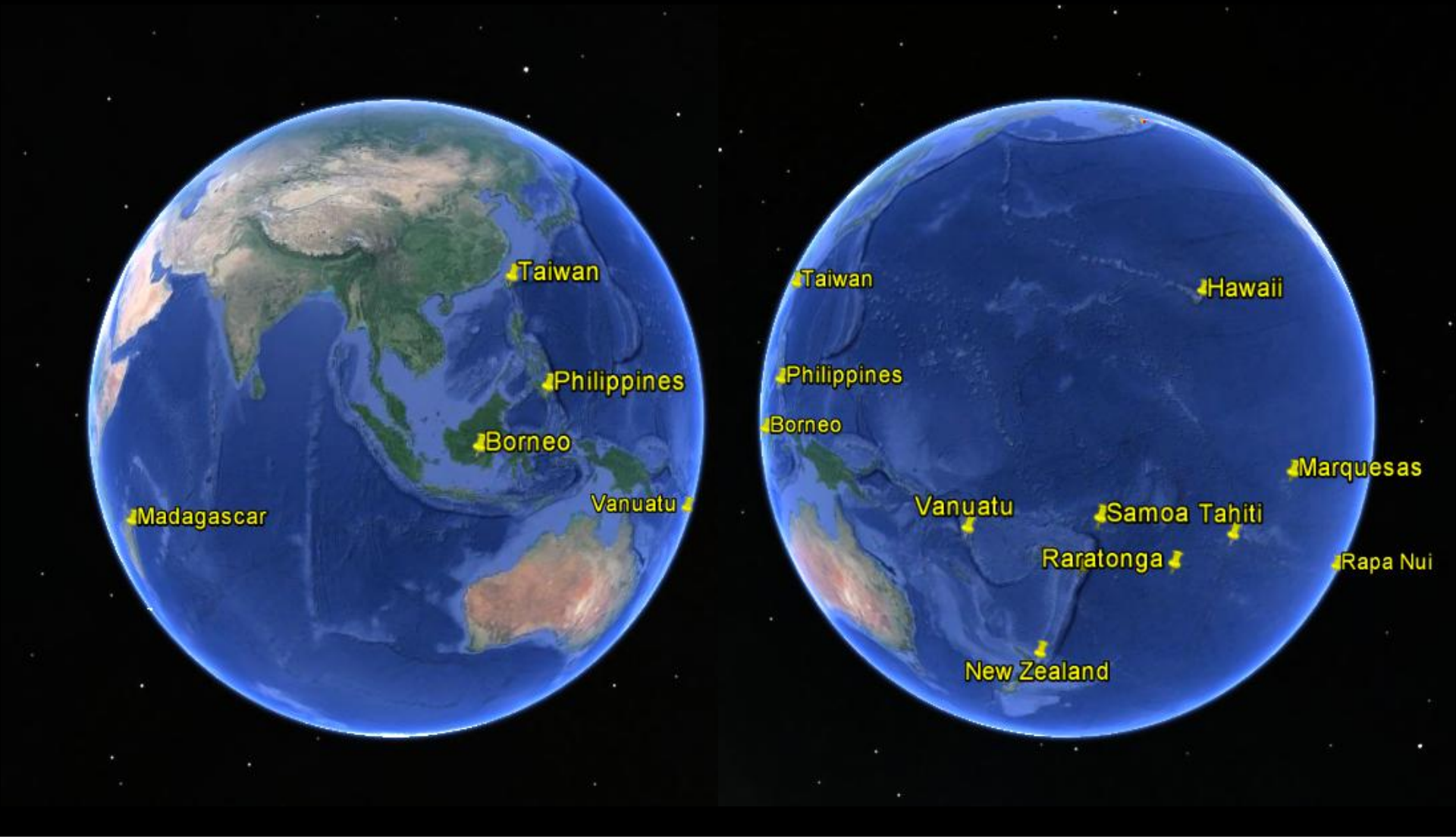
But turn back the clock to pre-colonial days and you see these languages covered just Europe and India, as you would expect. The total extent was about 8300 km, from Iceland to Assam.

Austronesia, an Empire of Related Languages



But in pre-colonial times, the geographic extent for what would come to be called the Austronesian Family of languages was much greater, about 22,000 km, all the way from Madagascar to Rapa Nui (Easter Island). Today Austronesian is #2 in number of languages (1291 per Wikipedia), #5 in number of speakers (over 400 million, or 5.9% of the world's population) The most widespread language sub-group is called Malayo-Polynesian, and it further divided into Western, Oceanic and a couple of Central subgroups.

Austronesia



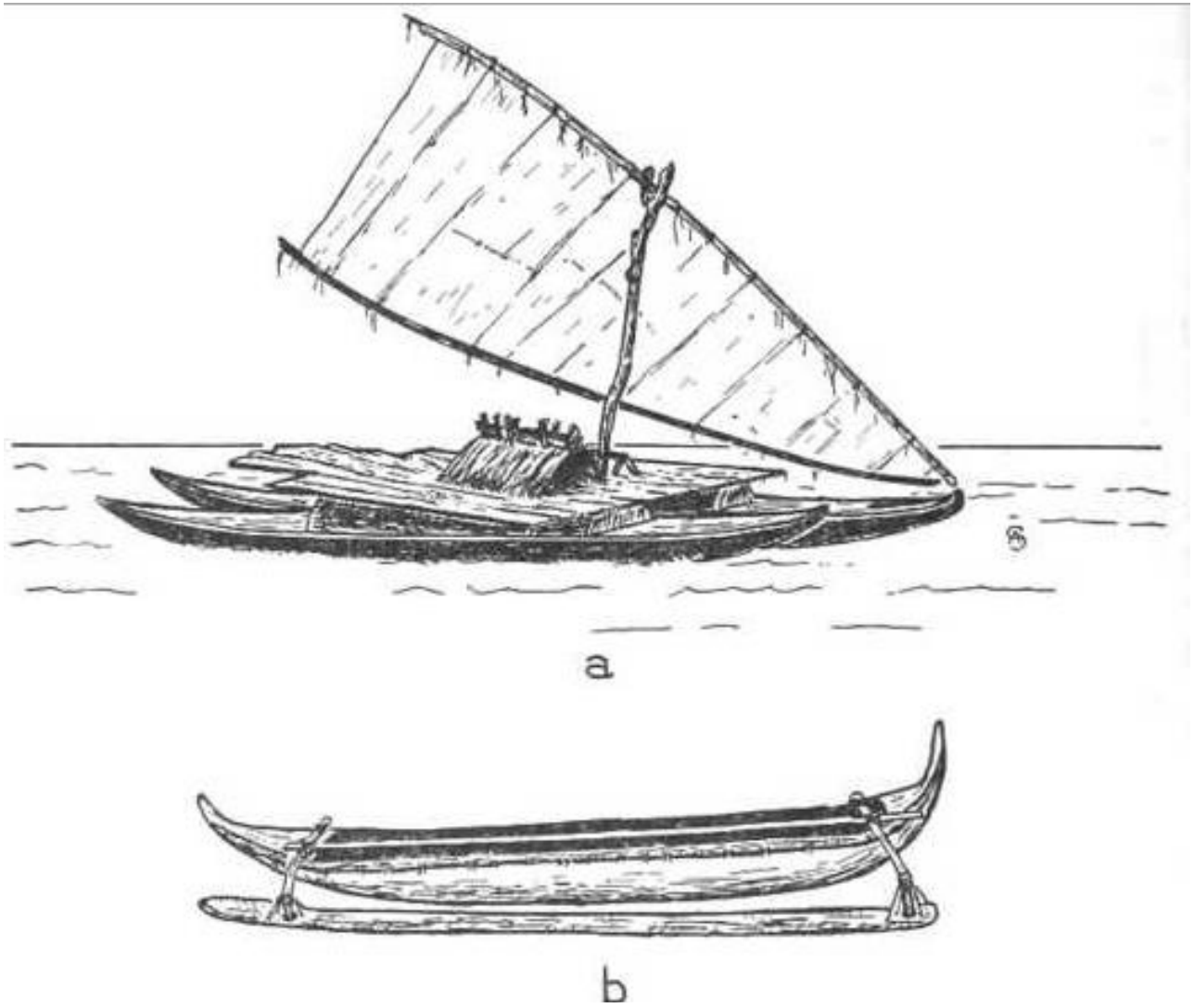
That 22,000km extent is enormous. You can't show it in a single hemisphere…it's spread over 206 degrees of longitude, and has been for a thousand years or so.

Austronesia

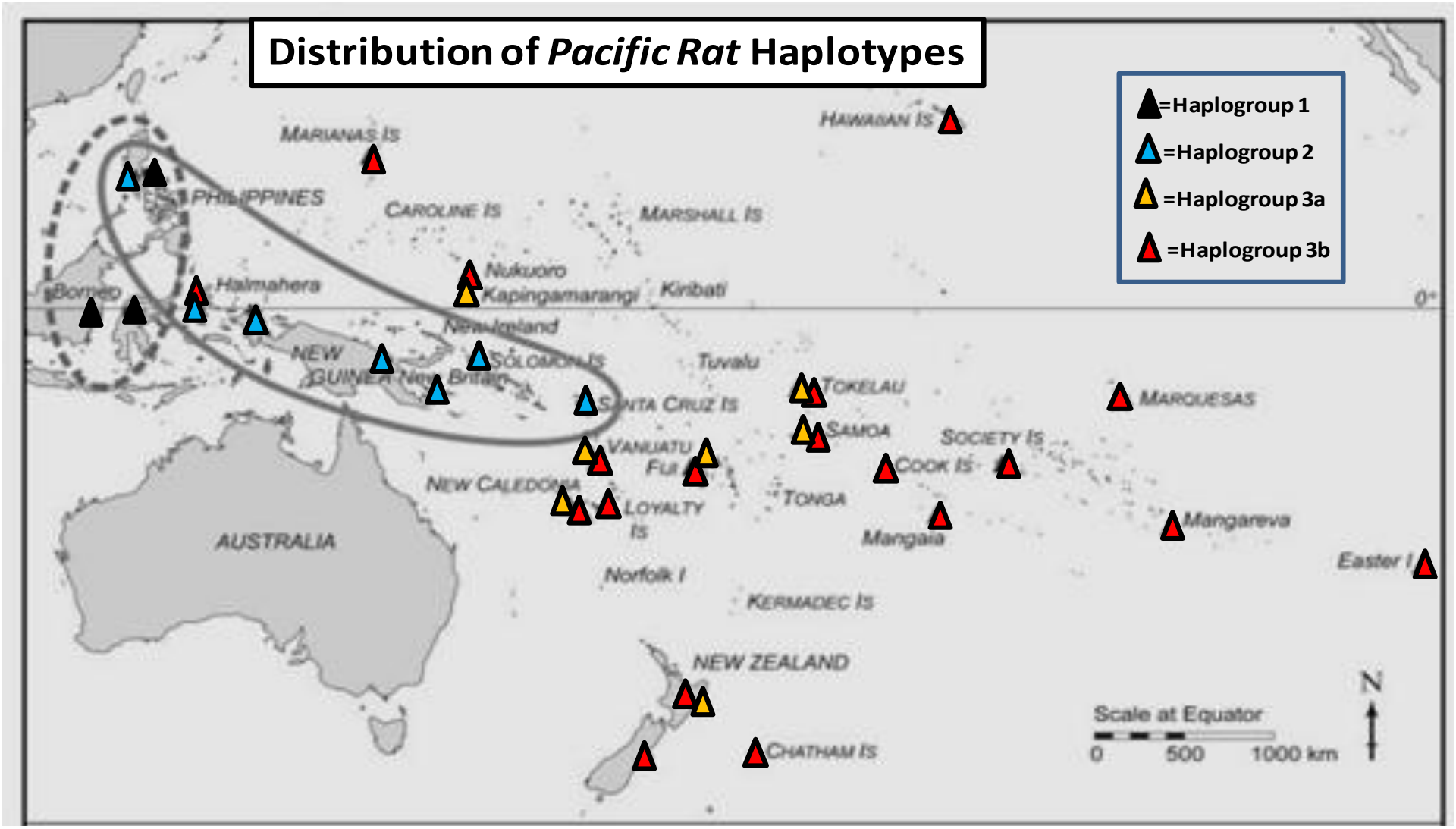
It appears that these languages originated on the island of Taiwan (or Formosa), where 9 out of the 10 sub-families exist, or did exist until recently. All the remaining languages, the ones we'll be discussing, are in the Malayo-Polynesian subfamily.

And because all of these languages are spoken on Islands, and because many of these islands were quite isolated until modern times, the Austronesian Family is particularly interesting to study.

The Technology that Settled Austronesia

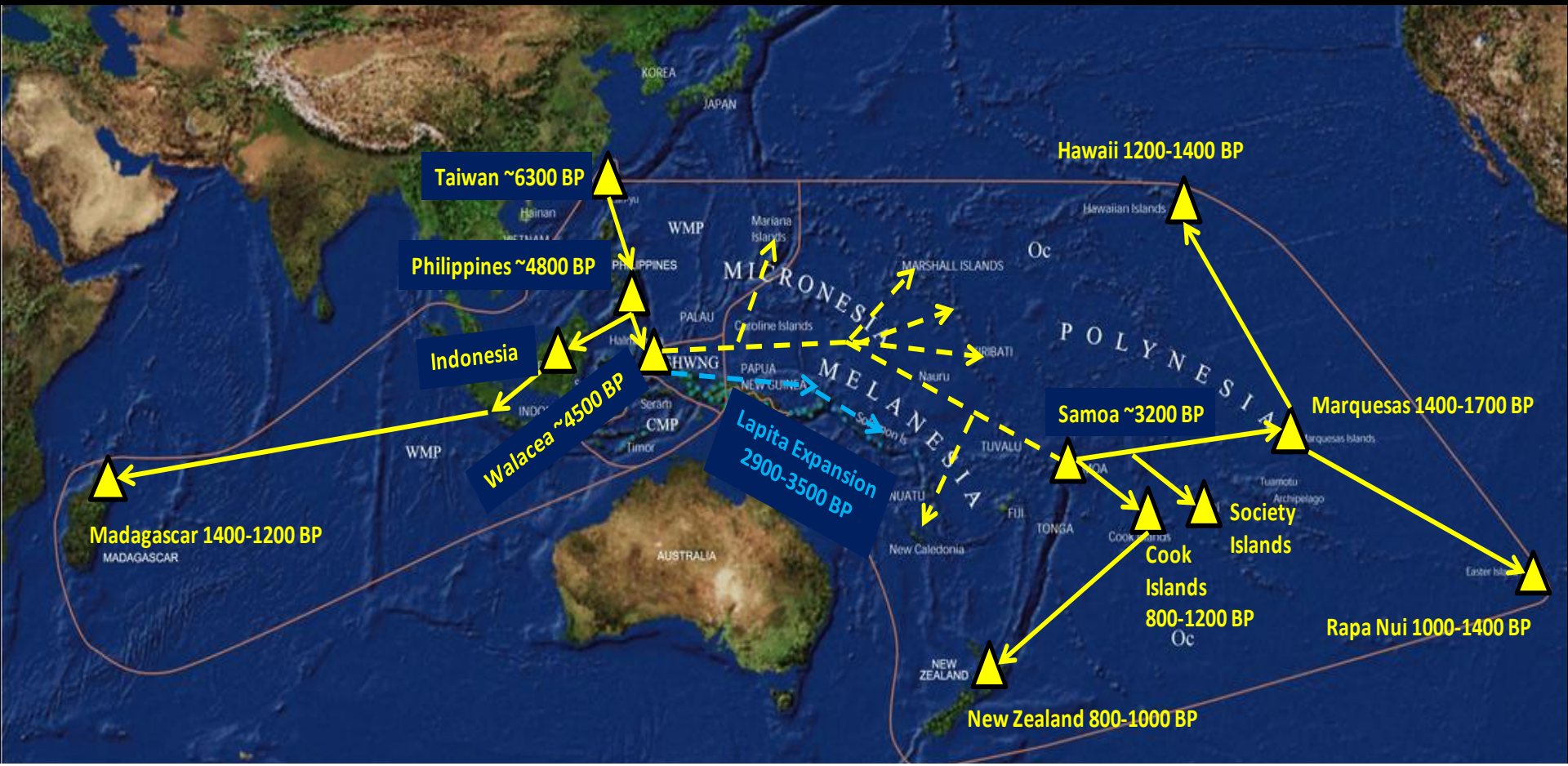


Here's the technology that made it possible for people to spread over such enormous distances: the outrigger canoe and double-canoe, or catamaran. These are such valuable inventions they are in use even today.



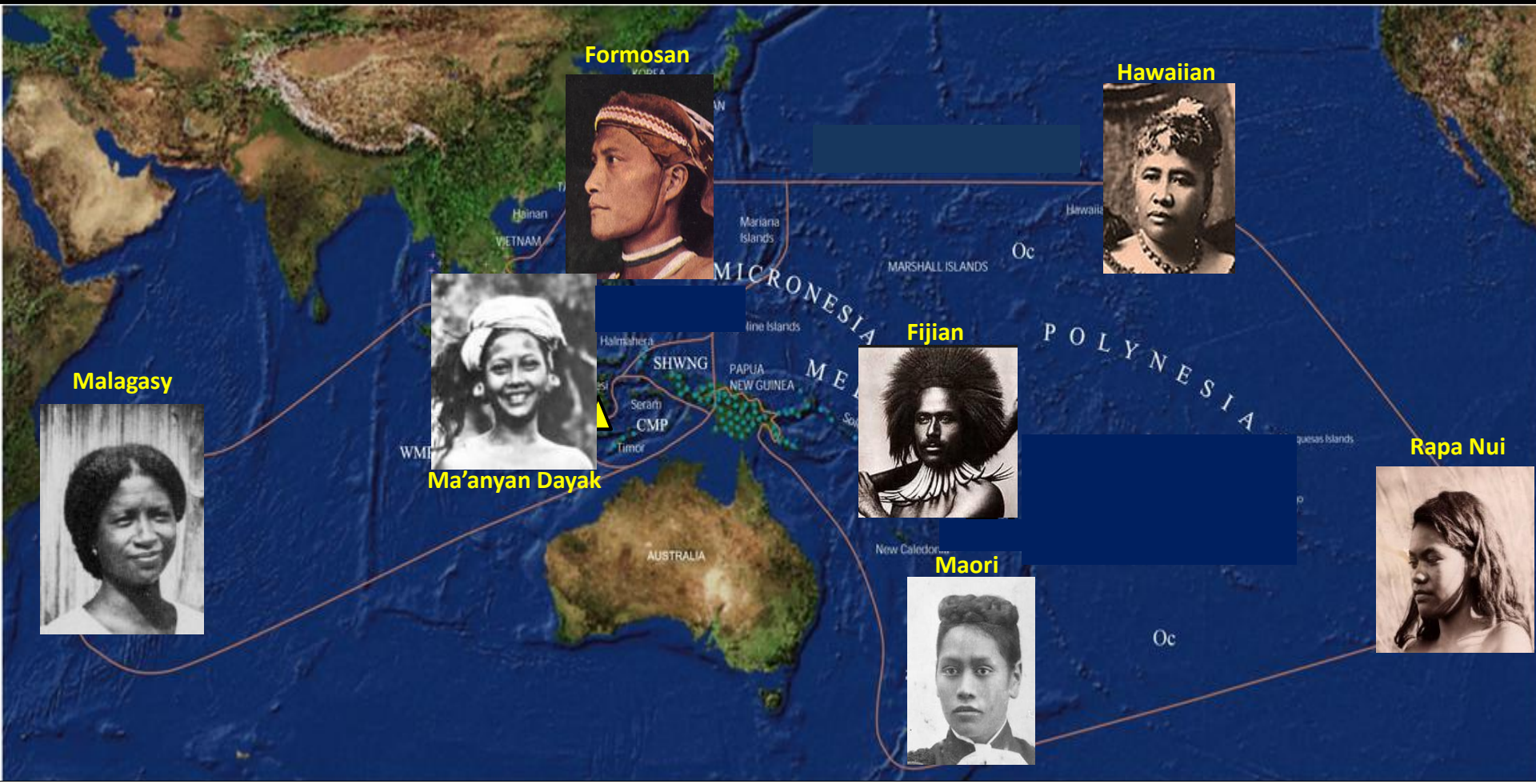
And these pacific explorers didn't venture forth by accident or because they got lost. They were exploring for new islands on which to settle. It was a determined effort, brought about by population pressures and human conflicts on tiny specks of land in a vast sea. The Pacific Rat [rattus exulans] didn't ride along accidentally, it was brought as a source of food. These rats are now distributed throughout the Pacific. Haplogroup 1 never moved East, only Haplogroup 2 made its way along the New Guinea coast, and only Haplogroup 3 travelled farther east, into Polynesia. This seems to imply that there were two separate eastward migrations, one along the shores of New Guinea, and the other across the wild blue ocean , both originating in an area that is between Indonesia and New Guinea.

Theory of Austronesian Expansion

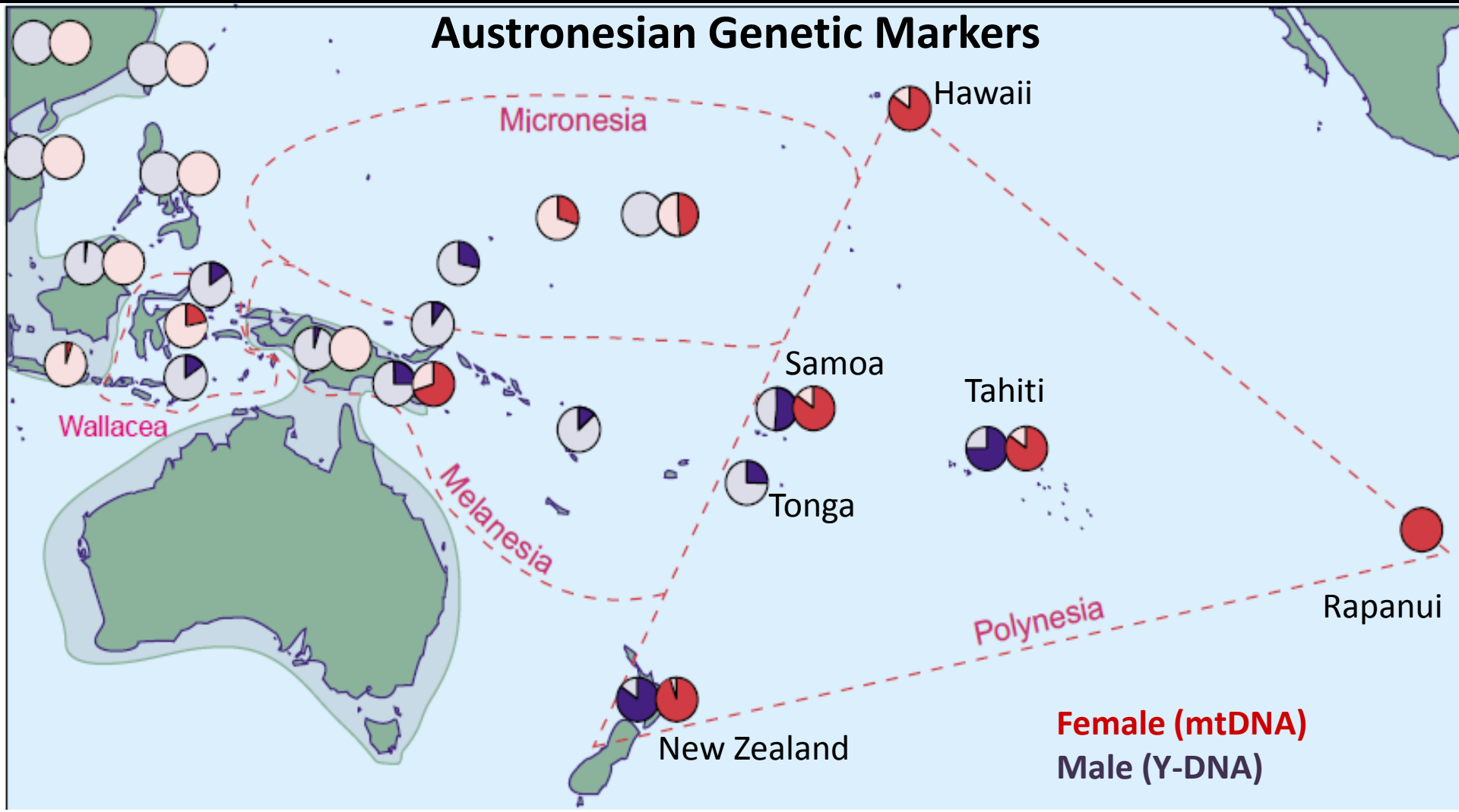


Combining archaeological, DNA and linguistic evidence, we can get this picture of how people spread through Austronesia. Polynesia was finally inhabited by about 1000 years ago, as far north as Hawaii, as far south as New Zealand, and as far east as Rapa Nui (Easter Island). Just a few hundred years before the pinnacle of eastward settlement, the western boundary of Austronesia was established with the arrival of Indonesian voyagers in Madagascar.

Austronesian People

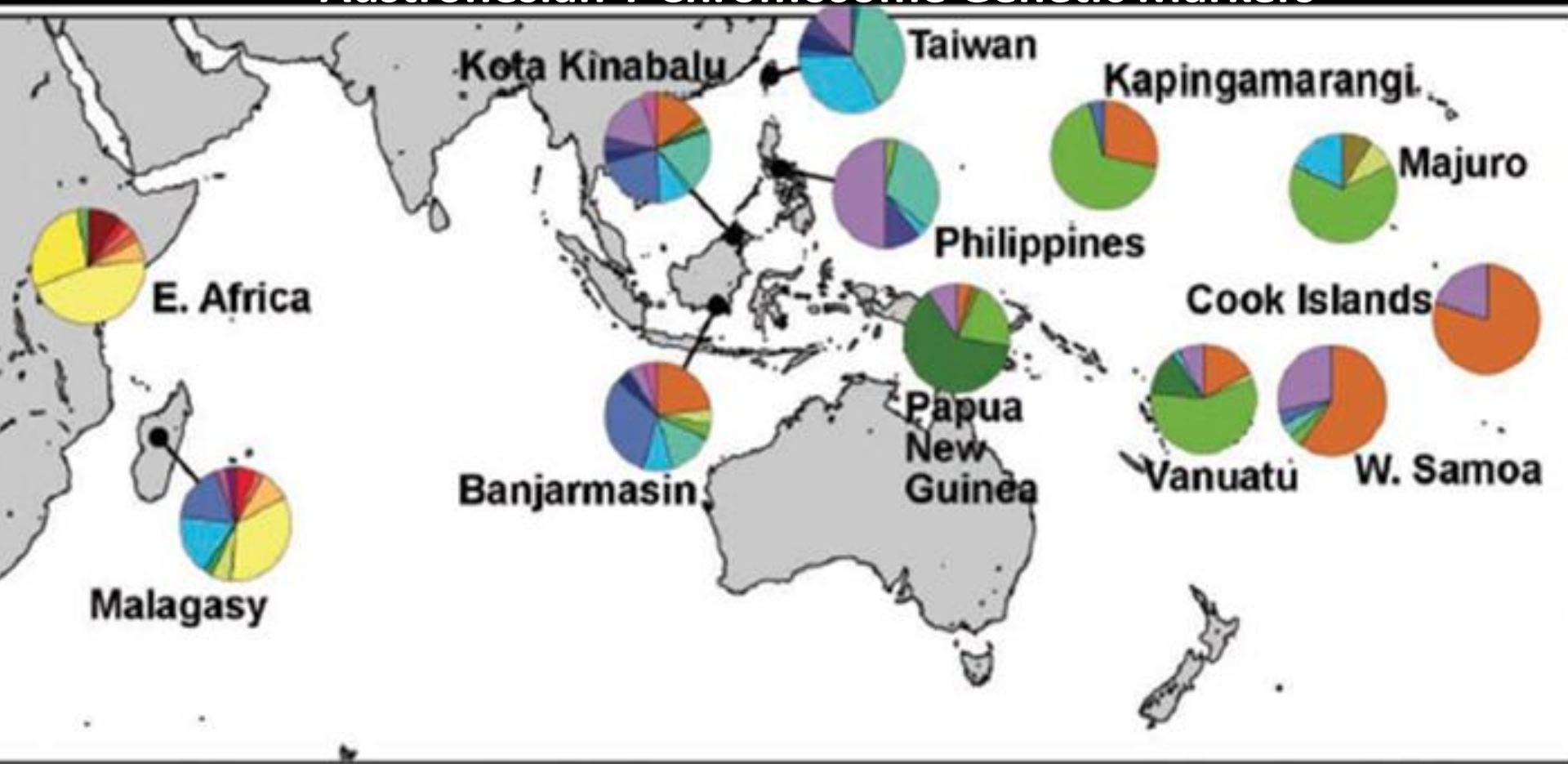


But the Austronesians seem - to the eye at least – to be completely unrelated to one another. The Polynesians in Rapa Nui, Hawaii and New Zealand all have a similar appearance. The Melanesians, like the Fijian warrior in this picture, look entirely different from the Polynesians.
But indigenous Taiwanese look even more different, and the lady from the island of Borneo looks – not surprisingly – southeast Asian. And finally, in Madagascar, most people look like Africans.
What ties them together most clearly is just the similarity of their languages.



When you examine the maternal (mitochondrial) and paternal (Y-chromosome) DNA characteristics of Pacific Islanders, you see clear evidence of Founder Effects, where a certain subset of people on one island sailed east to settle a new island, and became the majority of its inhabitants. But notice this: the genetic signal for Polynesians does not show up today in The Philippines or Taiwan. It can be traced back as far as Wallacea only. If it's true that Austronesian languages originated in Taiwan, it's a clear example of how genes and languages don't necessarily travel together through time.

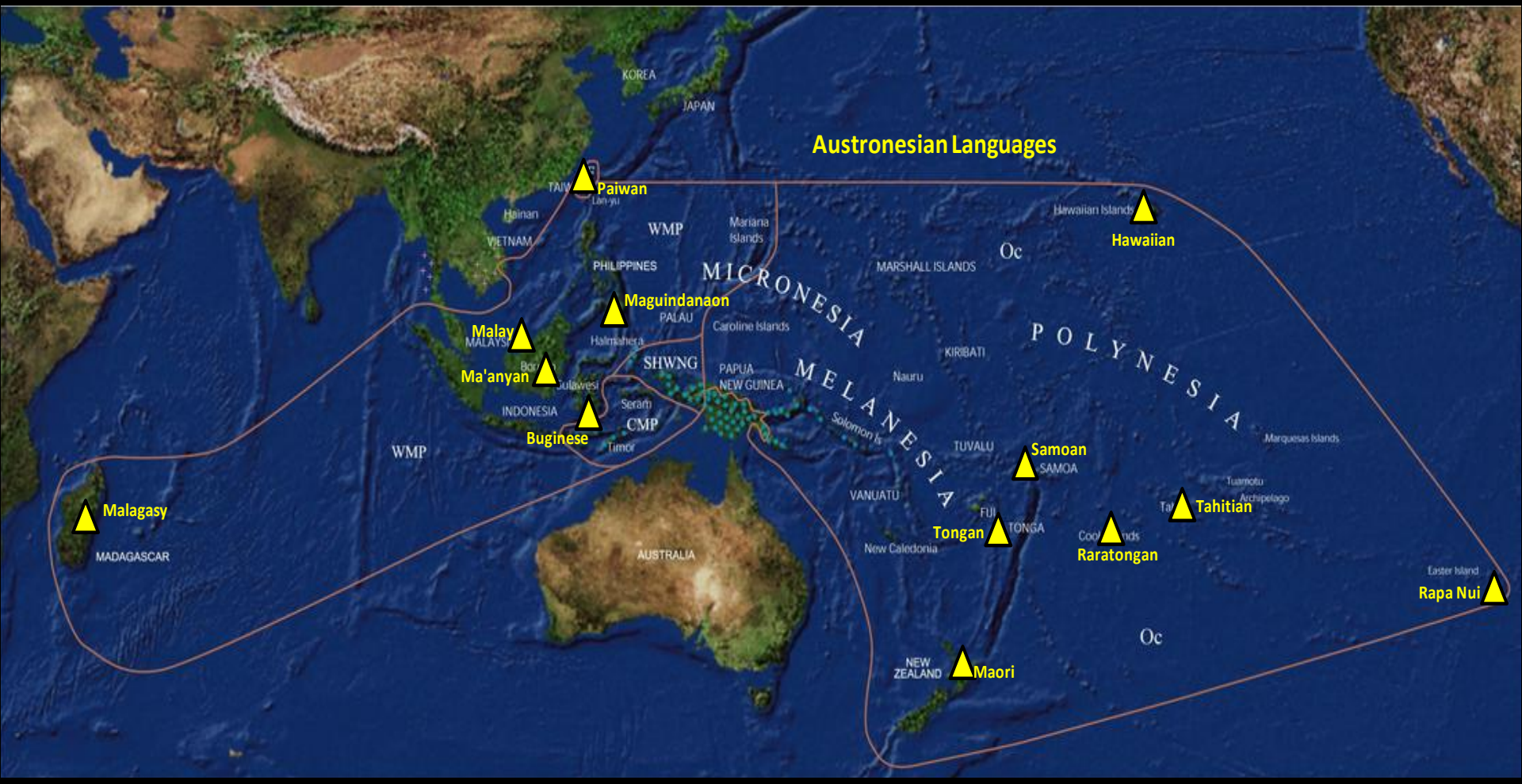
Austronesian Y Chromosome Genetic Markers



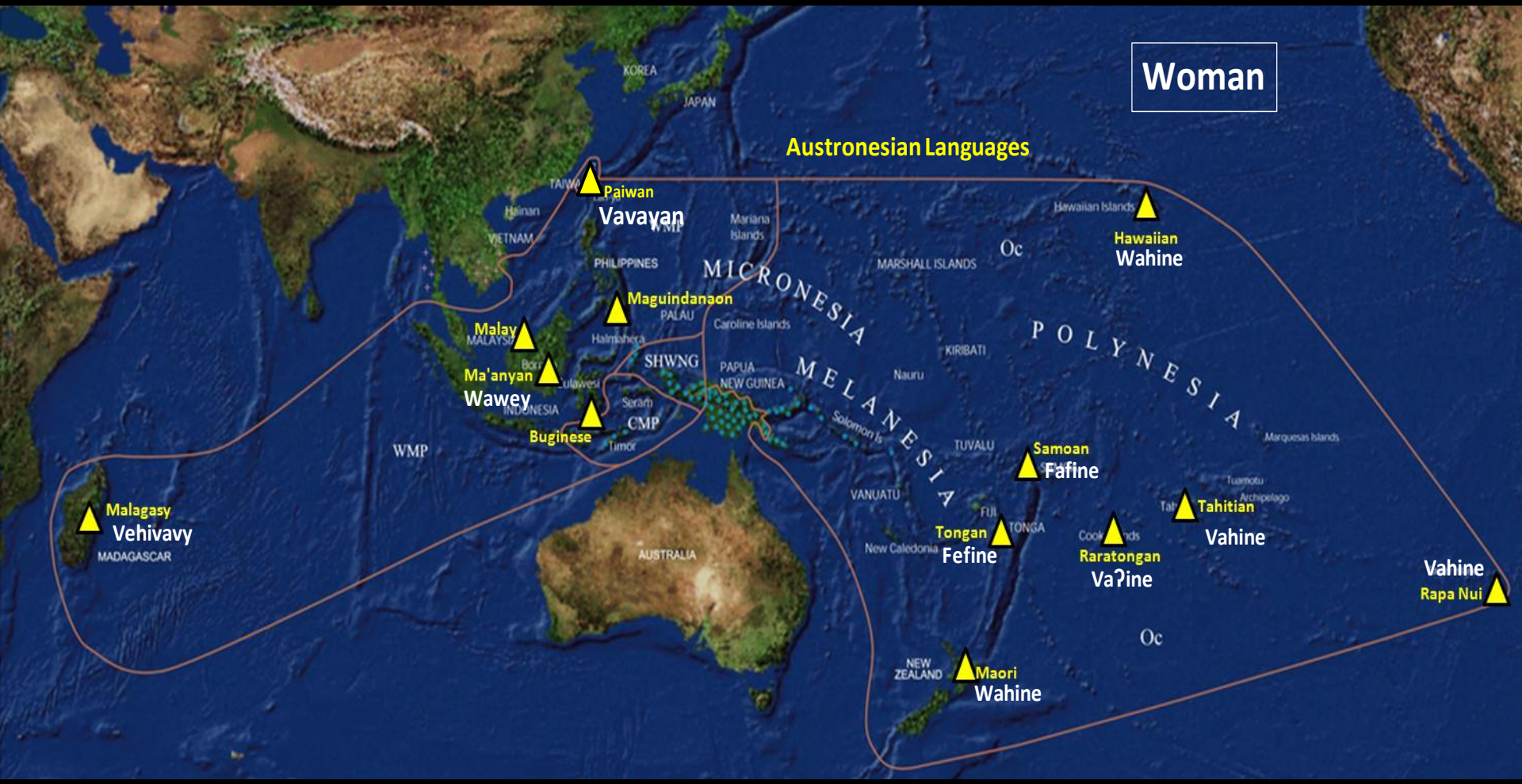
This illustration of the percentages of certain Y-Chromosome Haplotypes found across much of Austronesia makes it clear why the Polynesians look different from the Aboriginal Formosans. Only haplotype O3 (purple) found in Taiwan has survived the trek to Samoa and the Cook Islands. But these Polynesians are dominated by Haplogroup C (orange) which is not present today in Taiwan or the Philippines, but does appear in Borneo. On the other side of the Austronesian world, we have the Malagasy, who have retained Haplogroup O1b (blue-green) from Taiwan and Borneo along with Haplogroup O2a (blue) from Borneo. (More about the Malagasy later).

So the genetics of these people at the antipodes of Austronesia seem quite unrelated, though their languages are not. How do we know their languages actually ARE related?

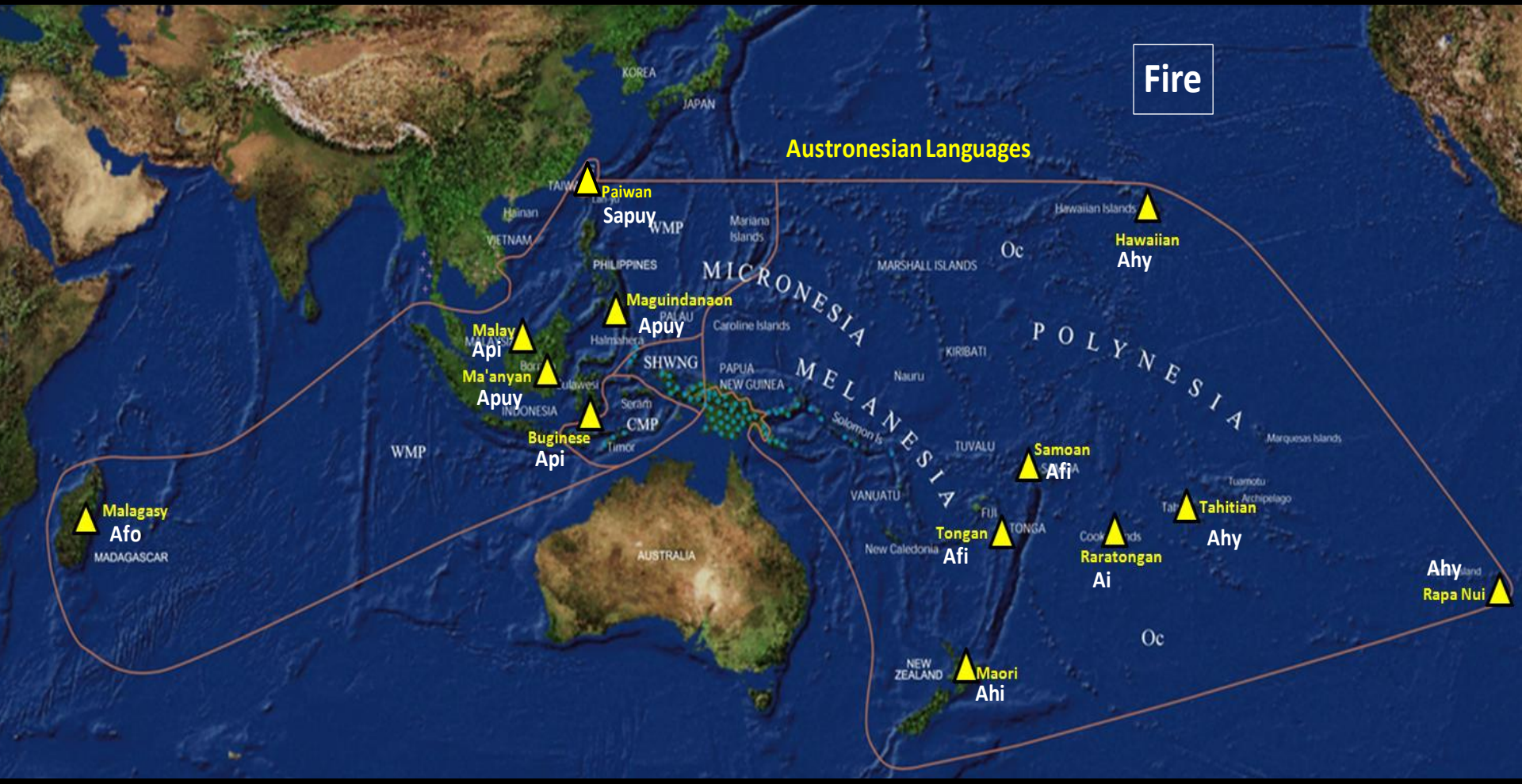
Some Austronesian Languages



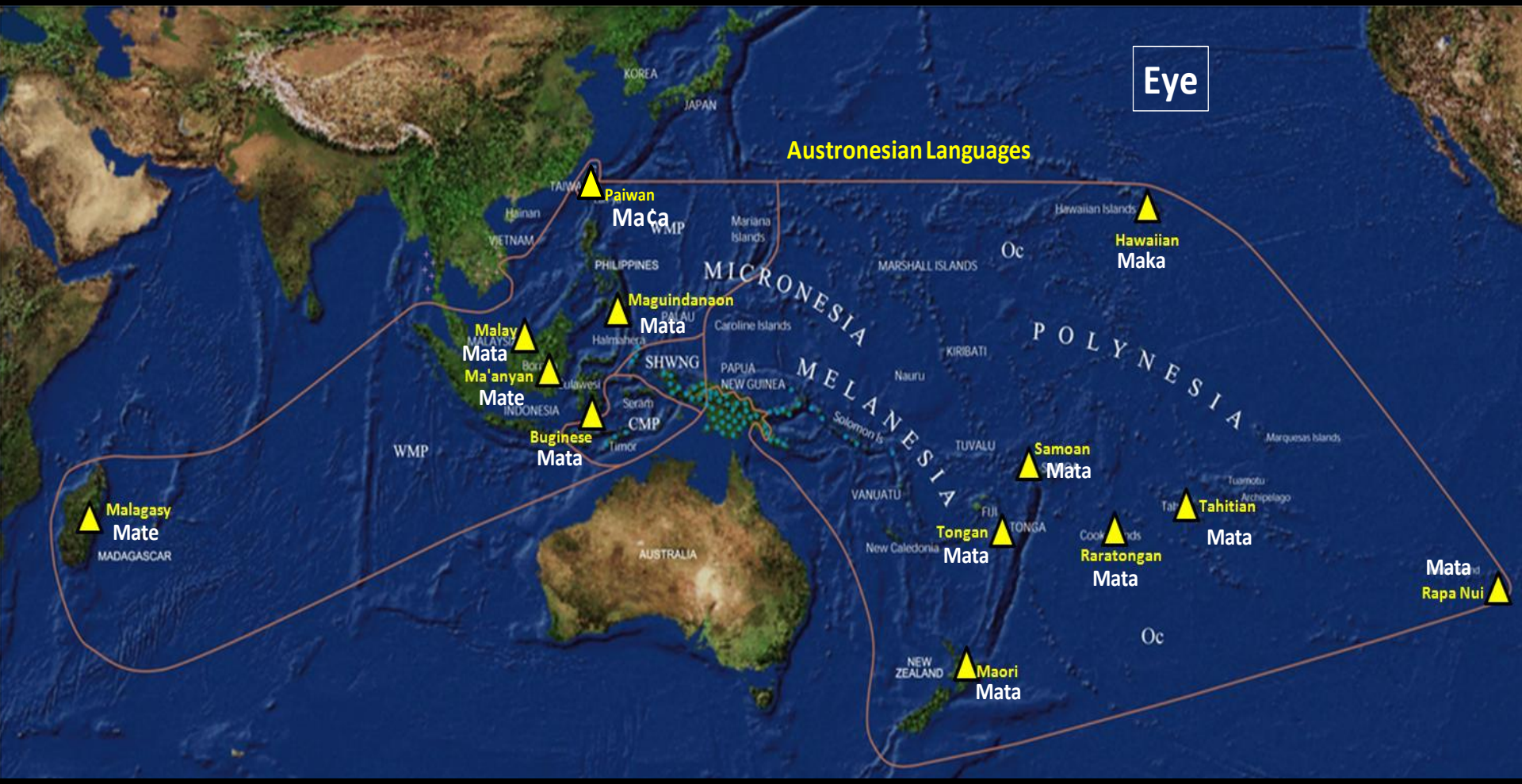
We can tell this by examining the characteristics of some Austronesian languages for which we have data. Linguists have catalogued and studied these languages in great technical detail for many, many years. But even as amateurs, we can examine the number of similar words with the same meaning in various languages, so-called cognates. I'm going to refer to this process as "Lexical Analysis".



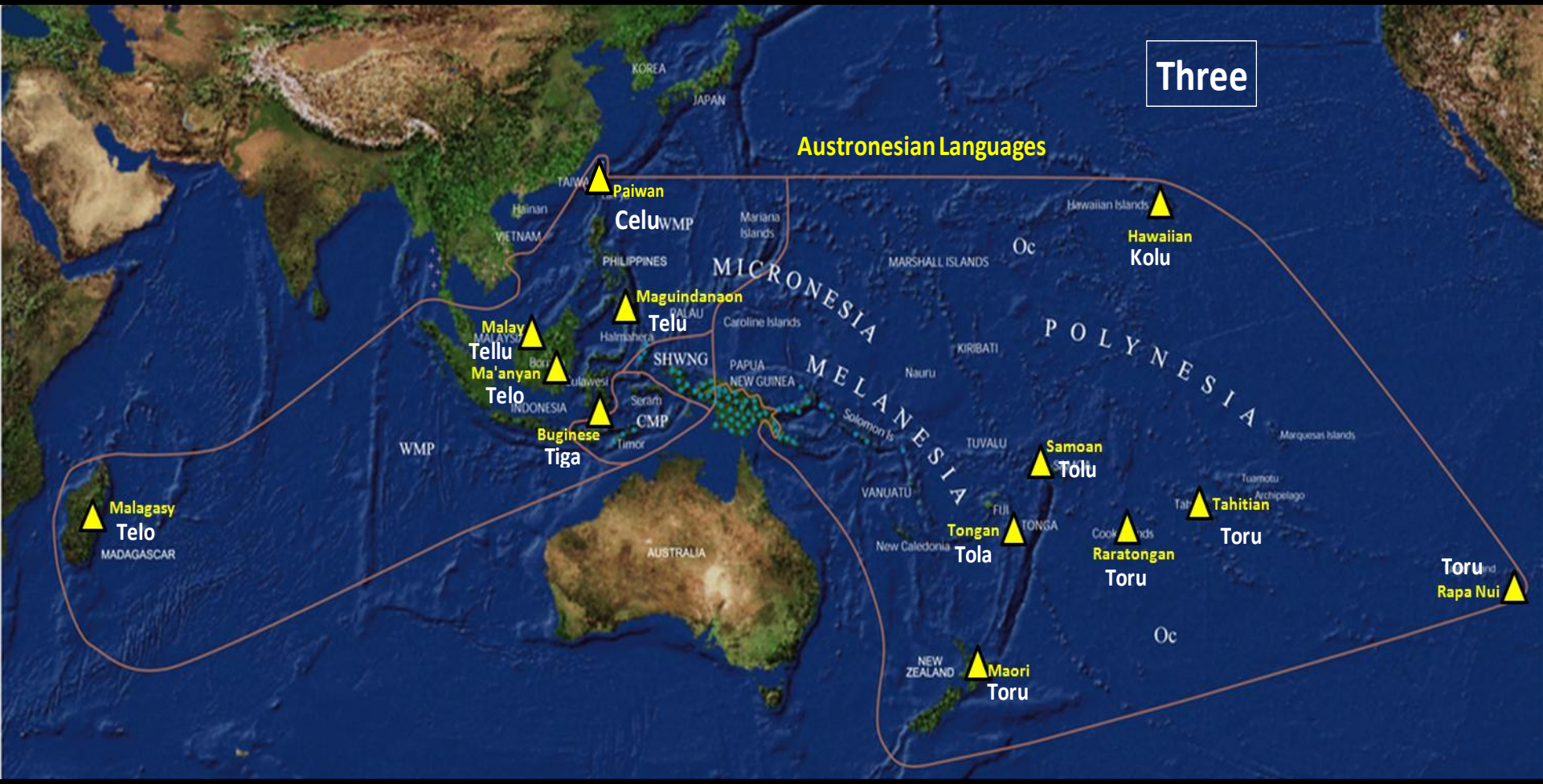
The word for "Woman" seems to be somewhat similar across the realm.



The word for “Fire” seems fairly universal, but with some sound changes.



“Eye” is pretty much the same everywhere you go.



And the number “Three” is pretty uniform across the oceans.

Western Malayo-Polynesian Languages

	Formosa →	Philipines →	Sulawesi →	N. Borneo →	S. Borneo →	S. Borneo →	Madagascar →
English	**Paiwan**	**Maguindanaon**	**Buginese**	**Malay**	**Ngaju**	**Ma'anyan**	**Malagasy**
sky	/kə-ḷəvəḷəv-an/	/langit/	/langit/	/langi/	/langit/	/langit/	/lanitra/
woman	/vavayan/	/babay/	/makunrai/	/perempuan/	/bawi/	/wawey/	/vehivavy/
fire	/sapui/	/apuy/	/api/	/api/	/apui/	/apuy/	/afo/
tongue	/səma/	/dila/	/jidah/	/lila/	/jela/	/lela/	/lela/
eye	/maȼa/	/mata/	/mata/	/mata/	/mate/	/mate/	/mate/
stone		/oatu/	/batu/	/batu/	/batu/	/watu/	/vatu/
One	/ita/	/isa/	/satu/	/sedi/	/ije/	/isa/	/isa/ or /iray/
Two	/usa/	/duoa/	/dua/	/dua/	/due/	/rueh/	/roa/
Three	/cəḷu/	/telu/	/tiga/	/tellu/	/telu/	/telo/	/telo/
Four	/səpac/	/pat/	/empat/	/epa/	/epat/	/efatro/	/efatra/

Here is a table of words selected for their uniformity across the western Malayo-Polynesian languages.
The words are not identical, they are *cognates*, words similar enough to one another to imply a causal relationship in their evolution.
As the languages moved from Taiwan to Madagascar, half of the words on this short list were essentially preserved.
Almost all were preserved from the Philippines onward.

Cognate Words in Polynesian Languages

Direction	Polynesia →	Polynesia →	Polynesia →	Polynesia →	Polynesia ↓	Polynesia ↑	Polynesia →
English	**Tongan**	**Sāmoan**	**Rarotongan**	**Tahitian**	**Māori**	**Hawai'ian**	**Rapanui**
sky	/laŋi/	/laŋi/	/raŋi/	/raʔi/	/raŋi/	/lani/	/raŋi/
woman	/fefine/	/fafine/	/vaʔine/	/vahine/	/wahine/	/wahine/	/vahine/
fire	/afi/	/afi/	/aì/	/auahi/	/ahi/	/ahi/	/ahi/
tongue	/'elelo/	/alelo/	/arero/	/arero/	/arero/	/alelo/	/arero/
eye	/mata/	/mata/	/mata/	/mata/	/mata/	/maka/	/mata/
stone	/maka/	/ma'a/	/pooàtu/	/'oofa'i/	/pōhatu/	/pohaku/	/ma'ea/
One	/taha/	/tasi/	/tai/	/tahi/	/tahi/	/kahi/	/tahi/
Two	/ua/	/iua/	/rua/	/piti/	/rua/	/lua/	/rua/
Three	/tola/	/tolu/	/toru/	/toru/	/toru/	/kolu/	/toru/
Four	/fa/	/fa/	/a/	/maha/	/wha/	/ha/	/ha/

And the batting average is even better across Polynesia, all the way from Tonga to Rapanui.
But these are "Cherry-picked" words. I like to call them "Golden words".
But what if we take a larger sample? Will some degree of relationship hold up?
And if so, can we make a measurement of how closely these languages are actually related?

A *Swadesh List* of 200 Selected Words

Serva's 200-Word Swadesh List (English):

1	all	21	cloud	41	far	61	good	81	in	101	narrow	121	root	141	smell	161	that	181	water
2	and	22	cold	42	fat	62	grass	82	kill	102	near	122	rope	142	smoke	162	there	182	we
3	animal	23	come	43	father	63	green	83	know	103	neck	123	rotten	143	smooth	163	they	183	wet
4	ashes	24	count	44	fear	64	guts	84	lake	104	new	124	rub	144	snake	164	thick	184	what
5	at	25	cut	45	feather	65	hair	85	laugh	105	night	125	salt	145	snow	165	thin	185	when
6	back	26	day	46	few	66	hand	86	leaf	106	nose	126	sand	146	some	166	think	186	where
7	bad	27	die	47	fight	67	he	87	left	107	not	127	say	147	spit	167	this	187	white
8	bark	28	dig	48	fire	68	head	88	leg	108	old	128	scratch	148	split	168	you (s)	188	who
9	because	29	dirty	49	fish	69	hear	89	lie	109	one	129	sea	149	squeeze	169	three	189	wide
10	belly	30	dog	50	five	70	heart	90	live	110	other	130	see	150	stab	170	throw	190	wife
11	big	31	drink	51	float	71	heavy	91	liver	111	person	131	seed	151	stand	171	tie	191	wind
12	bird	32	dry	52	flow	72	here	92	long	112	play	132	sew	152	star	172	tongue	192	wing
13	bite	33	dull	53	flower	73	hit	93	louse	113	pull	133	sharp	153	stick	173	tooth	193	wipe
14	black	34	dust	54	fly	74	hold	94	man	114	push	134	short	154	stone	174	tree	194	with
15	blood	35	ear	55	fog	75	how	95	many	115	rain	135	sing	155	straight	175	turn	195	woman
16	blow	36	earth	56	foot	76	hunt	96	meat	116	red	136	sit	156	suck	176	two	196	woods
17	bone	37	eat	57	four	77	husband	97	mother	117	right	137	skin	157	sun	177	vomit	197	worm
18	breathe	38	egg	58	freeze	78	I	98	mountain	118	right (hand)	138	sky	158	swell	178	walk	198	you (pl)
19	burn	39	eye	59	fruit	79	ice	99	mouth	119	river	139	sleep	159	swim	179	warm	199	year
20	child	40	fall	60	give	80	if	100	name	120	road	140	small	160	tail	180	wash	200	yellow

This is what's called a Swadesh List. It's named after Lexical Analysis pioneer Morris Swadesh, who first generated a similar list in 1952. This list attempts to sample a language using words that are thought to be universally understood in all languages. This 200-word version contains a wide range of words that have been found and understood in many languages. It is a popular tool for measuring the similarity between related languages. There are a lot of problems with these lists that I don't have time to discuss in this lecture. But basically, they work pretty well. So for now, let's use these lists – carefully – to see what we can learn about the Austronesian languages.

Lexical Distance: A way of Specifying Language Relationships

Lexical Distance =

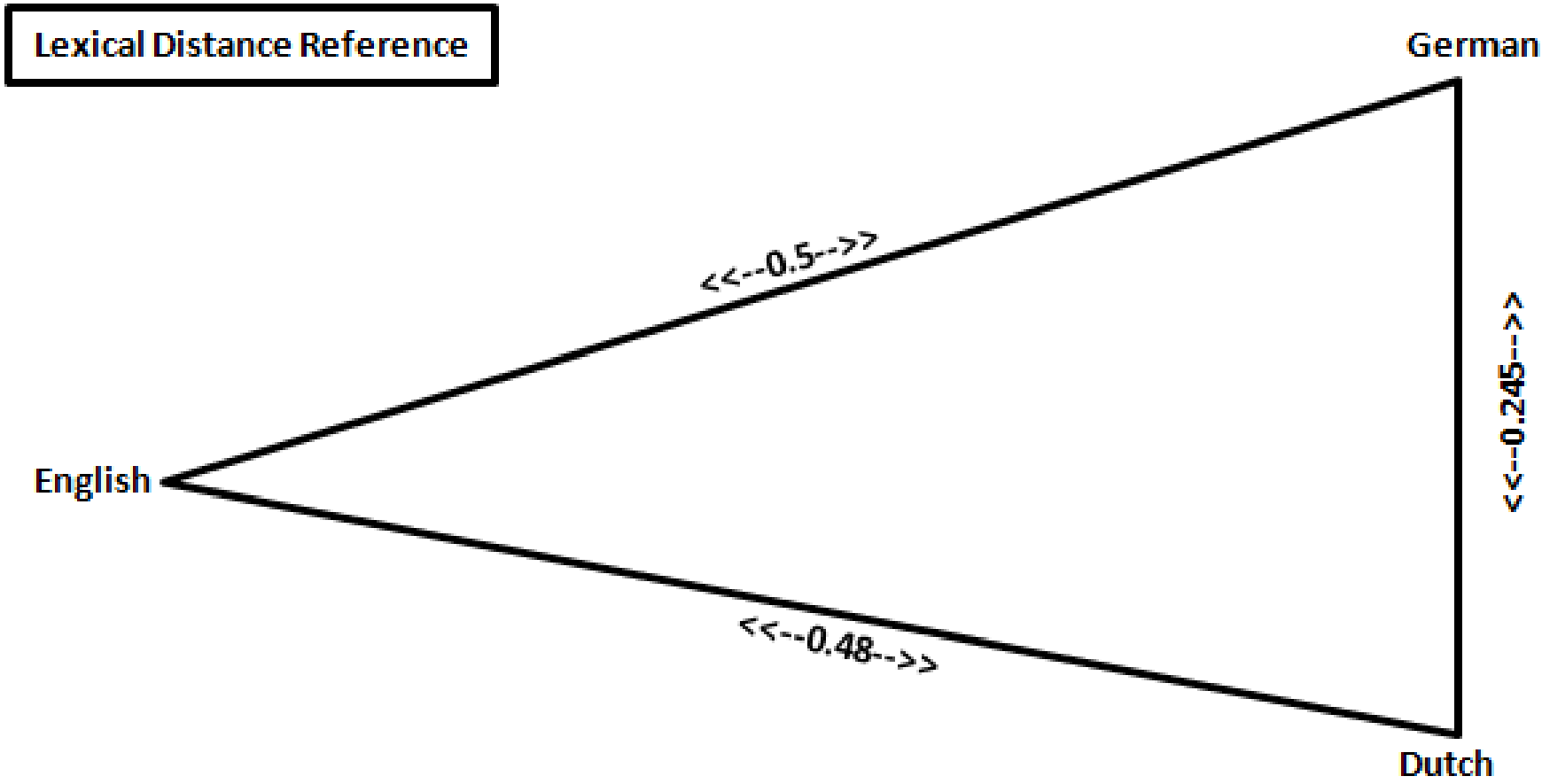
$$1 - \frac{\text{Number of Cognate Differences Between 2 Languages or Dialects}}{\text{Number of Cognate Sets in All Languages Being Compared}}$$

e.g. 1- (50 cognates / 200 sets) => Lexical Distance = 1 – 0.25 = 0.75

Lexical Distance = 1 means 0% Cognacy of Words from Swadesh List

Lexical Distance = 0 means 100% Cognacy of Words from Swadesh List

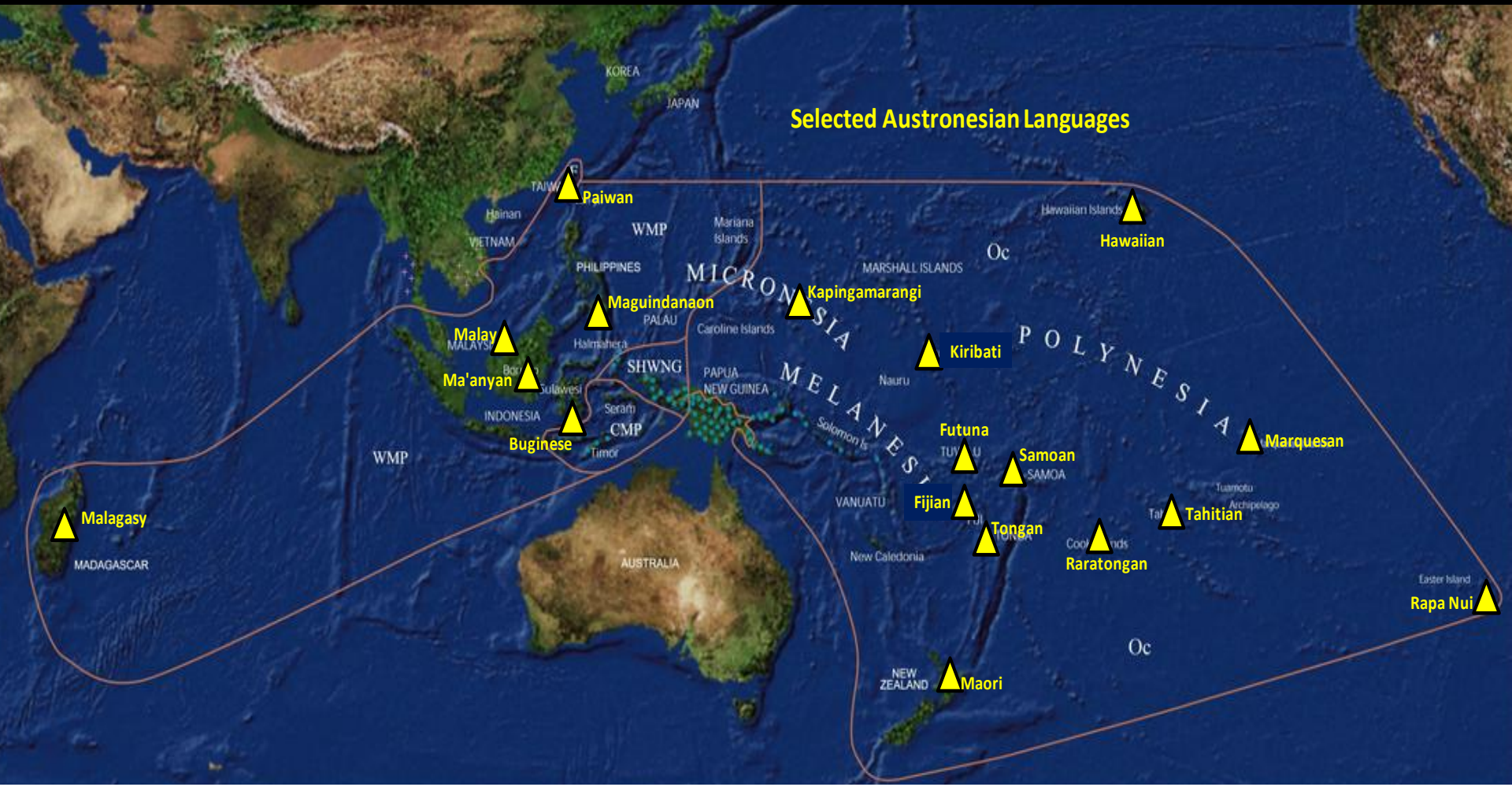
But first, let's take a moment to define a term: Lexical Distance. By counting the number of cognates between all the words on Swadesh Lists for two languages, we can derive a useful number called "Lexical Distance" between them. For example, if two dialects of languages share 50 words in a list of 200, the Lexical Distance is 1 minus 50 over 200 equals 0.75. Complete matching is a lexical distance of zero; no matching is a lexical distance of one.
Defining language differences as distances is useful because it is both intuitive and amenable to technical analysis, as we will see later on.



Lexical Distance = 1 means 0% Cognacy of Words from Swadesh List

To get a feel for what lexical distance means to someone listening to a language, let's use some familiar languages. About half the English words on the list are cognate with German or Dutch, leading to lexical distance of about 0.5.
About three quarters of Dutch words are cognate with German, leading to a lexical distance of about 0.25.
We'll use this diagram as a yardstick for visualizing the Lexical Distances between other languages.
Realize that none of these languages are mutually intelligible. There are a lot more words and some important differences in syntax. But this Lexical Distance number can still be a rough indicator of relatedness.
Notice that you can't simply add two lexical distances to get the resultant distance to a third language.

Selected Austronesian Languages



Let's apply Lexical Distance Analysis to some Austronesian words for which I have data.

Austronesian

Basic Vocabulary Database

Language: Samoan

Entries:

ID:	Word:	Item:	Annotation:	Cognacy:	Loan:
1	hand	lima		1	
2	left	tau-aŋavale		31	
3	right	tau-matau		5	
4	leg/foot	vae		1	
5	to walk	savali			
5	to walk	saele		6	
6	road/path	ala		1	
7	to come	sau		8	
7	to come	ō mai		1	
8	to turn	fāliu		2	
9	to swim	ʔaʔau		5	
10	dirty	ʔeleʔelea		10	
11	dust	pl̩efu		1	
12	skin	paʔu		27	
12	skin	ʻili		1	
12	skin	ʻiliola			
13	back	tua		12	
14	belly	manava		9	
15	bone	ivi		2, 5	
16	intestines	ŋāʔau		12	
17	liver	ate		1	
18	breast	susu		1	
19	shoulder	tauʔau		227	
20	to know, be knowledgeable	ilo-a		13	
21	to think	māfaufau			
21	to think	manatu			
22	to fear	mataʔu		1	
22	to fear	fefe			

Today, a remarkable amount of data is publicly-available on Austronesian languages.
The results of decades of field work my many researchers has been compiled by Professor Richard Blust and others at the University of Auckland and University of Hawaii.
In this example, we see data on the Samoan language as spoken today.

Data from ABVB Organized into a Table for Language Comparison

376 ENGLISH	228 TONGAN	303 SAMOAN	359 MARQUESAN	242 MODERN TAHITIAN	204 18th Century TAHITIAN	242 HAWAIIAN	326 RAPANUI	350 RARATONGAN	226 MAORI
1 hand	1 nima	1 lima	1 ìma	1 rima	1 rima	1 lima	1 rima	1 kapu 1 rima	1 ringa(ringa)
2 left 2 left	2 hema	2 tau-aʔavale	2 akeake 'ima a'ea'e	2 'aui	2 'aio 2 aruri	2 hema	2 maui	2 kauī 2 kauvī 2 mauī	2 mauii
3 right 3 right	3 mataʔu	3 tau-matau	3 atamai 'ima oko	3 'atau	3 'atau	3 'ākau	3 mata'u	3 katau	3 katau 3 matau
4 leg/foot	4 vaʔe	4 vae	4 vae	4 'aavae	4 'aavae	4 wāwae	4 ba'e 4 va'e	4 tapuae 4 vae	4 waewae
5 to walk	5 ʔalu	5 savali 5 saele	5 hèe 5	5 haere 5 ori	5 haere	5 hele	5 haere	5 àere	5 haere-a-waewae
6 road/path	6 hala	6 ala	6 vaà/nui	6 poroomu 6 puruumu 6 'ee'a 6 ara	6 ara	6 ala	6 ara	6 ara	6 ara
7 to come 7 to come	7 haʔu 7 ō mai	7 sau 7 o mai	7 tihe 7 he'e mai	7 haere mai	7 haere mai	7 [hele] mai	7 tu'u	7 tātā'i	7 haere mai
8 to turn word inf	8 tāafeafe liliu	8 faliu	8 kavií	8 tiipu'u		8 wili 8 huli	8 teka 8 taviri	8 'akapa'e 8 'iki'iki 8 pa'e	8 huri
9 to swim	9 kakau	9 ʔaʔau	9 kau	9 'au	9 'au	9 'au	9 kau	9 kau	9 kauhoe
10 dirty 10 dirty	10 ʔuli	10 ʔeleʔelea	10 paàpaà hava	10 repo	10 repo 10 repoa	10 lepo	10 haba/haba 10 hava 10 one	10 repo 10 àvaàva	10 paru
11 dust	11 efu	11 p|efu	11 'epo	11 hu'ahuaa 11 repo puehu	11 one 11 repo	11 'ehu 11 'e'a	11 huŋa 11 ŋa/rahu/ 11 huga 11 garahu 11 pugaehu	11 puèu	11 pu|ehu 11 puehu
12 skin	12 kili	12 paʔu 12 'ili 12 'iliola	12 kiì 12 12	12 'iri	12 'iri	12 'ili	12 kiri	12 kiri	12 kiri

I combined their data for major Polynesian words into a table for word-by-word comparison. Notice that the entries are multi-valued in many cases, because the researchers obtained synonyms for many of the words.

Data from ABVB Organized into Cognate Sets

ENGLISH	42 KIRIBATI	126 KAPINGAMARAN	61 FIJIAN	137 FUTUNA-ANIWA	134 TONGAN	193 SAMOAN	204 MARQUESA	MODERN 176 TAHITIAN	18th Century 207 TAHITIAN	184 HAWAIIAN	189 RAPANUI	232 RARATONGA	184 MAORI
hand		1 rima		1 rima	1 nima	1 lima	1 ìma	1 rima	1 rima	1 lima	1 rima	1 rima	1 ringa(ring
left			2 i mawī	2 masui							2 maui	2 mauī	2 mauii
right		3 tau	3 i matau	3 matau	3 mataʔu	3 tau-matau		3 'atau	3 'atau	3 'ākau	3 mata'u	3 katau	3 katau
leg/foot	4 wáe	4 vae		4 vae	4 vaʔe	4 vae	4 vae	4 'aavae	4 'aavae	4 wāwae	4 va'e	4 vae	4 waewae
to walk		5 haere						5 haere	5 haere	5 hele	5 haere	5 àere	5 haere-a-w
road/path		6 ala	6 sala		6 hala	6 ala		6 ara	6 ara	6 ala	6 ara	6 ara	6 ara
to come				7 lako mai	7 hmai [ṃai]	7 o mai	7 he'e mai	7 haere mai	7 haere mai	7 haere mai			7 haere ma
to turn										8 huli			8 huri
to swim		9 kau			9 kaukau	9 kakau	9 kau	9 'au	9 'au	9 'au	9 kau	9 kau	9 kauhoe
dirty								10 repo	10 repo	10 lepo		10 repo	
dirty							10 hava				10 hava		
dust					11 efu	11 pjefu		11 repo puehu	11 repo puehu	11 'ehu	11 pugaehu	11 puèu	11 pujehu
dust							11 'epo	11 repo puehu	11 repo				
skin		12 kiri	12 kuli-na	12 kiri	12 kili	12 'ili	12 kiî	12 'iri	12 'iri	12 'ili	12 kiri	12 kiri	12 kiri
back		13 dua		13 bua	13 tuʔa	13 tua	13 tua	12 tua	12 tue	12 kua	13 tu'a ivi	13 mokotua	13 tuaraa
belly							14 kopu	14 'oopuu	14 'oopuu	14 'ōpū	14 kopu	14 koopuu	
belly				14 manava		14 manava					14 manava	14 manava	
bone		15 iwi		15 ivi		15 ivi	15 ivi	15 ivi	15 ivi	15 iwi	15 ivi	15 ivi	15 iwi
intestines					16 ŋākau	16 ʔaʔau		16 'aa'au	16 'aa'au	16 na'au		16 ngaakau	16 wheekau
intestines							16 koekoe				16 kokoma		
liver	17 áto	17 ate	17 yate-na	17 ate	17 ʔate	17 ate	17 ate			17 ake	17 ate	17 ate	17 ate
breast		18 uu		18 u	18 huhu	18 susu	18 u		18 uu	18 ū	18 u'u	18 uu	18 uu
shoulder							19 pauhihi			19 po'ohiwi	19 kapuhivi	19 pakuìvi	19 pakihiwi
shoulder								19 tapono	19 tapono				
to know		20 iroa		20 iroa	20 ʔilo	20 ilo-a							
to know	20 atá							20 'ite	20 'ite	20 'ike		20 kite	
to think				21 manatunea		21 manatu	21 maákau	21 mana'o	21 mana'o	21 mana'o	21 mana'u		
to think												21 maàra	21 whakaaro
to fear		22 madagu		22 mataku		22 mataʔu	22 meta'u		22 mata'u	22 maka'u	22 mataku	22 mataku	22 mataku
to fear								22 ri'ari'a	22 ri'ari'a		22 riaria		
blood		23 toto		23 toto	23 toto	23 toto	23 toto	23 toto	23 toto	23 koko	23 toto	23 toto	23 toto
head							24 upoko	24 upo'o	24 upo'o	24 po'o	24 puoko	24 upoko	24 upoko
head			24 ulu-na	24 uru	24 ʔulu	24 ulu	24 u'u						
neck								25 'a'ii	25 'a'ii	25 'ā'ī			
neck							25 kaakii					25 kakii	
neck		25 ua				25 ua							25 ua
hair		26 hulu			26 louʔulu			26 rouru	26 rouru		26 huru/huru	26 rauru	

Here is a portion of those cognate sets. Horizontal rows highlighted in yellow are sets of cognate words, sometimes more than one word per meaning. For example, the meaning "dirty" (no. 10) has two cognate sets: "repo" and "hava". The flow of words can be visualized. "Lima" (no. 1) survives with only slight modifications throughout Eastern Polynesia. "Hava" (10) and "koeloe/kokoma" (16) are unique to Marquesan and Rapanui. "Tapono" is found only in 18th century and modern Tahitian. "uu" and similar forms signify "breast" everywhere except in modern Tahitian, where it has been re[placed by the word "titti", probably borrowed from English sailors]. Continued

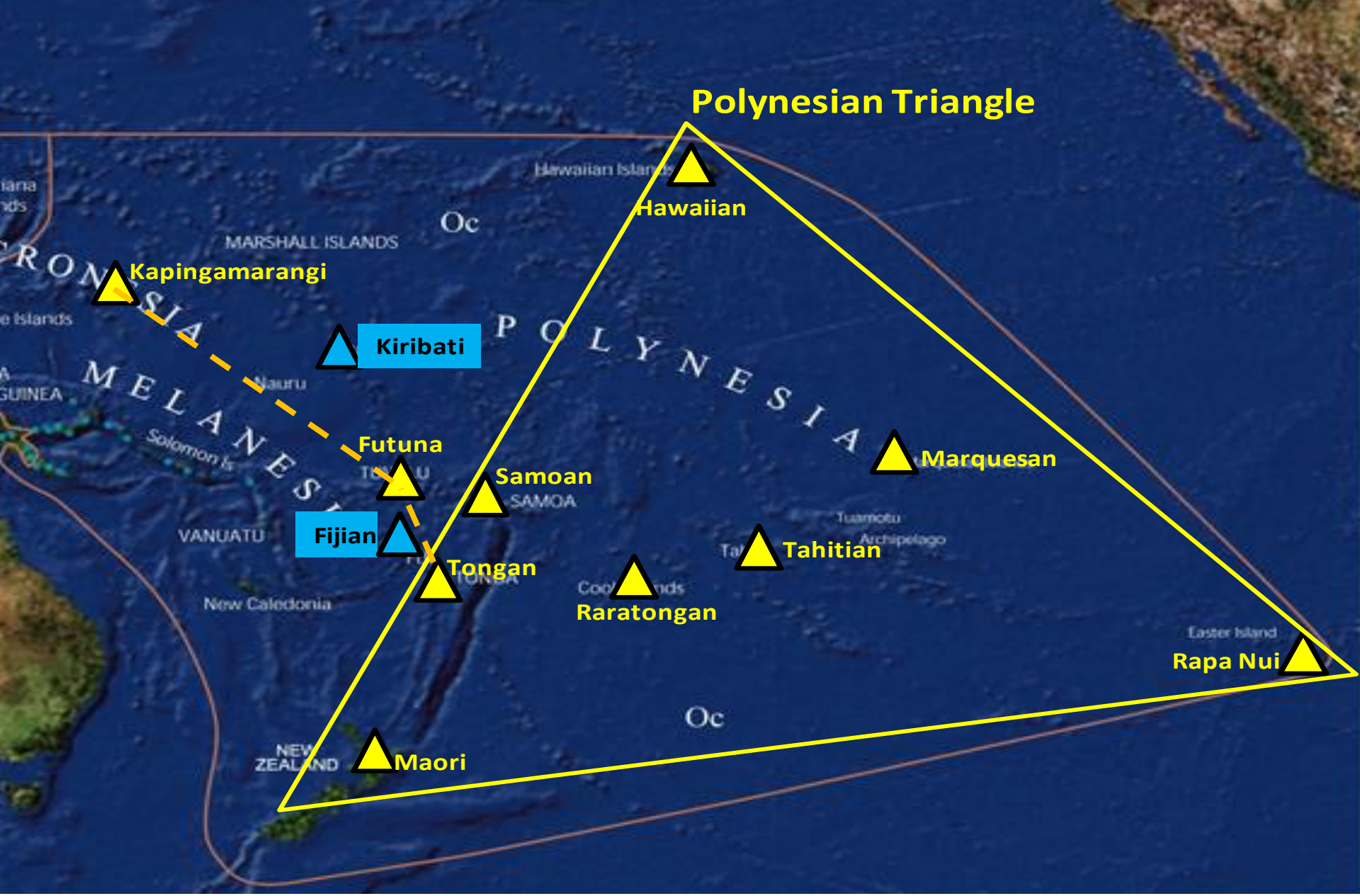
The Modern Tahitian word "mana'o" is cognate to several other languages, but was missing from the 18th century data, even though it, or something very similar to it, had to have been present in the 18th century. So I added it to the table (21, in red) to further reconstruct the 18th century Tahitian language. By analyzing the cognate connections between these columns, we can calculate a measure of Lexical Distance between languages. And by comparing the forms of each cognate set's members across languages, we can follow the language evolution.

Lexical Distance Matrix for Selected Polynesian and Melanesian Languages

	KIRIBATI	KAPINGA	FIJIAN	FUTUNA	TONGAN	SAMOAN	MARQUESAN	TAHITIAN	18c TAHITI	HAWAIIAN	RAPANUI	RARATON	MAORI
KIRIBATI		0.93	0.94	0.91	0.92	0.90	0.91	0.95	0.92	0.92	0.92	0.91	0.92
KAPINGAMARANGI	0.93		0.88	0.75	0.76	0.71	0.73	0.78	0.74	0.75	0.74	0.70	0.74
FIJIAN	0.94	0.88		0.86	0.86	0.85	0.89	0.91	0.88	0.88	0.89	0.86	0.88
FUTUNA-ANIWA	0.91	0.75	0.86		0.75	0.68	0.71	0.79	0.73	0.75	0.72	0.68	0.72
TONGAN	0.92	0.76	0.86	0.75		0.66	0.75	0.78	0.73	0.76	0.73	0.71	0.74
SAMOAN	0.90	0.71	0.85	0.68	0.66		0.65	0.72	0.64	0.67	0.65	0.60	0.65
MARQUESAN	0.91	0.73	0.89	0.71	0.75	0.65		0.68	0.63	0.61	0.62	0.55	0.63
TAHITIAN	0.95	0.78	0.91	0.79	0.78	0.72	0.68		0.52	0.67	0.69	0.62	0.68
18c TAHITIAN	0.92	0.74	0.88	0.73	0.73	0.64	0.63	0.52		0.63	0.64	0.57	0.63
HAWAIIAN	0.92	0.75	0.88	0.75	0.76	0.67	0.61	0.67	0.63		0.63	0.61	0.63
RAPANUI	0.92	0.74	0.89	0.72	0.73	0.65	0.62	0.69	0.64	0.63		0.58	0.63
RARATONGAN	0.91	0.70	0.86	0.68	0.71	0.60	0.55	0.62	0.57	0.61	0.58		0.56
MAORI	0.92	0.74	0.88	0.72	0.74	0.65	0.63	0.68	0.63	0.63	0.63	0.56	

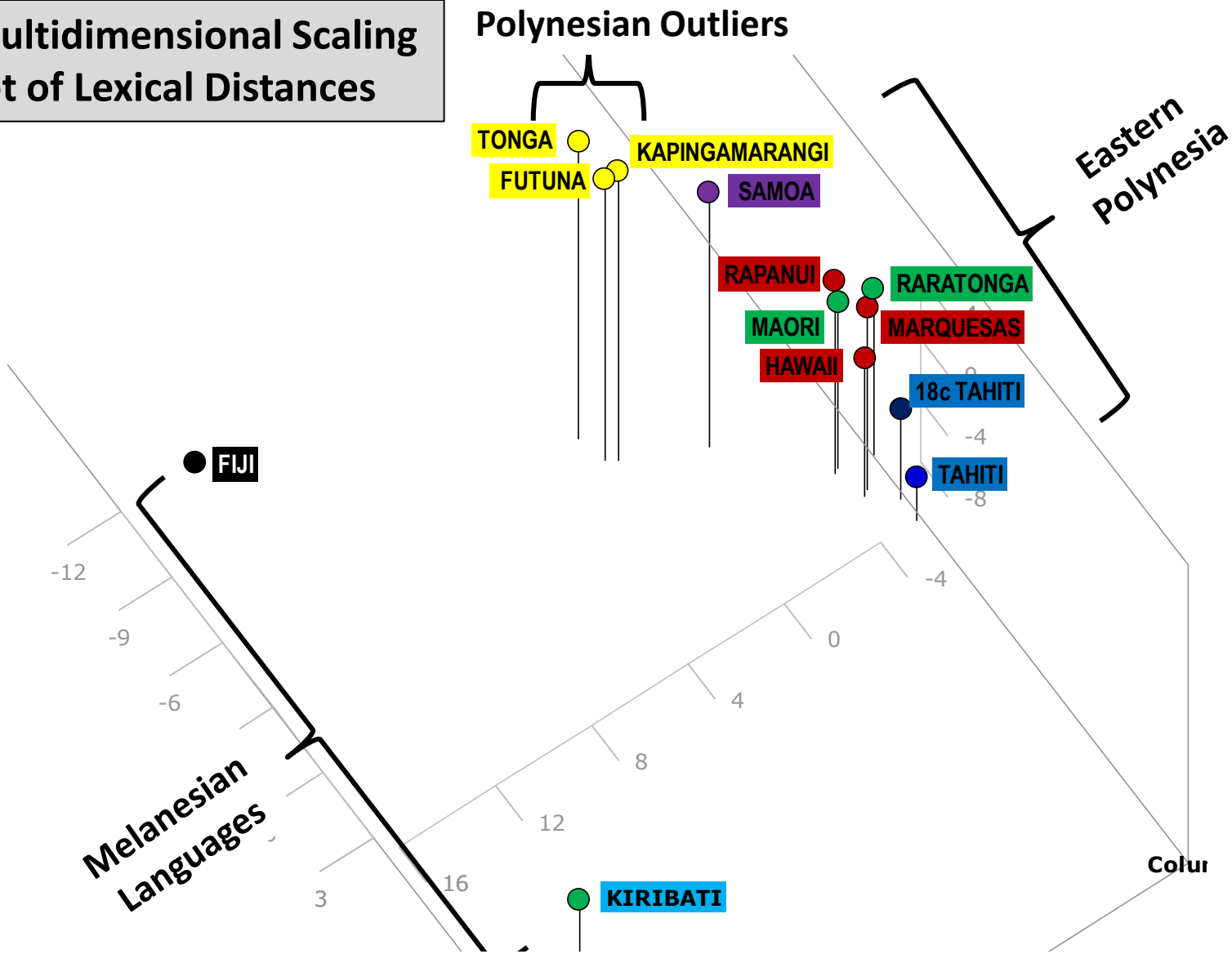
355 Cognate Sets

The lexical distances between languages computed from the preceding table of 13 languages and 206 meanings are organized in this 13 by 13 matrix.



These languages are found mostly within the so-called "Polynesian Triangle", but there are a couple of what are called "Polynesian Outliers" included (Futuna and Kapingamarangi). The languages of Fiji and Kiribati, which belong to the Melanesian family, are included for comparison.

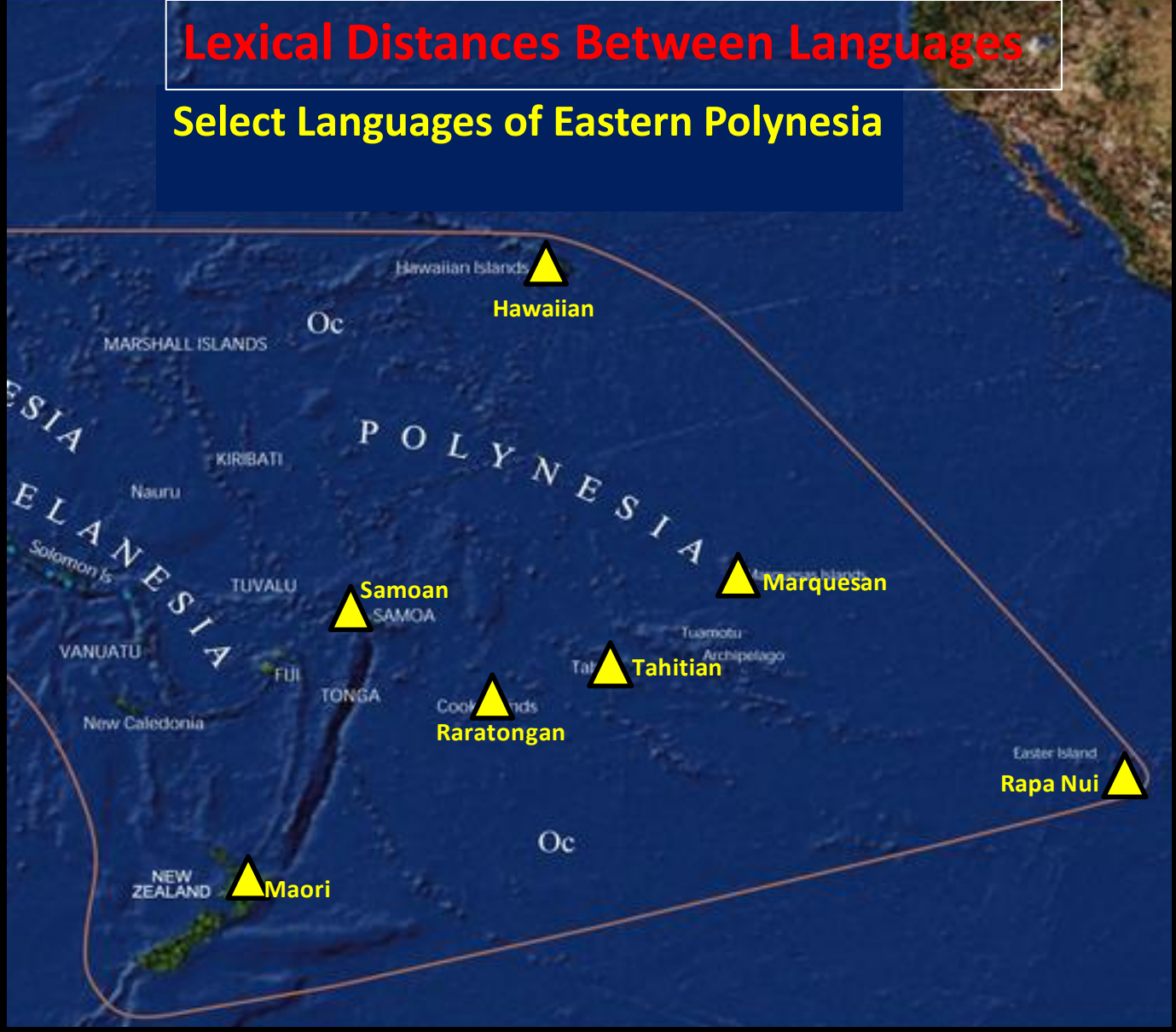
3D Multidimensional Scaling Plot of Lexical Distances



Our brains are not capable of visualizing all the data in the 13 full dimensions of the matrix. So we turn to a technique called Multidimensional Scaling (MDS), which allows us to view an approximate representation of this 13-dimensional distance matrix in three dimensions. The two Melanesian languages from Fiji and Kiribati stand far from the Polynesian languages. Notice that the Polynesian outliers Futuna and Kapingamarangi along with Tongan stand apart from the main group of eastern Polynesian languages. And Samoan is sort of halfway in between the outliers and the Eastern Polynesian languages.

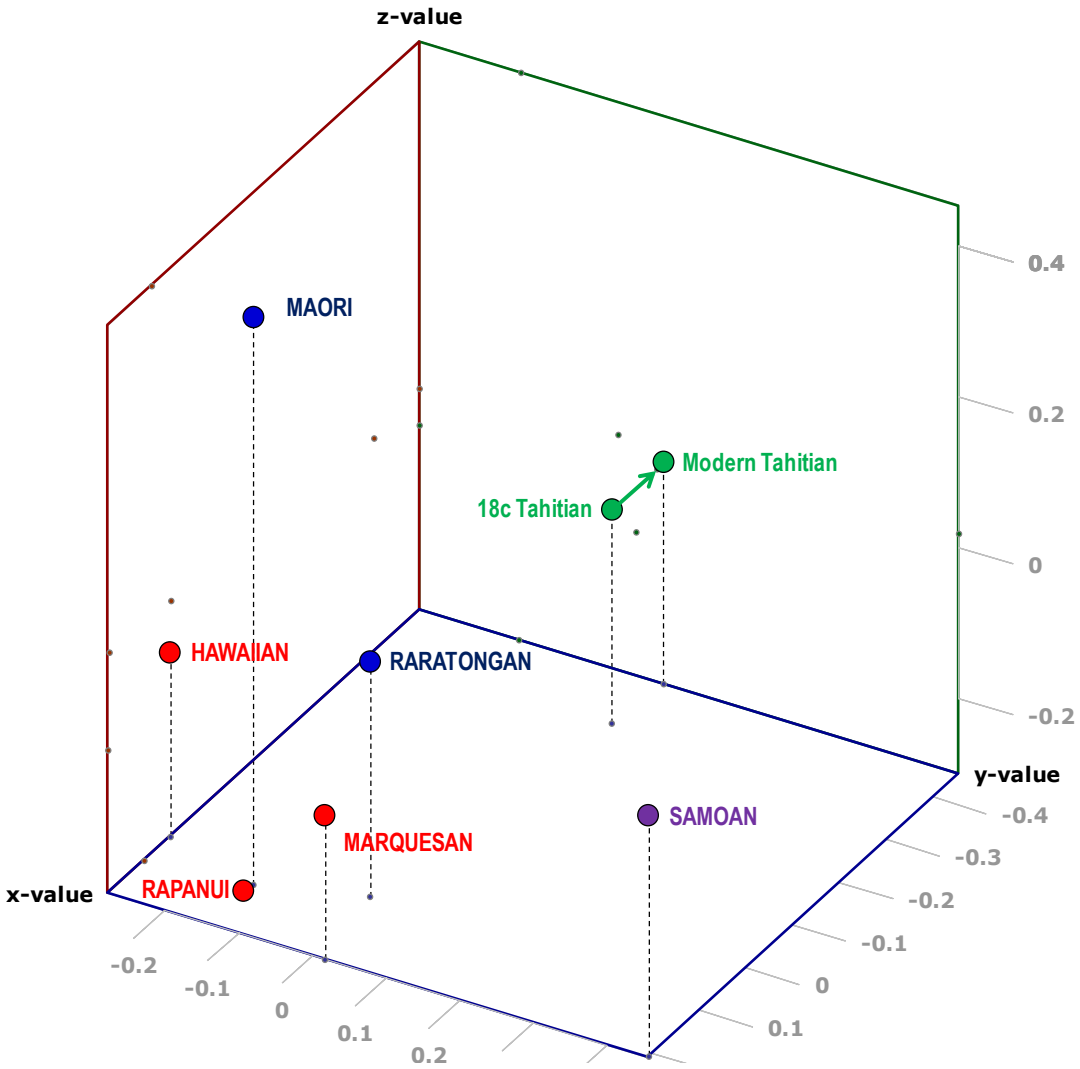
Lexical Distances Between Languages

Select Languages of Eastern Polynesia

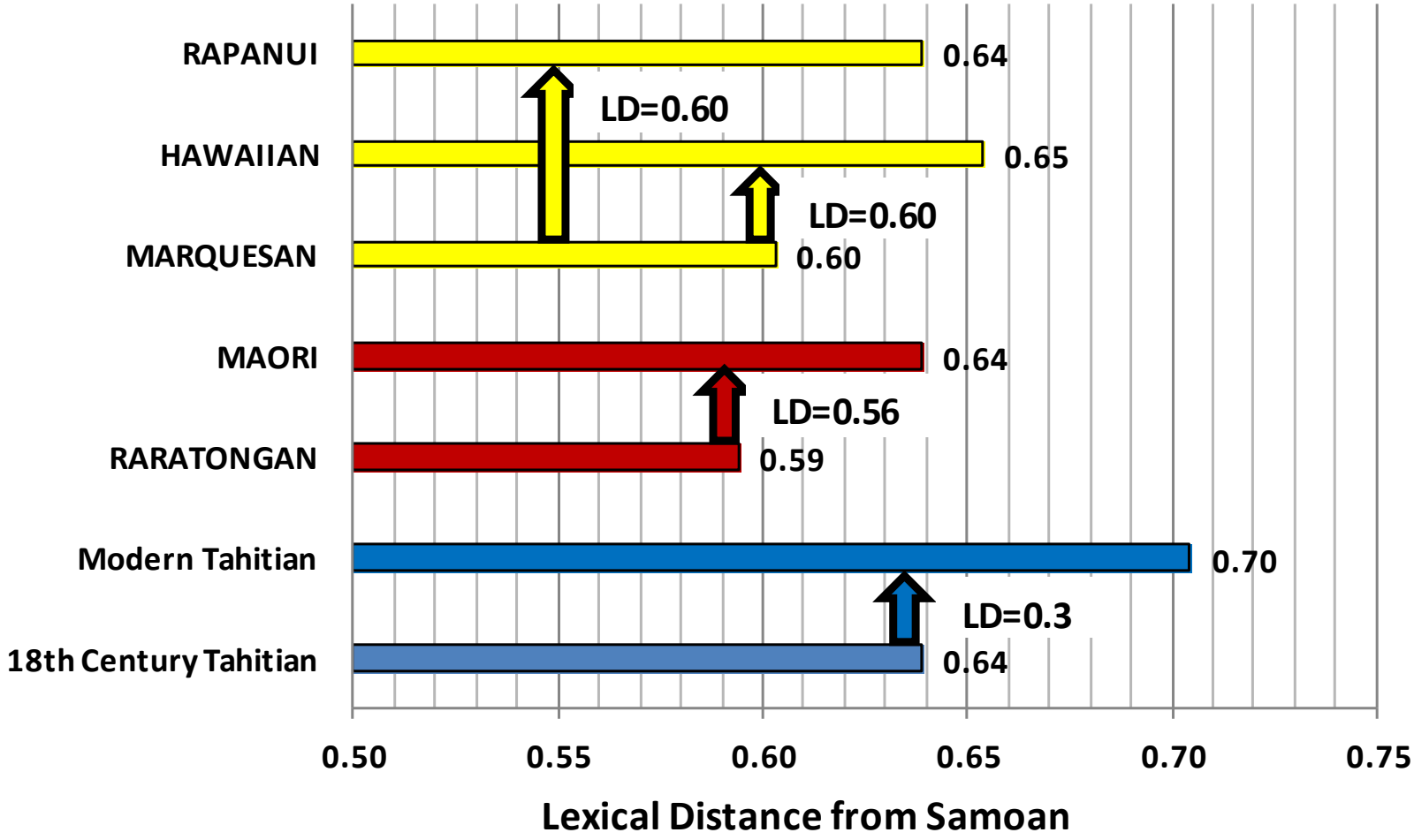


Let's now focus on this group of major Polynesian Languages. They are all fairly closely related. Average Lexical Distance between any language and any other in this region is just a little bit more than English to German. Can this tell us anything about the settlement of Polynesia?

Polynesian Language Relationships per Multi-dimensional Scaling of Lexical Distance Matrix



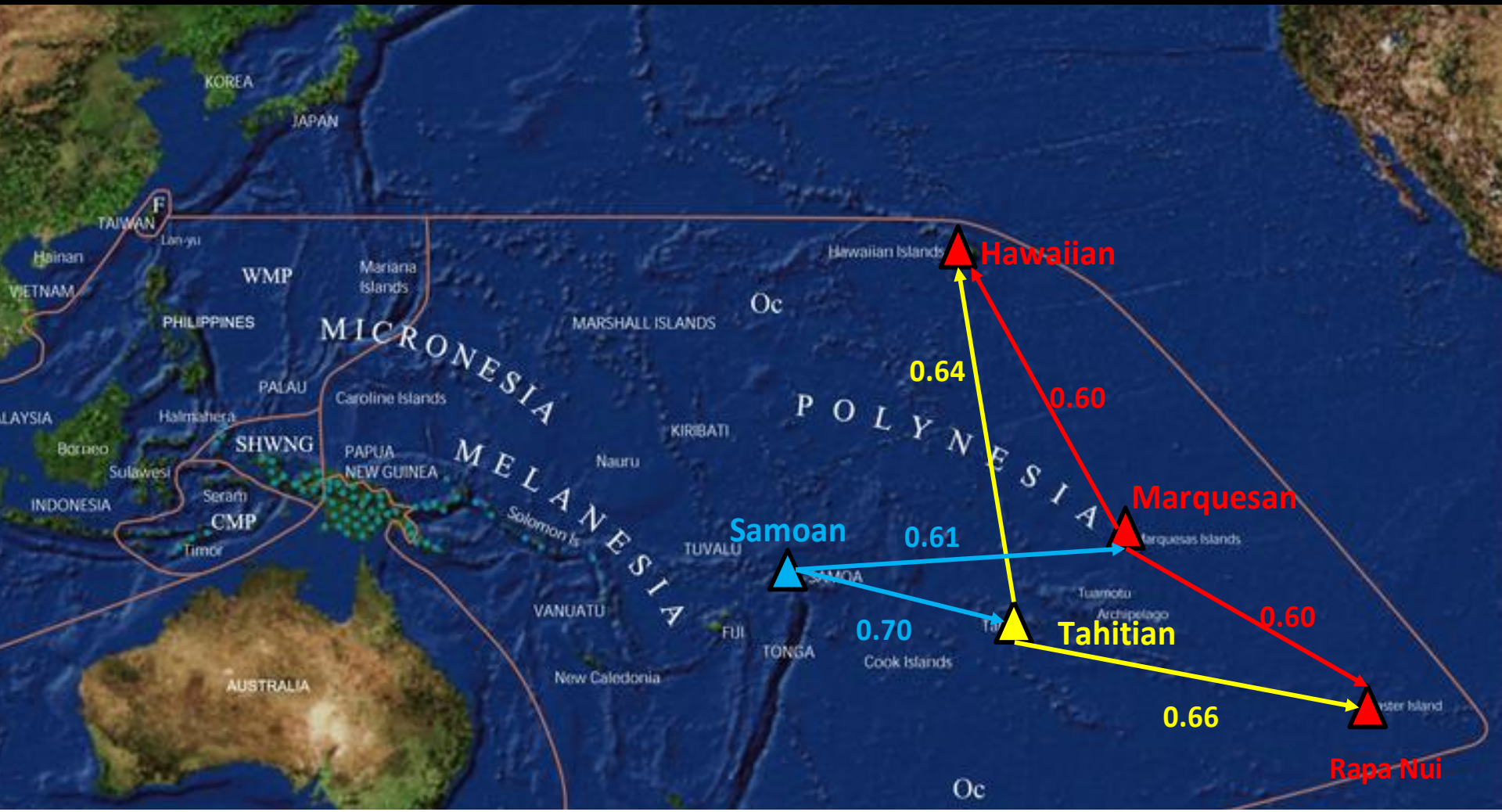
Notice that four branches appear to emerge from this 3D MDS plot: Samoan, Tahitian, Rartongan/Maori, and Marquesan/Hawaiian/Rapanui. This grouping gives us a first-pass estimate as to how these languages may have descended from a common ancestral tongue.



When we plot the lexical distances of three of these groups relative to the modern Samoan language, we see there was a very large change in Tahitian between the 18th century and the present. It increased from 0.64 to 0.70, the largest LD to Samoan. Raratongan and Maori show similar changes in lexical distance to Samoan, 0.59 for Raratongan to 0.64 for Maori. And it's the same for Hawaiian and Rapa Nui compared to Marquesan.

Also notice that 18th century Tahitian is closer to Samoan than is Modern Tahitian, as you would expect since Modern Tahitian is descended from 18th century Tahitian. So Maori is farther away than Raratongan; Hawaiian and Rapanui are farther away than Marquesan. This must imply an order of descent within these groups.

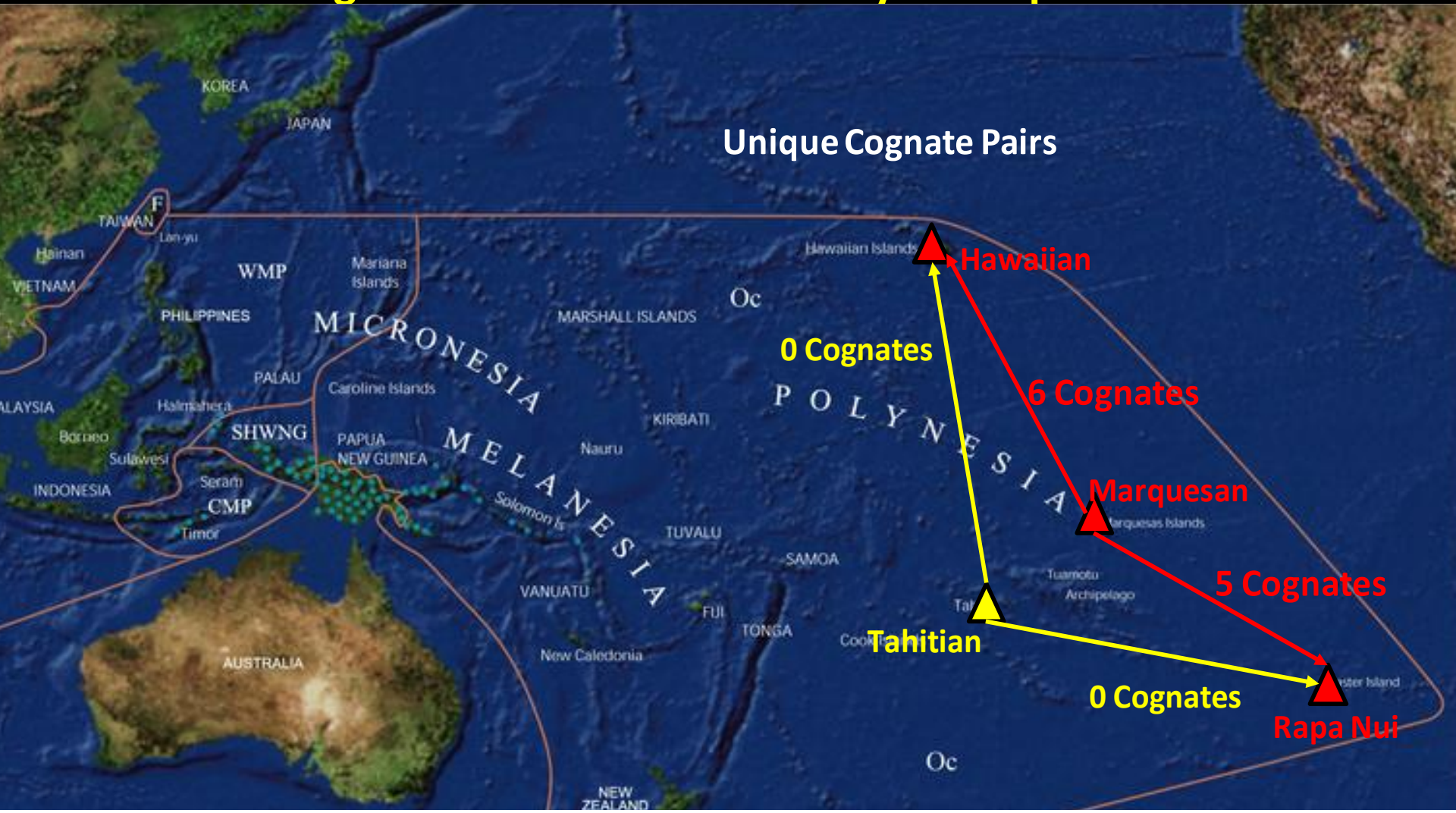
Which Migration Route is More Likely?:Marquesas or Tahiti?



One of the clear implications of the previous plots is that Hawaii and Rapa Nui were settled from The Marquesas. But there is an alternate theory for the settlement of Hawaii and Rapa Nui, one that assumes Tahiti was the source of settlement. The lexical distance numbers show that Marquesan is slightly closer to these two final settlements than it Tahitian. But this is not very compelling evidence by itself. The MDS evidence is fairly convincing, but not ironclad.

Can we actually learn anything more certain from our data about the source of these settlements?

Which Migration Route is More Likely?:Marquesas or Tahiti?



Yes we can! But to do that we need to dig into the details of the cognate words. It turns out that there are 6 cognate sets whose only members are Marquesan and Hawaiian, and 5 sets whose only members are Marquesan and Rapa Nui. When we look for these rare two-language cognate sets between Tahitian and Hawaiian and Tahitian and Rapa Nui, we come up empty. This is pretty strong evidence that words coined in the Marquesas, or which survived from the original Prototype language only in the Marquesas, were propagated to Hawaii and Rapa Nui. There is no evidence that such a thing happened from Tahiti. So our cognate-counting detective work tends to support the standard theory of migration.

Sidney Parkinson
c. 1745 – 26 January 1771

Vocabulary of the language of Otaheite, with remarks SYDNEY PARKINSON, Draughtsman to JOSEPH BANKS, Esq. on his late Expedition with Dr. SOLANDER, round the World. (London, 1773).

Typical Parkinson Illustration:

While we are talking about Tahiti, and before we do any more analysis, let's take a look at the observations of a young amateur linguist upon first contact with the people of Tahiti. This young Scotsman was named Sidney Parkinson. He journeyed to the South Pacific with Captain Cook on his first voyage, working as an illustrator for the soon-to-be-famous naturalist Joseph Banks. Parkinson collected many Tahitian words , of which 124 appear in the Austronesian Basic Vocabulary Database. Because of Parkinson's linguistic observations, and from other sources, linguists have been able to piece together an estimate of what Tahitian was like when spoken in the 18th century, before it was much influenced by Europeans. The language has changed substantially since that time.

Parkinson's Observations on the Tahitian Language

"REMARKS on the Otaheitean Language.
The language is very soft, having a great number of vowels, diphthongs, and triphthongs. Every word, almost, begins with a vowel, which they most commonly drop.
The natives could not repeat, after us, the sounds of the letters, Q, X, and Z, without great difficulty; G, K, and S, they could not pronounce at all. Many of the names of the people of our ship having the G, K, or S, in them, they could not approach nearer the sound of them than as follows:

Toote, for Cook …
Tolano — Solander …
Treene — Green …
Patine — Parkinson

"They have various sounds peculiar to themselves, which none of us could Imitate; some of them they pronounced like B and L mingled together; others between B and P, and T and D. Some like Bh, Lh, and Dh.

Evolution of the Basic Tahitian Vocabulary over ~240 Years

18th Century Tahitian –vs- Modern Tahitian

…About a Quarter of the Words Have Changed

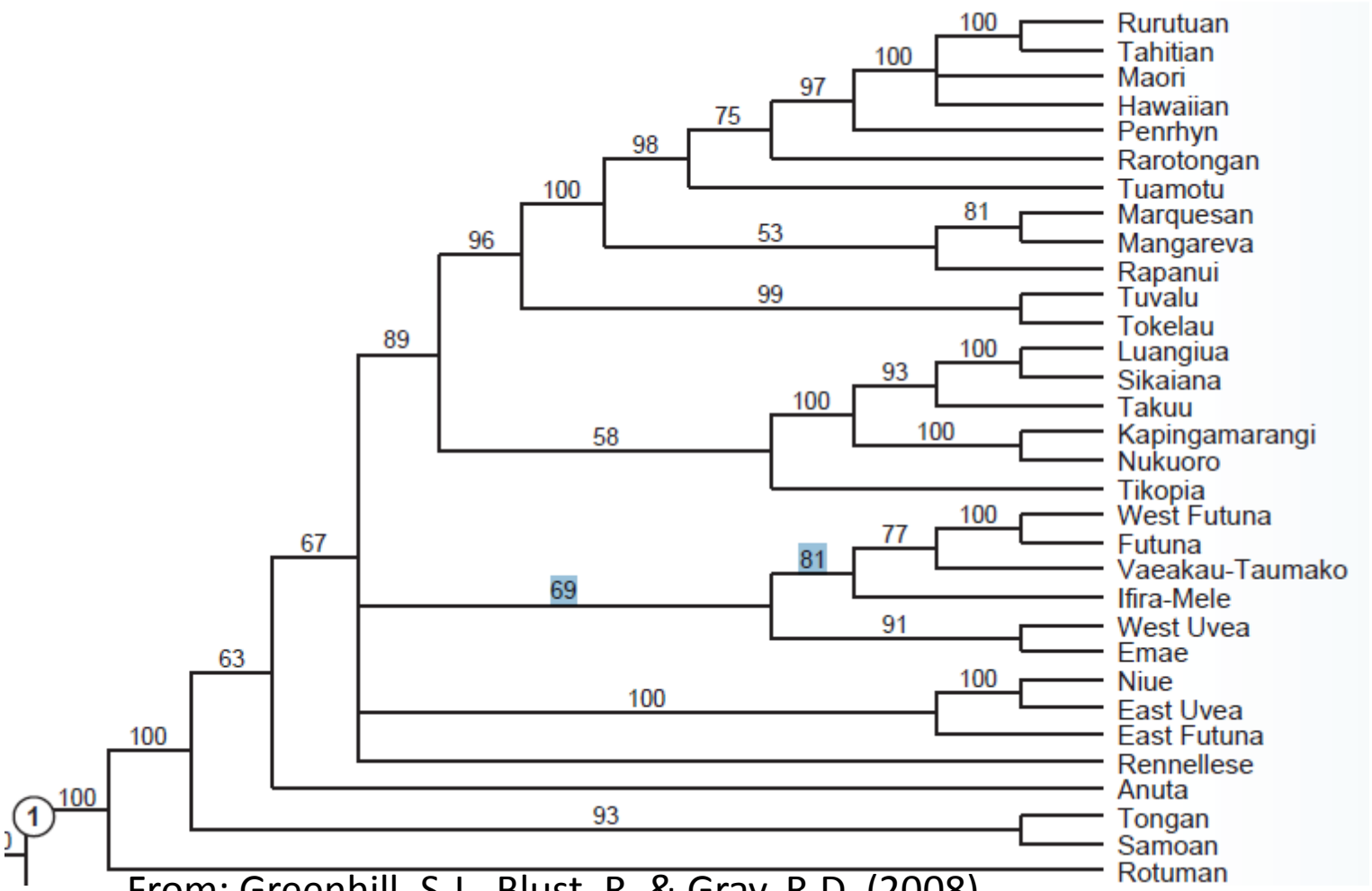
	Identical	Evolved	Innovated	Synonym	Abandoned	Borrowed
% of Comparisons:	**58%**	**16%**	**19%**	**1%**	**5%**	**2%**

Evolved: *teiaha >> teimaha*
Innovated: *mata'u >> ri'ari'a*
Synonym: "Salt" *miti* >> "Sea" *miti*
Abandoned: *tuumuu >>*_____
Borrowed: "Father" *metua taane >>> paapaa*

Analyzed from: *The Austronesian Basic Vocabulary Database,* 2014

Because we have this observation of 18th century Tahitian (then an unwritten language), we can see exactly how the words on our list have changed over the years. This clearly illustrates an important point: languages evolve *in situ*, not just in a settlement relocation. Over time, the lexical distance in the place of origin will approach ONE, whether spoken by settlers on a new island or those who remained behind.

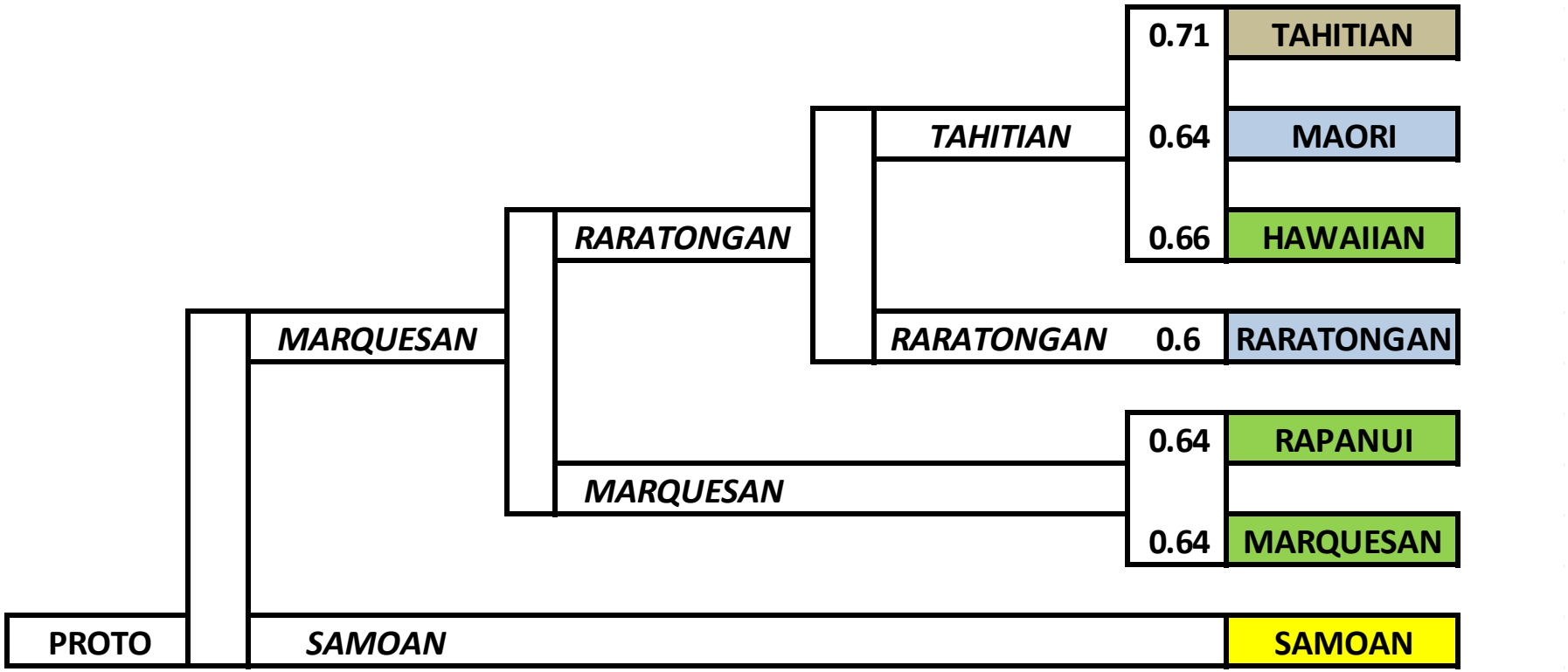
Phylogenic Tree Synthesized from Lexical Distance Data



From: Greenhill, S.J., Blust. R, & Gray, R.D. (2008).
http://language.psy.auckland.ac.nz/austronesian/tree.php

It is possible to arrange languages in something called a Phylogenic Tree. This is similar to trees generated in genetics, or as an attempt to indicate the relationships between various species of plants or animals. But there are some real pitfalls to doing this, and it's easy to be misled by what seems to be a scientifically-derived diagram of relationships.
This tree is the effort by Greenhill, Blust and Gray (the same people who compiled the ABVB list). Let's call this the GBG tree and take a close look at the Polynesian part of it.

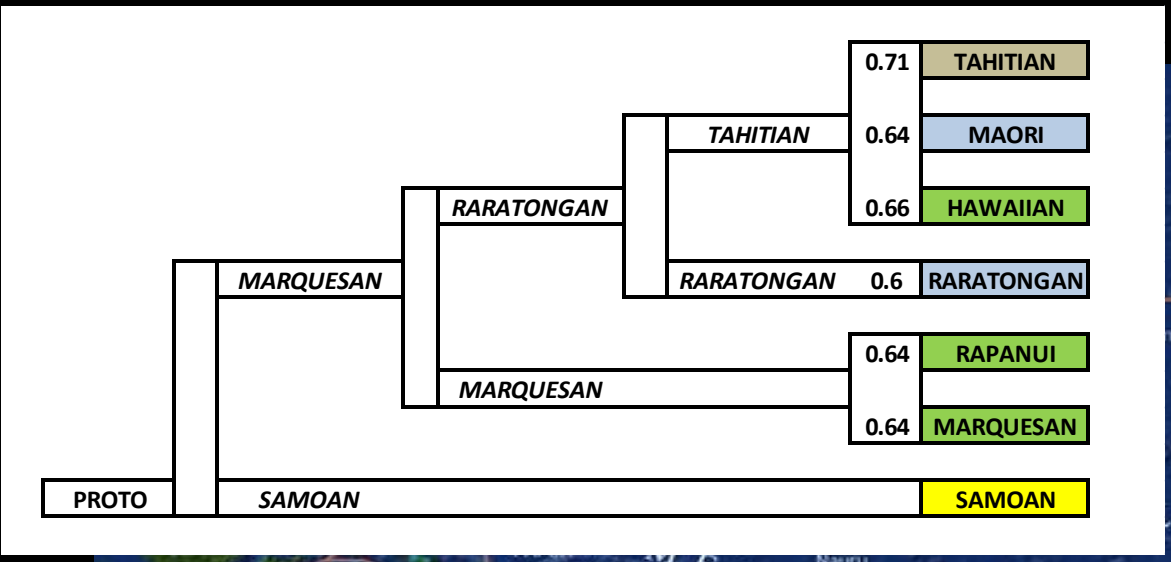
Simplified GBG Phylogeny of Polynesian Languages

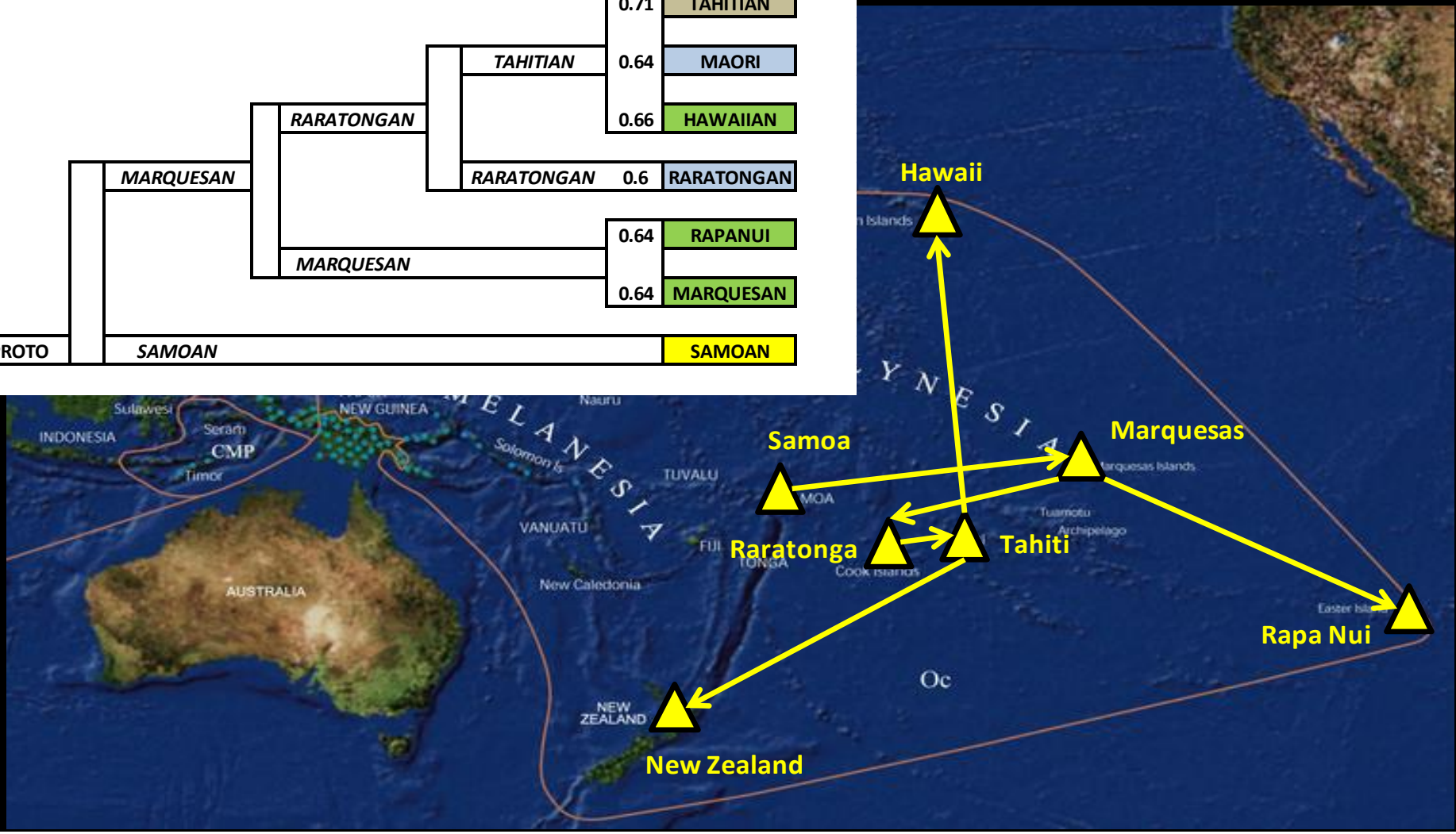


Here I have simplified the GBG tree to show the portion of their tree applicable to our list of major Polynesian languages.

This tree is synthesized by a computer based on lexical distances, and seems to imply that Raratongan was the progenitor of Tahitian, Maori and Hawaiian. (Numbers are Lexical Distances from Samoan).

Migration Route Inferred from GBG Synthesized Tree

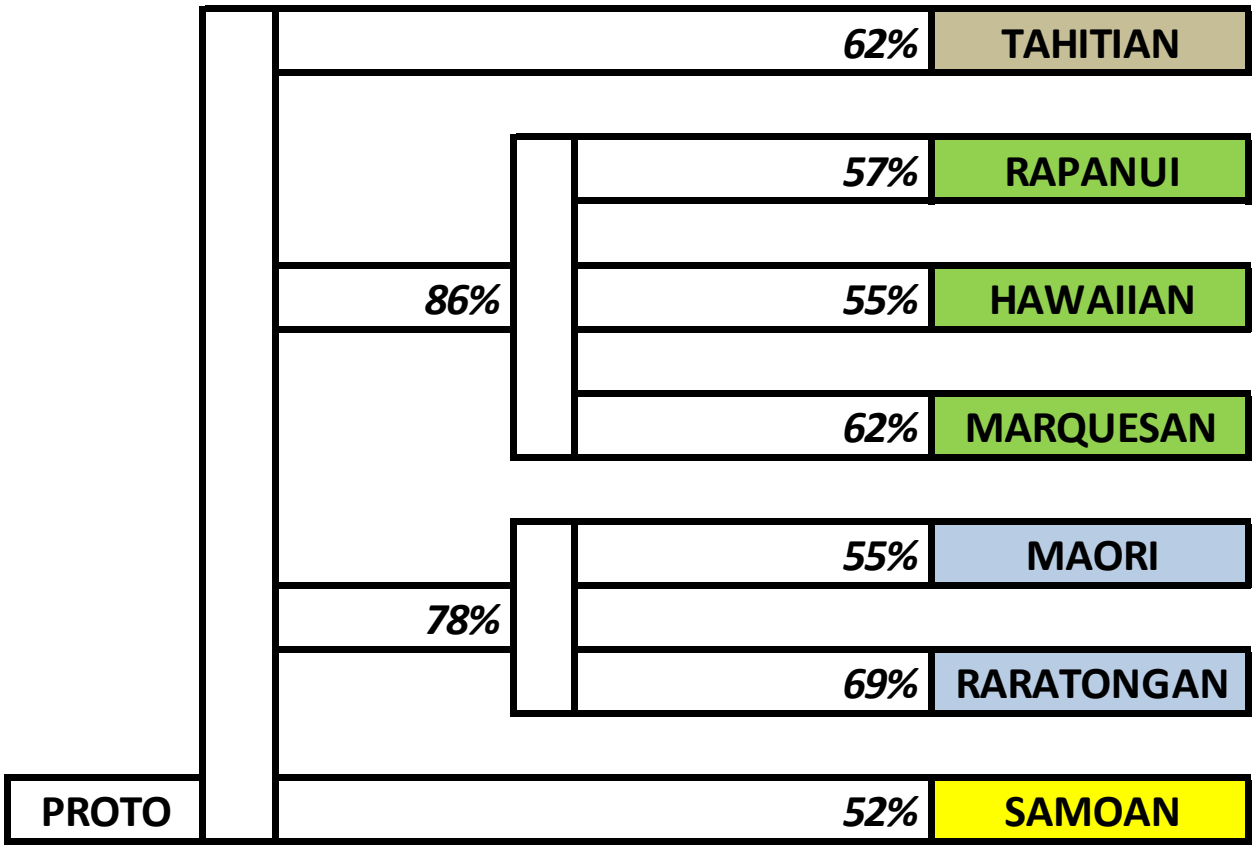




But as you can see, this sequence makes almost no sense. According to it, settlers would have to have backtracked from the Marquesas to Raratonga before settlement could have proceeded. And we've already seen that settlement of Hawaii from Tahiti is extremely unlikely. This is surely not the correct settlement pattern. In Eastern Polynesia, the GBG tree fails the reality check.

Simplified Phylogeny of Polynesian Languages

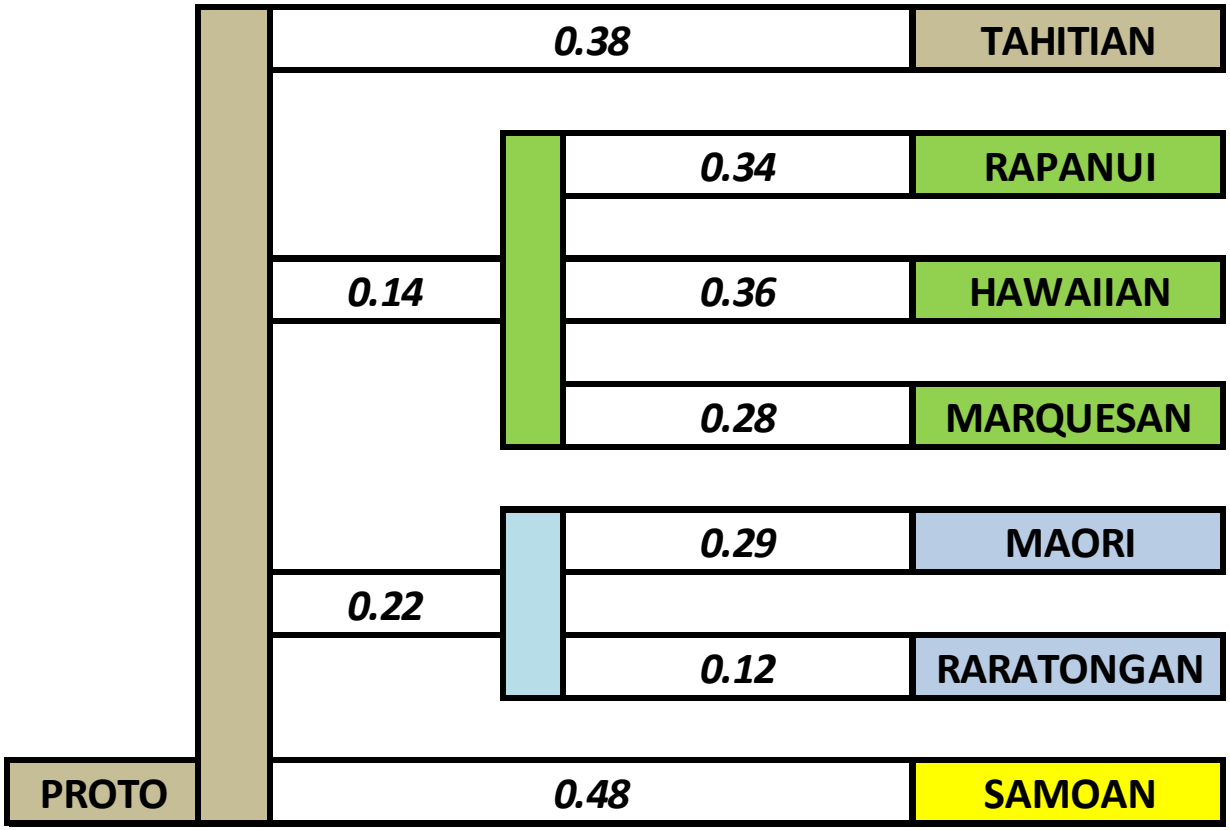
Tree with survival percentages:



Here is a tree I put together based on the "rough draft" from the Multidimensional Scaling Analysis 3D plot. This tree is based on human judgment, not automated calculation.
It has three major branches: Raratongan, Marquesan, and Tahitian.
The percentages of cognates surviving from each previous node are shown.

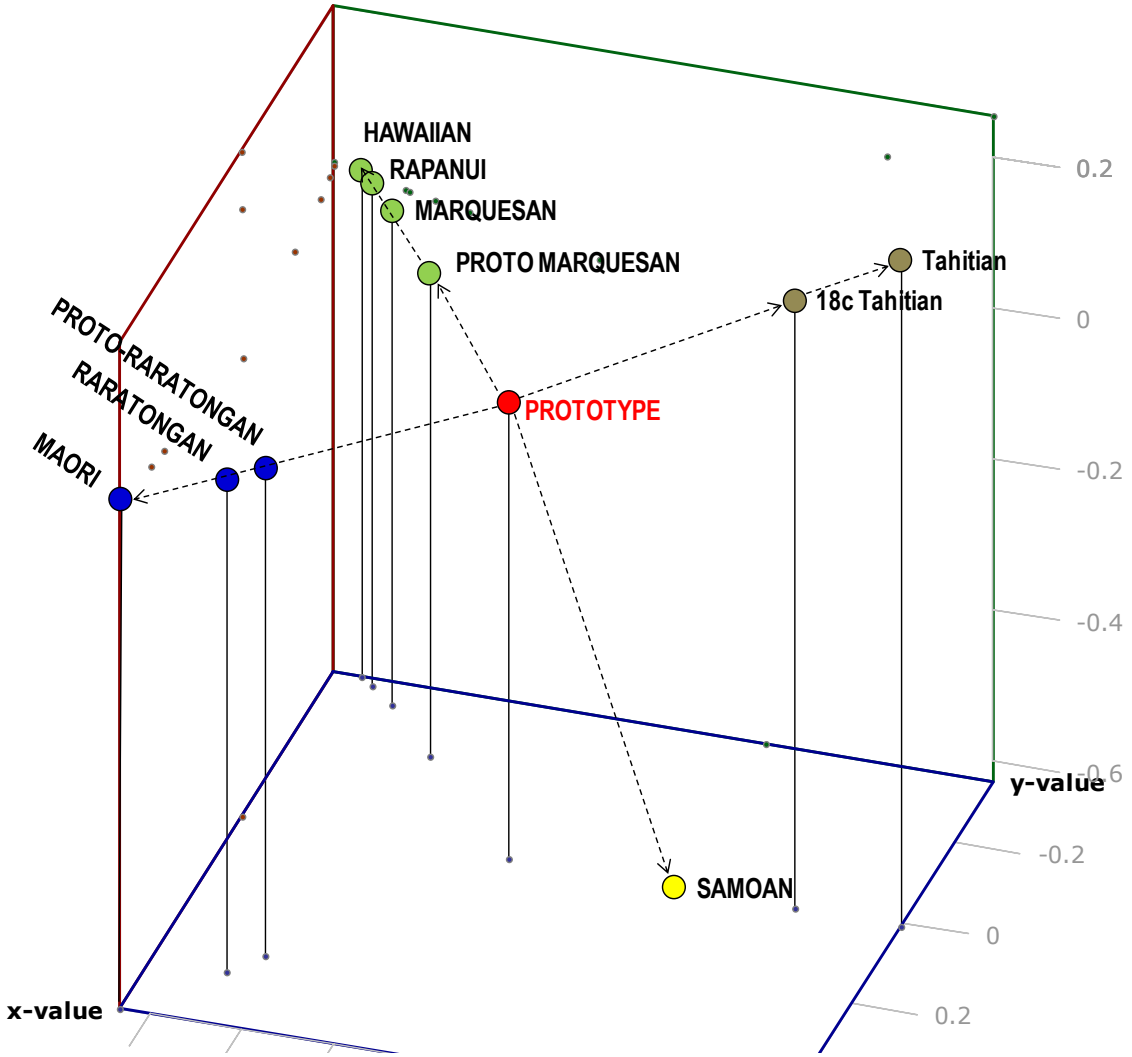
Simplified Phylogeny of Polynesian Languages

Tree with re-calculated Lexical Distances



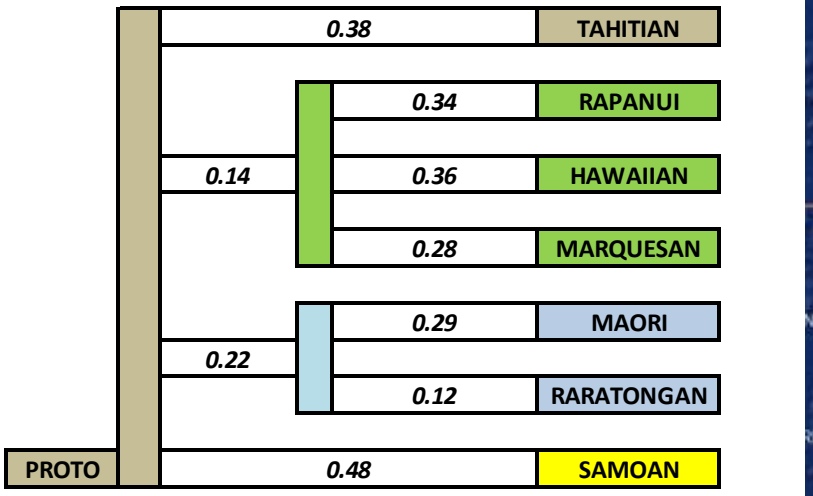
Here is the same tree, but with those percentages of surviving cognates converted to Lexical Distances. Each distance in this case is NOT from Samoan, but from reconstructed prototype languages. These prototypes were constructed by summing all the cognates of the presumed descendant languages and eliminating overlaps. Blue is the proto language for Raratongan and Maori; green is for Marquesan, Hawaiian and Rapanui; brown is the grand prototype for all the languages.

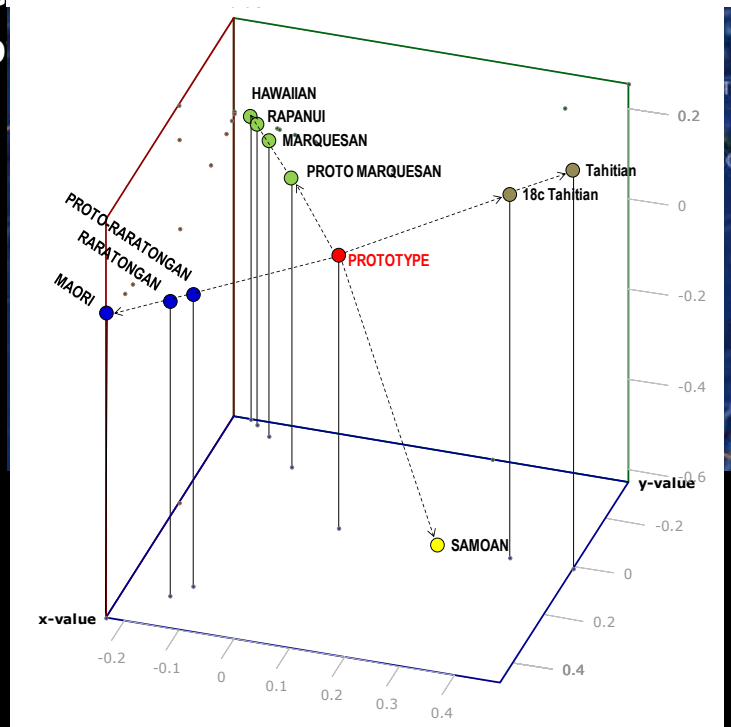
Complete MDS of Polynesian Languages

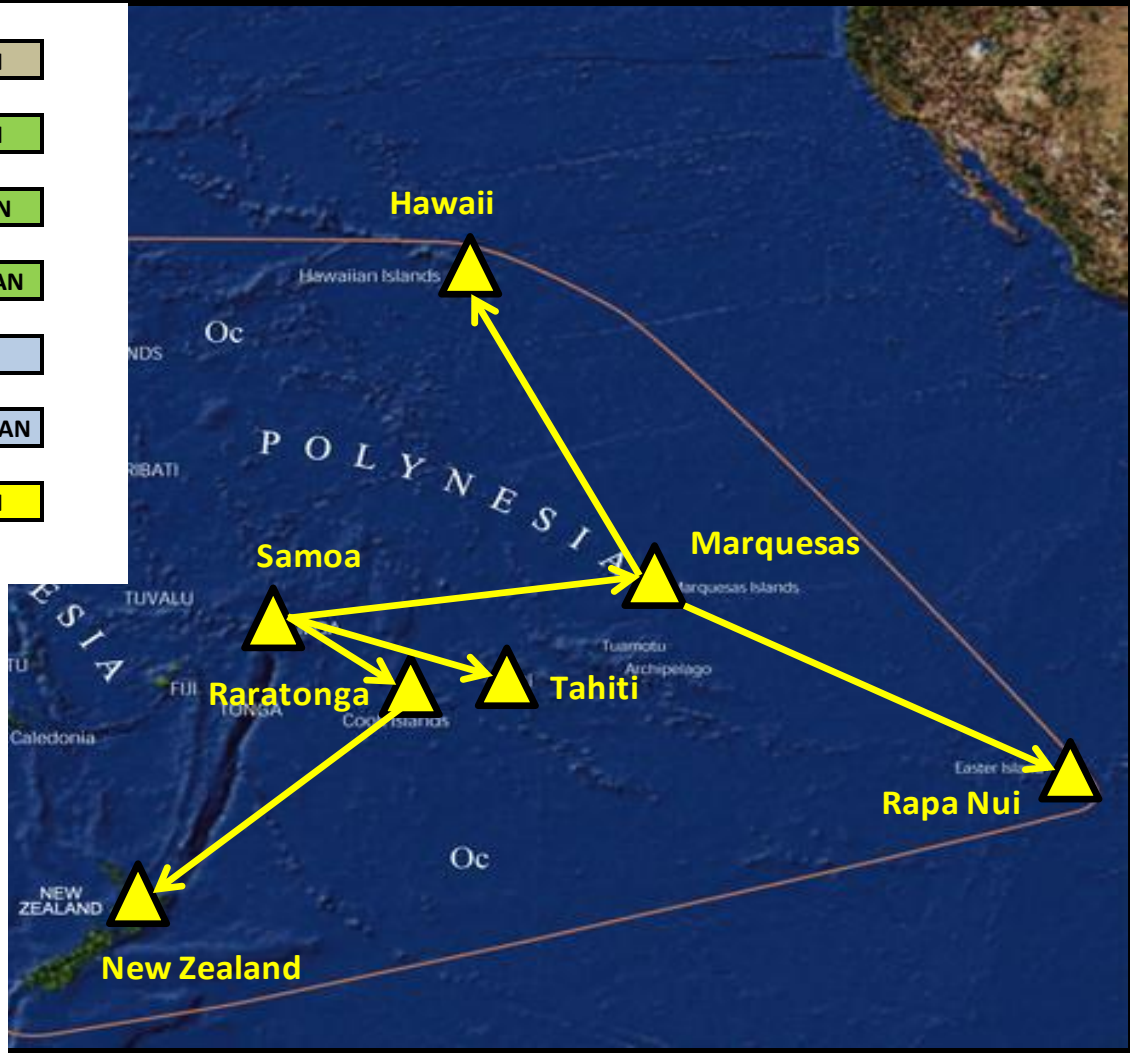


When we generated a new matrix based on the tree with a grand prototype language (red) and sub-prototype languages, we get a very clean, intuitive 3D MDS plot of the lexical distances. The radial distances from the prototype language in the center (red) are proportional to the lexical distances from each language to the prototype. What is NOT shown in this plot are actual lexical distances between modern languages, only the distances each of them is from the prototype.

Migration route Inferred from Standard Model







The most likely physical locations for the speakers of these prototype languages were: Samoa (brown); Raratonga (blue); Marquesas (green).

Summary

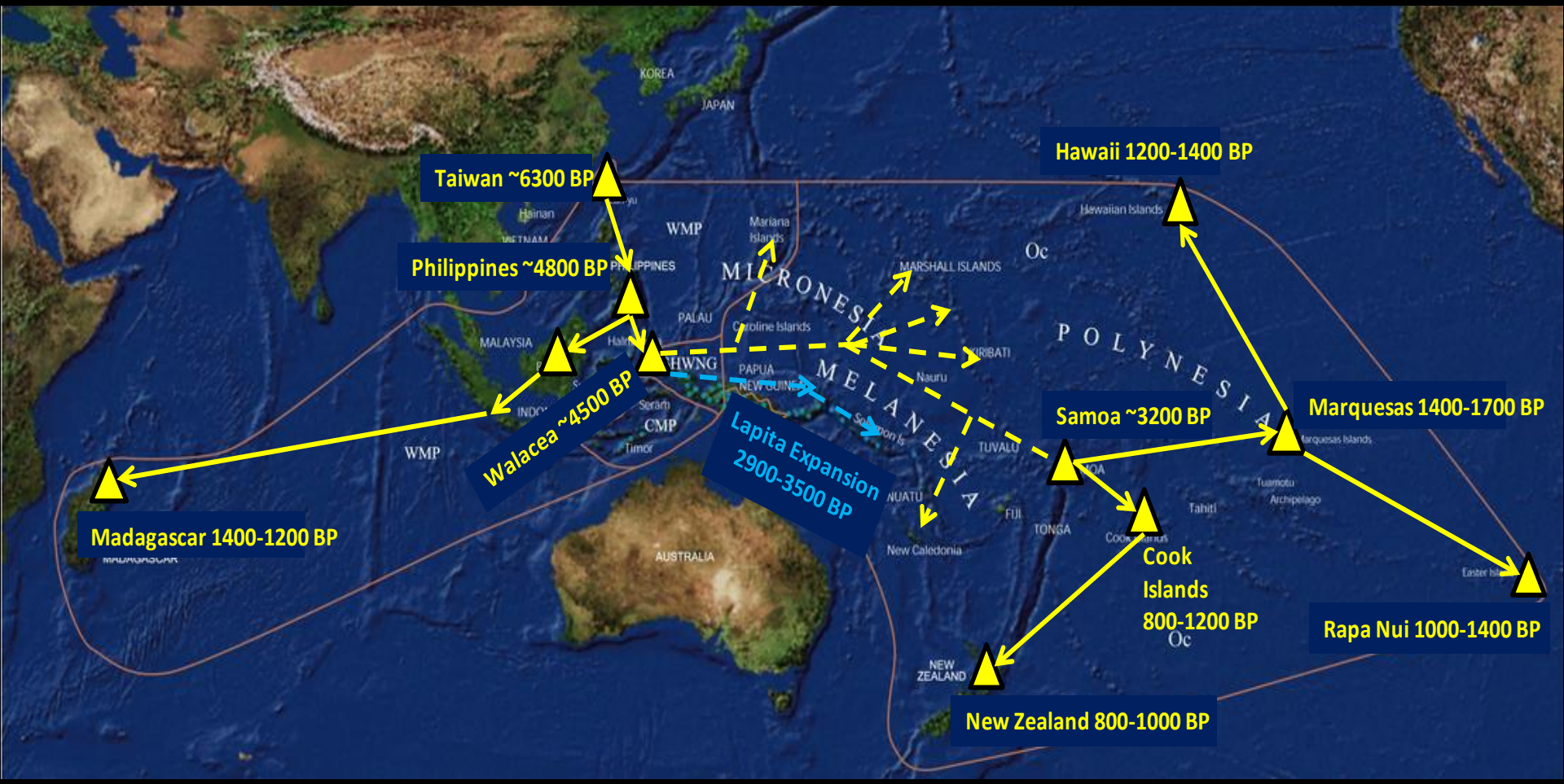
- Austronesia is an “Empire” of Language Similarity Only
- Original Human Settlement of Eastern Polynesia Occurred between 3200 and 1000 years ago
- Careful Lexical Analysis shows Polynesian Language Groupings are Consistent with the Standard Theory of Population Expansion
- Specifically: Pioneers emerged from Samoa to inhabit Raratonga, Tahiti, and The Marquesas. Later Voyages brought People to Hawaii, Rapanui, and New Zealand.

References:

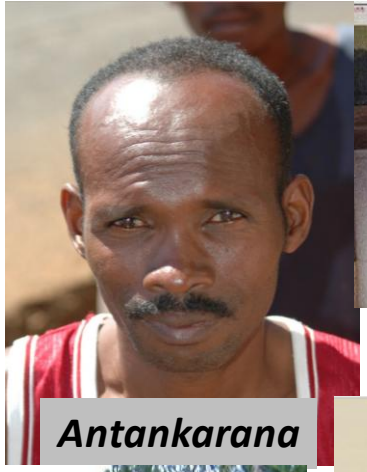
Kirch, P V. and R. C. Green, "Hawaiki, Ancestral Polynesia" (2001)
Diamond, J., "Guns, Germs and Steel" (1997)
Matisoo- Smith, E. and J.H. Robins, "Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat," PNAS (2004)
Hurles, M.E. et al. "Untangling Oceanic settlement: the edge of the knowable," Trends in Ecology & Evolution 01/2003; 18(10):531-540
Hurles, M. E. et al., "The Dual Origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from Maternal and Paternal Lineages," in Am. J. Hum. Genet. 76:894-901, 2005.
Greenhill, S.J., Blust. R, & Gray, R.D. (2008), "The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics.," *Evolutionary Bioinformatics, 4:271-283.*
http://language.psy.auckland.ac.nz/austronesian/
Young, F. W. , "Multidimensional Scaling,"
http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html
Kamakura, W. "Kamakura's Free Software Downloads: Metric MDS,"
http://wak2.web.rice.edu/bio/WagnerKamakuraDownloads.htm

www.roryvantuyl.com/Linguistics.html

The Austronesian Expansion



Polynesia is only part of Austronesia. Let's shift gears now and look at the Western Malayo-Polynesian half of the picture. Specifically, let's look at the westernmost island in Austronesia: Madagascar, and its people, the Malagasy. When people brought the first permanent settlement to Madagascar about 1200-1400 years ago, they brought with them a culture, their genes, and of course an Austronesian language.


Antankarana


Tsimihety


Sakalava


Betsimisaraka


Sihanaka


Betsileo


Merina


Antambohoaka


Antaimoro

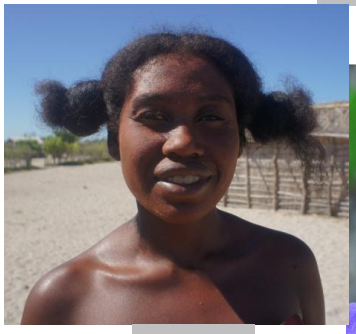

Antaisaka


Antanosy


Mahafaly


Bara


Masikoro


Vezo


Antandroy

These are Malagasy people from 16 of the 20 or so ethnic groups of Madagascar. If you were transported to Madagascar without knowledge of where you were going, you'd surely think you had landed in Africa once you saw the people. But if you were familiar with African language sounds, you would be totally mystified by what you heard: these people may look African, but they sure don't sound African! And you wouldn't be the first person to be confused by this situation.

In 1603, Dutch Merchant Frederick Houtman noticed the Malagasy natives spoke a language "...very similar to Malay."

In 1500, Portuguese navigator Diego Diaz became the first European to visit Madagascar. Other European traders followed, most of them headed to and from the East Indies. In 1603, Dutch Merchant Frederick Houtman noticed the Malagasy natives spoke a language "…very similar to Malay."

And in 1613, Portuguese Jesuit Fr. Luis Mariano wrote that he noticed the similarity of Malagasy speech to *the languages of Southeast Asia.*

He traveled up and down the coasts of Madagascar noting :

"their language...is the same throughout the island... the natives of the South and North understand each other with ease."

But why would people who looked like This...

Talk like people who looked like This?

So Europeans were puzzled from the earliest days as to why people who looked like *This (left)*… would talk like people who looked like *This (right)*?
The eminent 19th century Malagasy scholar Alfred Grandidier called this "Le plus belle énigme du monde" – "The most beautiful mystery in the world."

The answer came clear to Europeans in 1777, when French merchant-explorer Nicholas Mayeur ventured into the central highlands and for the first time discovered the Merina people.

He wrote...

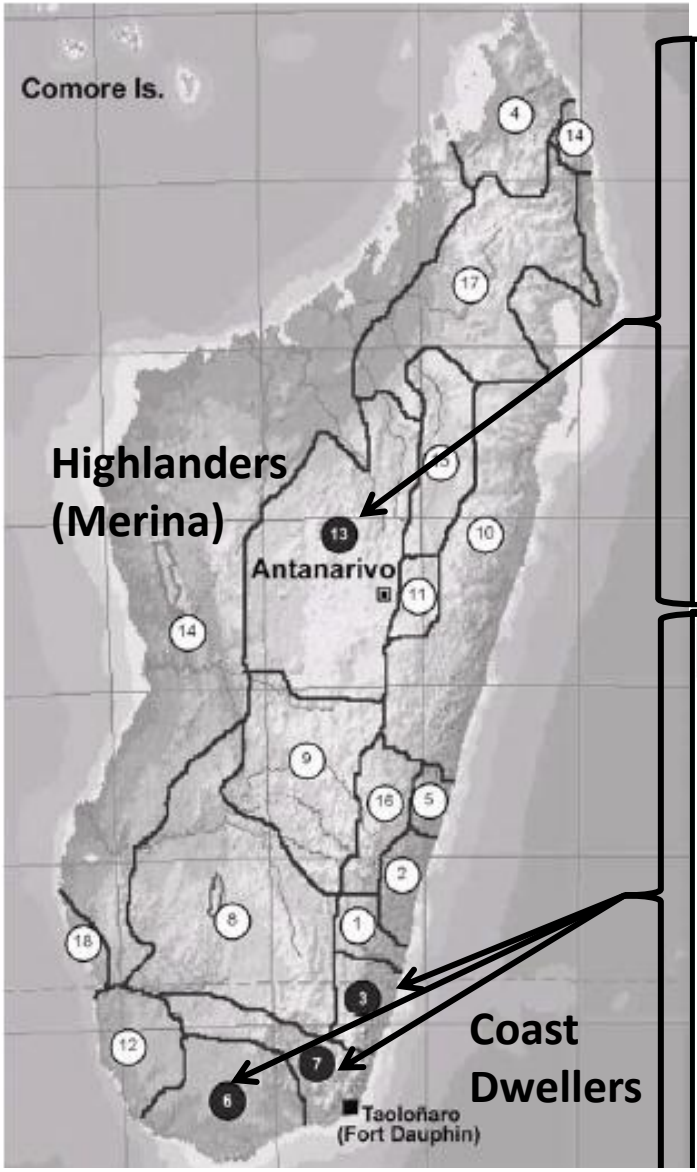
"In the interior of this great island entirely surrounded by savage peoples there is more enlightenment, more industry and a more active administration than on the coasts where the inhabitants are in constant relations with foreigners."

The Merina People

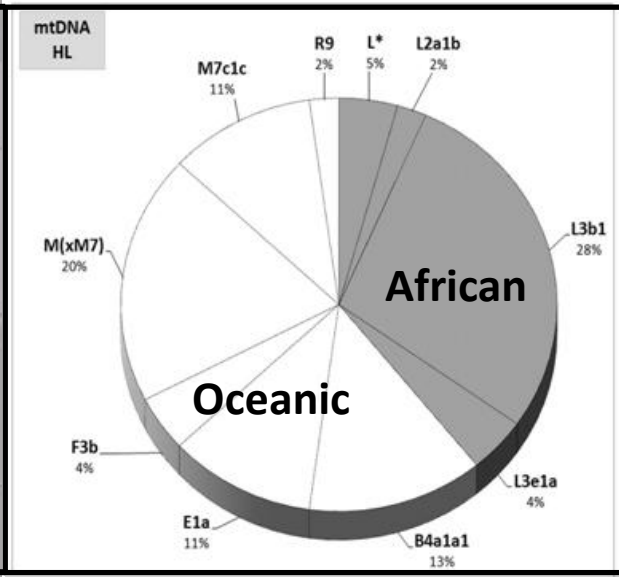
King Andrianampoinimerina (1745–1810)

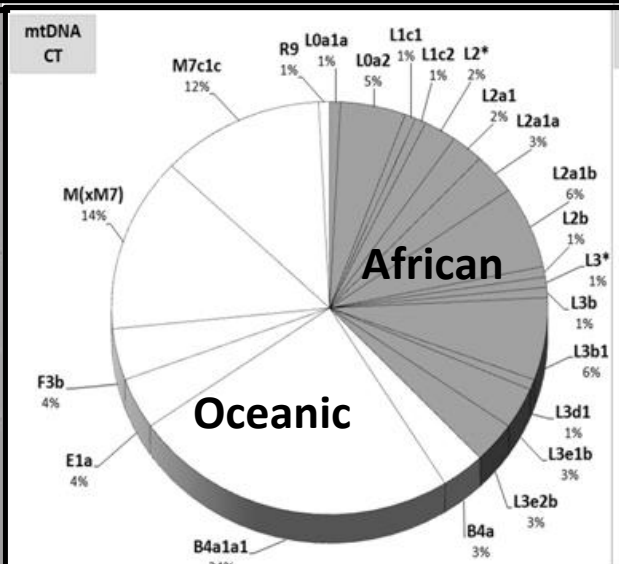
Modern Merina

Here is the Imerina Kingdom in the highlands of Madagascar. It was unified in the 18th century by the great Merina king Andrianampoinamerina. Many Merina - especially the upper classes - don't look like they came from Africa at all. Could they have come from the Malay Peninsula as Houtman hypothesized? Or from SE Asia as Mariano supposed? And did they bring the Malagasy language with them?

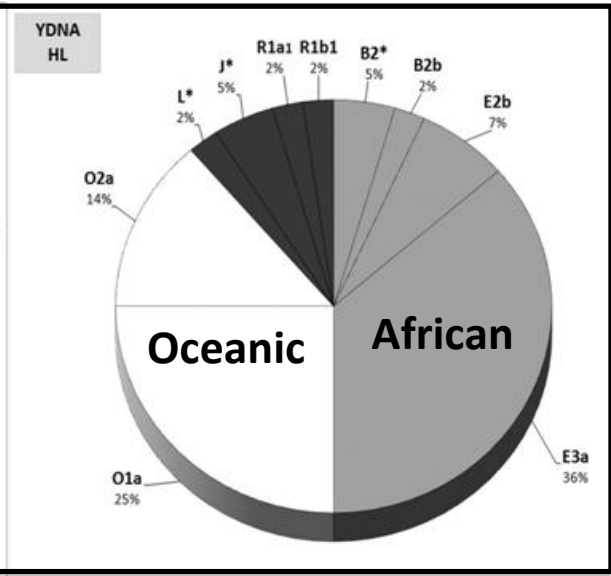


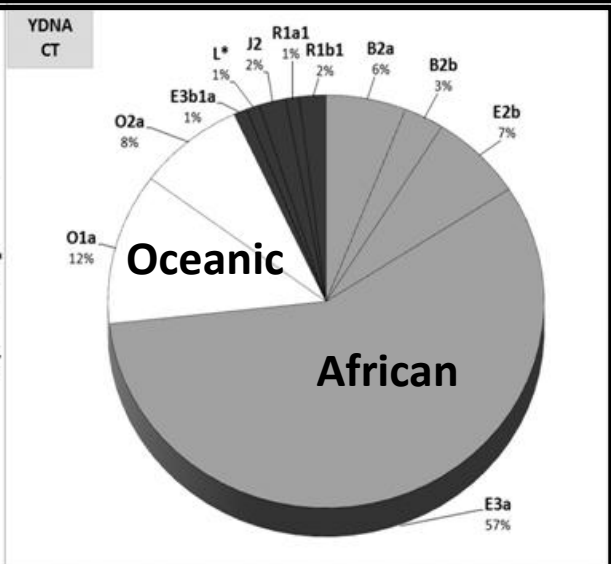
Female Line [mtDNA]





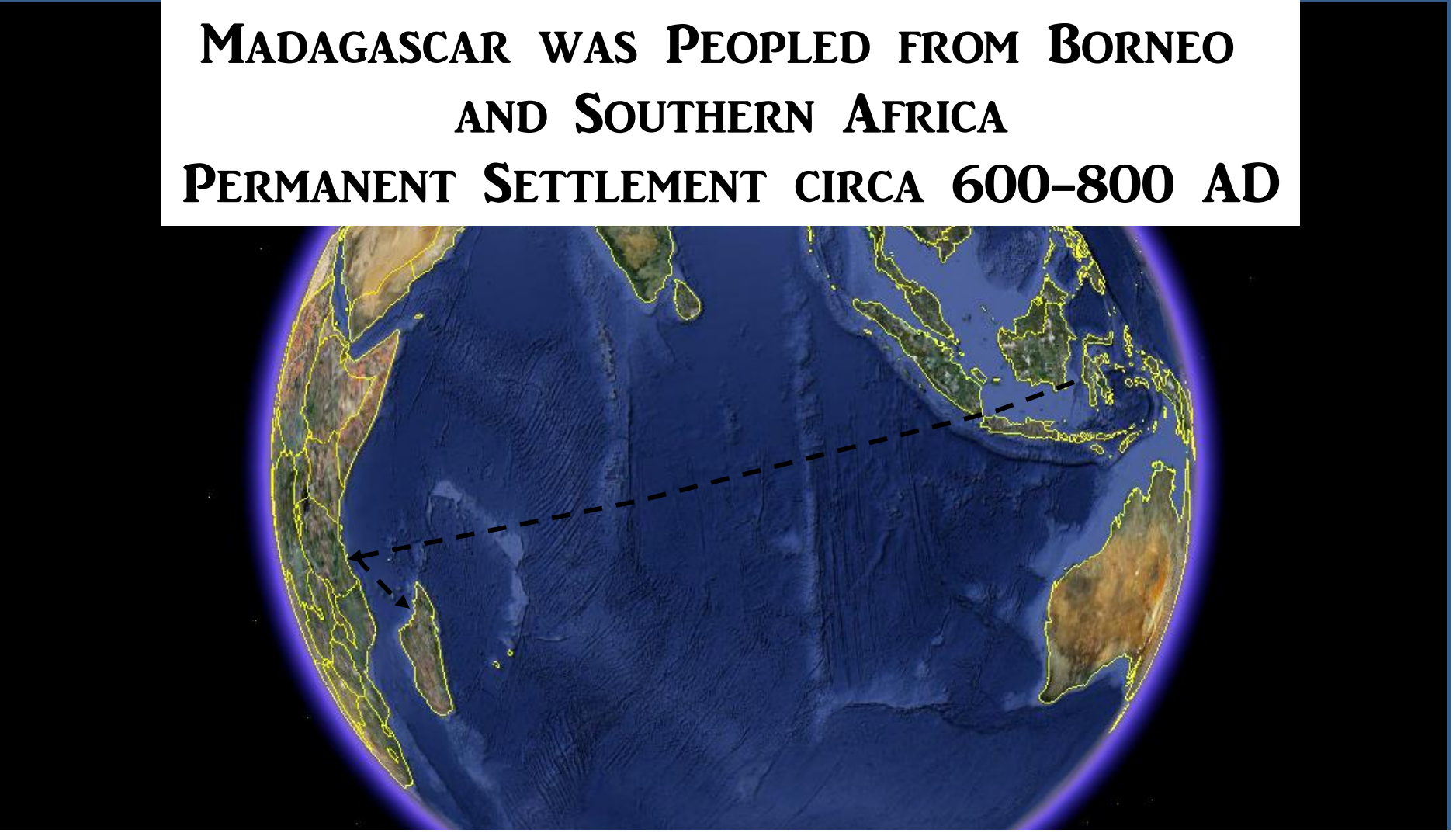
Male Line [Y-DNA]





Here's what the most recent DNA study tells us: Both the Merina and the coastal dwellers from the southern and eastern sections share both African and Oceanic heritage. In fact, the female line (mtDNA) is quite similar for both the Merina and the coastal dwellers: about 60% Oceanic, 40% African. In the male line, the Merina are about 50% African, while the coast dwellers are about three-quarters African.

Madagascar was Peopled from Borneo and Southern Africa
Permanent Settlement circa 600–800 AD



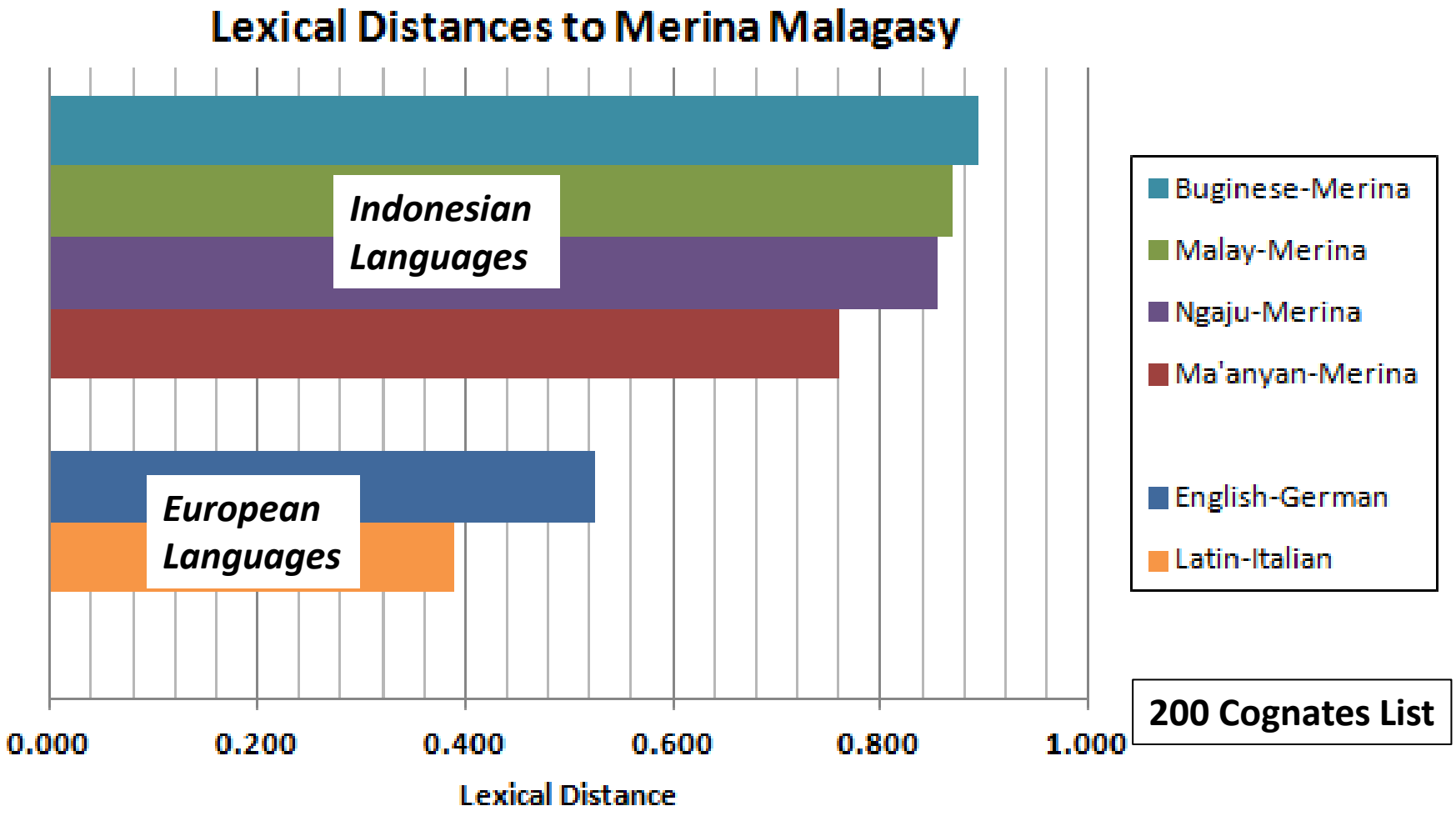
theory holds that the Indonesians sailed from Borneo to mainland Africa, where they hooked up with some tal Africans, and they then migrated together to Madagascar Given the sailing expertise of the Pacific ronesians, it's no mystery how they could have sailed to Madagascar from Borneo. The big question is why n't this place settled much earlier from Africa? Well, Southern Africa wasn't settled by the Bantu until about 1500 years ago, about the same time the Indonesians sailed on the scene. And the Bantu did not at that have the sailing technology of the Indonesians. So hitching a ride from expert Indonesian sailors would been an effective way for them to get to Madagascar.

Indonesian Languages Cognate to Merina Malagasy

And as you've probably figured out, the Malagasy language comes from Indonesia: probably the island of Borneo. The most closely related language is that of the Ma'anyan Dayak people. But the Ma'anyan are land-bound and no longer go to sea. So the actual sailors who set off for the west may have been ancestors of the Ma'anyan who lived near the coast.

Indonesian Languages Cognate to Merina Malagasy

…But Not Very Close



None of the modern Indonesian languages that contain words similar to Malagasy are particularly closely related to it. We see that Ma'anyan and Malagasy are not nearly as close as English to German or Latin to Italian.

All the data I used in preparing the Malagasy portion of this presentation was generated from field research directed by Prof. Maurizio Serva, an Italian physicist and expert on Madagascar and its languages. He compiled his data into 200-word Swadesh lists of 23 Malagasy dialects, arranged by location. And he pioneered the use of technical analysis of this data.

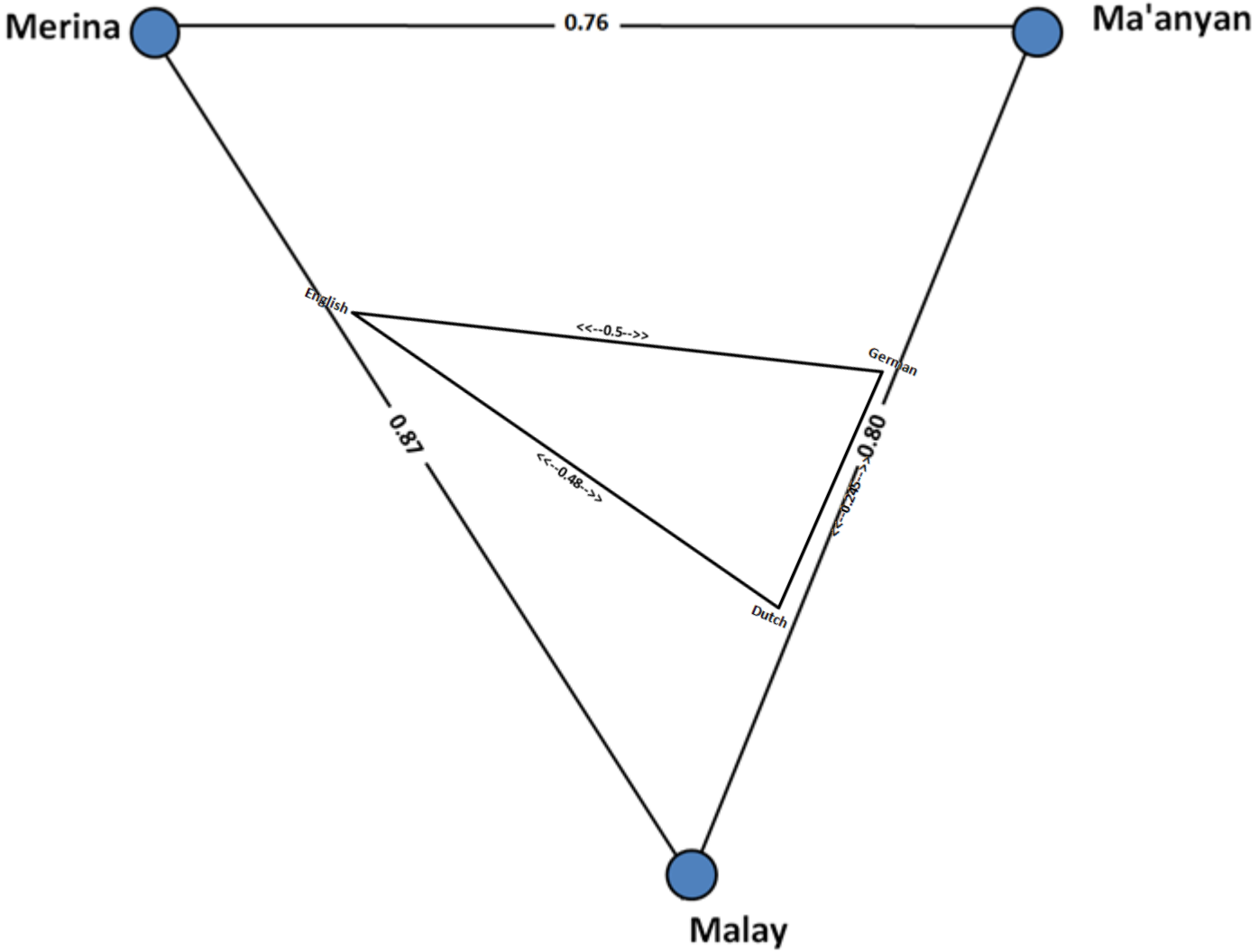
Serva's Malagasy Word List: 200 Words; 23 Dialects

Cognate to Ma'anyan & Merina | Cognate to Merina

	ENGLISH	Ma'anyan	MERINA (ANTANANARIVO)	Ma'ayan/Merina Cognate?	ANTAMBOHOAKA (MANANJARY)	ANTAISAKA (YANGAINDRANO)	ANTAIMORO (MANAKARA)	ZAFISORO (FARAFANGANA)	BARA (BETROKA)	BETSILEO (FIANARANTSOA)
1	all	KATULUH	rehetra	0	izy marobe	aby	iaby	daholo	aby	aby
2	and	ANRI	sy	0	de	sy	da	sy	da	sy
3	animal	SATUA	biby	0	biby	biby	biby	biby	biby	biby
4	ashes	WALENUM	lavenona	0	lavenona	lakevo	lakevogna	lakevo	lakevo	lavenona
5	at	HANG	any	0	agny	agny	agny	agny	agny	any
6	back (of a	PUNGUNG	lamosina	0	lamosigna	lamosy	lamosigna	lamosy	lambosy	lamosina
7	bad	PAMAEH ATEI	ratsy	0	ratsy	ratsy	ratsy	ratsy	ratsy	ratsy
8	bark	KUDIT KAYU	hodikazo	0	hodina kakazo	oditrazo	hodikazo	hoditrazo	hodikazo	hodikazo
9	because	DAYA	satria	0	satria	satria	satria	satria	satria	satria
10	belly	WUNTUNG	kibo	0	kibo	troky	kibo	troky	troky	troka
11	big	HANTE	lehibe	0	agnona be	zakabe	zakabe	zakabe	foloay	lehibe
12	bird	WURUNG	vorona	1	vorogna	voro	vorogna	voro	voro	vorona
190	wife	MATUE WAWEY	vady	0	vady	viavy	vady	vady	vady	vady
191	wind	RIWUT	rivotra	1	agnina	rivotry	rivotry	rivotry	rivotry	rivotra
192	wing	ELAT	elatra	1	elatra	elatry	elatry	elatry	elatry	elatra
193	wipe	NGABARASIS	mamafa	0	mamafa	mamafa	mamafa	mamafa	mamafa	mamafa
194	with	ANDRY	amin ny	0	amin ny	ame	amin ny	amin ny	miaraky	amin ny
195	woman	WAWEY	vehivavy	1	viavy	viavy	viavy	viavy	apela	apela
196	woods	JUMPUN	ala	0	ala	hazo	ala	ala	ala	ala
197	worm	SAASING	kankana	0	viky	haka	ankagna	aka	hanka	kankana

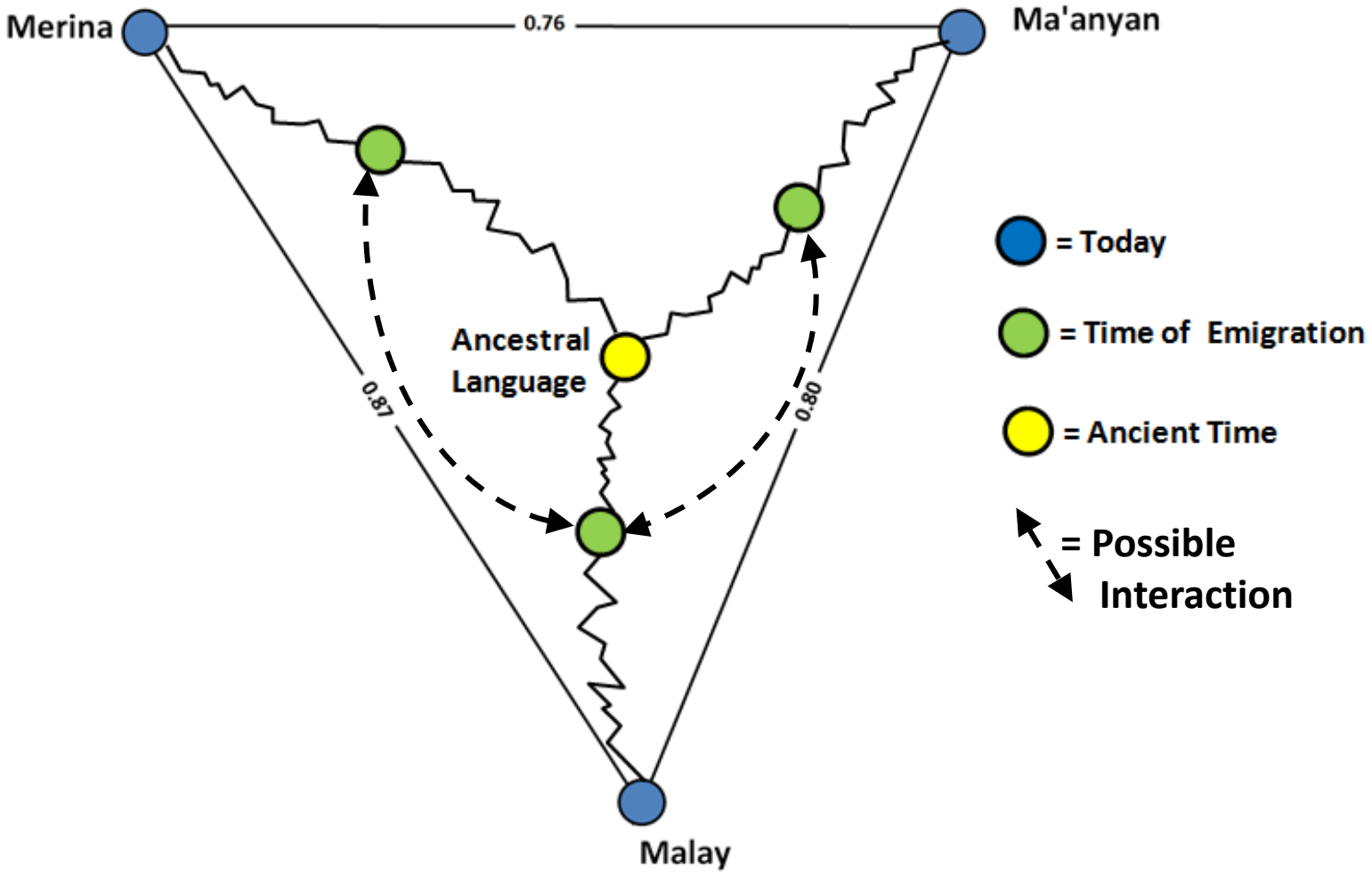
Serva and his collaborators spoke to Malagasy people from 23 different towns and 20 different tribal groups and assembled these lists. What is really good about these lists is their identification of the LOCATION each speaker came from. This turns out to be more important than tribal affiliation when it comes to language. Notice that I have marked in pink the words on this list I judge to be COGNATE to Merina Malagasy. Those highlighted in blue are cognate to Ma'anyan as well as Merina.

Lexical Distances from Merina to some Cognate Languages



Here is a diagram of the lexical distances between Merina Malagasy and some Indonesian languages for which Swadesh Lists are available. Although Ma'anyan is the closest to Merina in terms of lexical distance, it is quite a bit farther away than German is from English, for example. Nevertheless, it is significantly closer to Merina than is Malay or any of the other Indonesian languages we've studied. Does this mean that the people who sailed from Indonesia to Madagascar were only Ma'anyan speakers? Or that Ma'anyan and Malay speakers all got in a boat together and sailed to Madagascar? Probably Not!

Relationships to Ancestral Language



More likely, this is what happened:The Indonesian languages and Merina Malagasy share a common ancient ancestral language. Sub-prototype languages evolved from this ancestral language by the time people sailed for Madagascar. Malagasy settlers and the forebears of the Malay and Ma'anyan speakers all spoke different languages, but these languages were much closer than they are today. These ancient languages diverged over time . Today's Merina Malagasy is a direct descendant of one of those ancient languages, Ma'anyan is a descendant of another. Merina is not directly descended from Ma'anyan.

Cognate Table for Merina and Indonesian Languages

ENGLISH	MERINA	59	MAANYAN	51	NGAJU DAYAK	25	BUGINESE	26	MALAY	23	MAGUINDANAON	17	MARANAO	17	MAKASSAR	18	COGNATES
die	MATT	1	MATEY	1	MATEI	1	MATE	1	MATI	1	MATAY	1	MATAY	1	MATE	1	6
eye	MASO	1	MATE	1	MATE	1	MATA	1	MATA	1	MATA	1	MATA	1	MATA	1	6
fire	AFO	1	APUY	1	APUI	1	API	1	API	1	APUY	1	APOI	1	PEPE		6
five	DIMY	1	DIME	1	LIME	1	LIMA	1	LIMA	1	LIMA	1	LIMA	1	LIMA	1	6
four	EFATRA	1	EPAT	1	EPAT	1	EPPA	1	EMPAT	1	PAT	1	PAT	1	APPA	1	6
sky	LANITRA	1	LANGIT	1	LANGIT	1	LANGI	1	LANGIT	1	LANGIT	1	LANGIT	1	LANGI	1	6
stone	VATO	1	WATU	1	BATU	1	BATU	1	BATU	1	OATU	1	BATO	1	BATU	1	6
bone	TAOLANA	1	TAULANG	1	TULANG	1	BUKKU		TULANG	1	TULAN	1	TOLAN	1	BUKU		5
liver	ATY	1	ATEY	1	ATEI	1	ISSUNG		HATI	1	ATAY	1	ATAY	1	ATE	1	5
three	TELO	1	TELO	1	TELU	1	TELLU	1	TIGA		TELU	1	TELO	1	TALLU	1	5
tongue	LELA	1	LELA	1	JELA	1	LILA	1	LIDAH		DILA	1	DILAQ	1	LILA	1	5
dig	MANGADY	1	NGADI	1	MANGALI	1	MAKKAE	1	MENGGALI	1	KAGKAL		KALI		MAKKALI	1	4
feather	VOLOMBORON	1	WULU	1	BULU	1	BULU	1	BULU	1	BUMBUL		BOLBOL		BULU	1	4
rain	ORANA	1	URAN	1	UJAN	1	BOSI		HUJAN		ULAN	1	ORAN	1	BOSI		4
road	LALANA	1	LALAN	1	KARATAK		LALENG	1	JALAN		LALAN	1	LALAN	1	AGANG		4
rope	TADY	1	TADI	1	TALI	1	TULU		TALI	1	SUMPAAN		TALI	1	OTERE		4
run	MASOANDRO	1	MATEANDRAU	1	MATA ANDAU	1	MATA ASSO	1	MATAHARI	1	SENANG		ALONGAN		MATA ALLO	1	4
year	TAONA	1	TAUN	1	NYELU		TAUNG	1	TAHUN	1	TUHAN	1	RAGON		TAUNG	1	4
bird	VORONA	1	WURUNG	1	BURUNG	1	MANU MANU		BURUNG	1	PAPANUK		PAPANOK		JANGANJANGANG		3
black	MAINTY	1	MAINTEM	1	BABILEM		MALOTONG		HITAM		MAITEM	1	MAITEM	1	LELENG		3
blood	RA	1	IRA	1	DAHA		DARRA	1	DARAH	1	LUGU		ROGOQ		CERA		3
earth	TANY	1	TANE	1	PETAK		TANA	1	TANAH	1	LUPA		BOTEQ		BUTTA		3
name	ANARANA	1	NGARAN	1	ARAN	1	ASENG		NAMA		NGALA		NGARAN	1	ARENG	1	3
smooth	MALAMA	1	MALINEY	1	MALISEN	1	MALONGO	1	LICIN		MATILAK		LANOQ		LACCU		3
that	IRY	1	IRU	1	TE		IATU	1	ITU	1	NAN		ANAN		ANTU		3
thin	MANIFY	1	MARIRIS		TIPIS		MANIPI	1	TIPIS		MANIPIS	1	MANIPIS	1	NIPISI	1	3
bite	MANAIKITRA	1	NGIKIT		MANGIRUT	1	OKKO		MENGIGIT	1	EBUT		KEKEB		ANNOKKO		2
day	ANDRO	1	ANDRAW	1	ANDAW	1	ESSO		HARI		GAY		ALONGAN		ALLO		2
cutr	TSINAY	1	SANAI	1	BAJAKAH KANAI		PIRU		USUS		TINAY	1	TINAQI		PARRU		2
hair (on the	VOLO	1	WULU	1	BALAW		WELUA	1	RAMBUT		BUK		BOK		GAMMA		2
hand	TANANA	1	TANGAN	1	LENGE		JARI		TANGAN	1	LIMA		LIMA		LIMA		2
kill	MAMONO	1	MUNU	1	PATEI		MABUNO	1	MEMBUNUH		GELA		BONOQ		AMBUNO	1	2
live	VELONA	1	WELUM	1	BELUM	1	TUWO		HIDUP		UYAG		OIAG		TALLASA		2
nare	ORONA	1	URUNG	1	URUNG	1	INE		HIDUNG		NGILUNG		NGIRONG		KAMURU		2
person	OLONA	1	ULUN	1	ULUN KALUNEN		TAU		ORANG	1	TAO		TAO		TAU		2
sou	MANJAITRA	1	IKAMIT		MITUR		MAJAI	1	MENJAHIT	1	PAMANAY		PAMANAQI		ANJAI	1	2
sleep	MATORY	1	MANRE	1	BATIRUH		MATINRO	1	TIDUR		TULUG		TOROG		ATTINRO		2
woman	VEHIVAVY	1	WAWEY	1	BAWI	1	MAKUNRAI		PEREMPUAN		BABAY		BEBAY		BANEA		2
breathe	MIAINA	1	MIHEWUK		MANAHASENG		MANNAWA	1	BERNAFAS		GINAOA		PENGGINAOA		AMAI		1
cloud	RAHONA	1	RAKUN	1	BAUN ANDAW		ELLUNG		AWAN		GABUN		GABON		RAMMANG		1
come	AVY	1	HAWI	1	DUMAH		TUMAI		DATANG		ANGAY		DANON		BATTU		1
dry	MAINA	1	MAEYANG	1	KEYANG		MARAKKO		KERING		MAMALA		MAGANO		KALOTORA		1
egg	ATODY	1	ATELUY	1	TANTELUH		ITTELLO		TELUR		LEMAN		ORAK		BAYAEO		1
far	LAVITRA	1	LAWIT	1	KEJAW		MABELA		JAUH		MAOATAN		MAOATAN		BELLA		1
fat	MATAVY	1	TAWE	1	ENYAK		LUNRA		LEMAK		LEMBU		SIBOQ		JAMEJAME		1
leaf	RAVINA	1	RAWEN	1	DAWEN		DAUNG		DAUN		LAON		RAQON		LEKO		1
lie	MANDRY	1	MANDRE	1	PENTER		TURU		BERBARING		EDTATAIGA		IGAQ		ANRABA		1
man	LEHILAHY	1	UPU		HATUE		ORANE		LELAKI	1	MAMA		MAMA		BURANE		1
mouth	VAVA	1	WAWA	1	NYAMA		TIMU		MULUT		NGALI		NGARIQ		BAWA	1	1
new	VAO	1	WAU	1	TAHETA		MABARU		BARU		BAGU		BEGO		BERU		1
night	ALINA	1	KAMALEM		ALEM	1	WENNI		MALAM		MAGABI		GAGAOIQI		BANNI		1
thir	ITY	1	ITI	1	TUH		IAE		INI		NIA		INI		ANNE		1
tua	ROA	1	RUEH	1	DUE		DUA		DUA		DUOA		DOA		RUWA		1
water	RANO	1	RANU	1	DANUM		UAE		AIR		IG		AIR		JENE		1
what	INONA	1	INUN	1	NARAI		AGA		APA		NGAYN		AI		APA		1

But it is nevertheless clear, when you look at the list of all 59 cognates to Merina (highlighted in yellow), that Ma'anyan is the closest relative to Merina Malagasy. There are 29 "Golden Words" on the list. They have cognates to 3 or more Indonesian languages. And there are 8 words on the list that are cognate to Merina from only languages other than Ma'anyan. This is further evidence that the true source language was NOT Ma'anyan, but rather a common ancestral tongue.

So Language analysis and DNA data confirm this theory:
Many Ancestors of the Malagasy People came from Indonesia

Though Derived from Indonesian Language, Merina Malagsy is farther removed from Indonesian than English is from German. Malagasy is unintelligible to Indonesians today.

But what about the various dialects of Malagasy spoken today? Can people throughout the island communicate well with speakers of other dialects?

How Well Do the Malagasy Communicate?

Here's what foreign auditors have said in historical times:

(1613) *"THEIR LANGUAGE...IS THE SAME THROUGHOUT THE ISLAND...THE NATIVES OF THE SOUTH AND NORTH UNDERSTAND EACH OTHER WITH EASE."* **--Fr. Luis Mariano**

(1777) *"...I WAS UNDERSTOOD EVERYWHERE. HOWEVER, I RECOGNIZED A DIFFERENCE IN THEIR WAY OF [PRONOUNCING] CERTAIN WORDS FROM ONE PROVINCE TO ANOTHER."*
--Nicolas Mayeur

How Well Do the Malagasy Communicate?

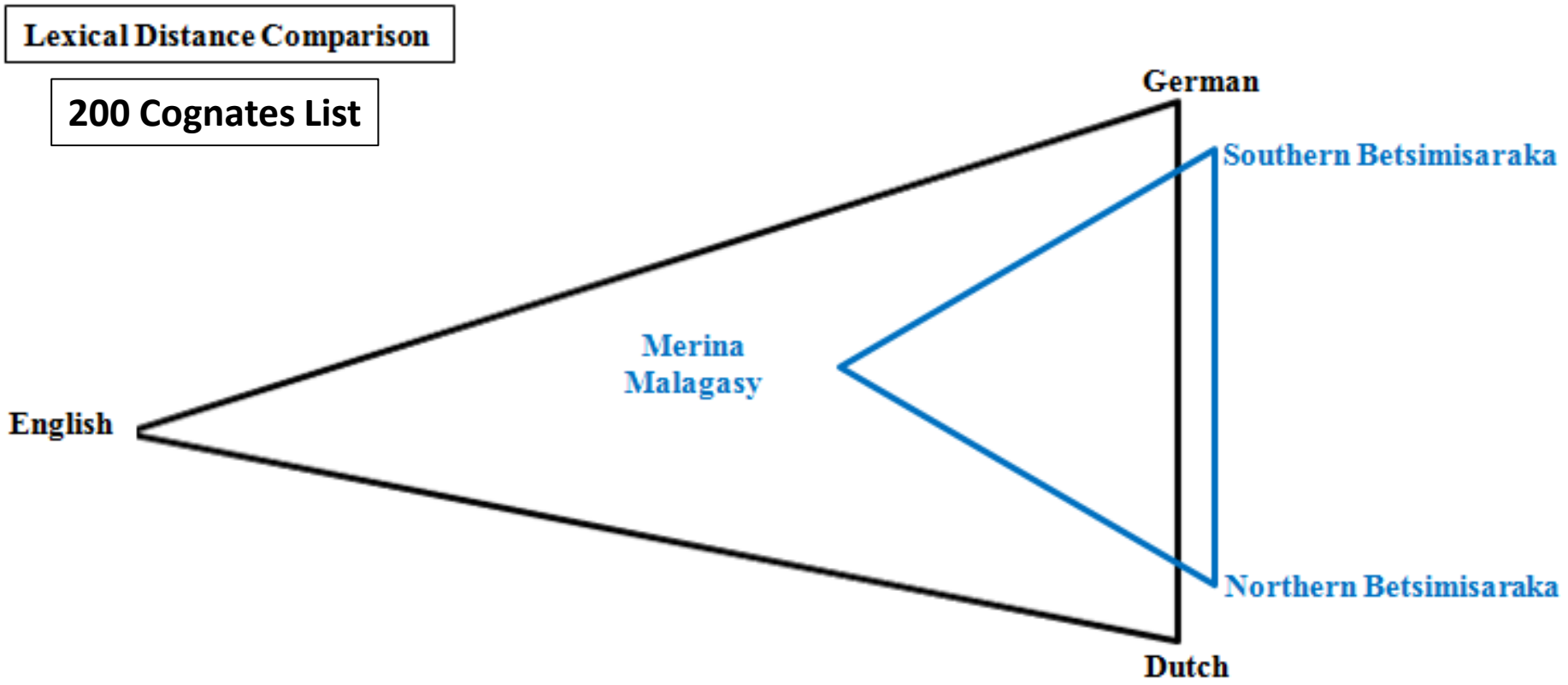
(2012) *"Based on my own experience of staying in a non-Merina region, I feel comfortable to claim that if two speakers from different regions distant from each other speak to each other, they typically have problems communicating if they only use their own speech varieties. However, in an actual situation, such speakers negotiate with words and expressions they know of other varieties, eventually establishing a form of communication."*

--Ritsuko Kikusawa, National Museum of Ethnology, Japan

(2011) *"Dialects from close regions are usually perceived as being similar by Malagasy people while distant dialects usually have a low degree of mutual intelligibility. Most of the people are able to understand the Merina dialect, which is the official language, but outside of the Imerina region only cultivated people are able to speak it."*

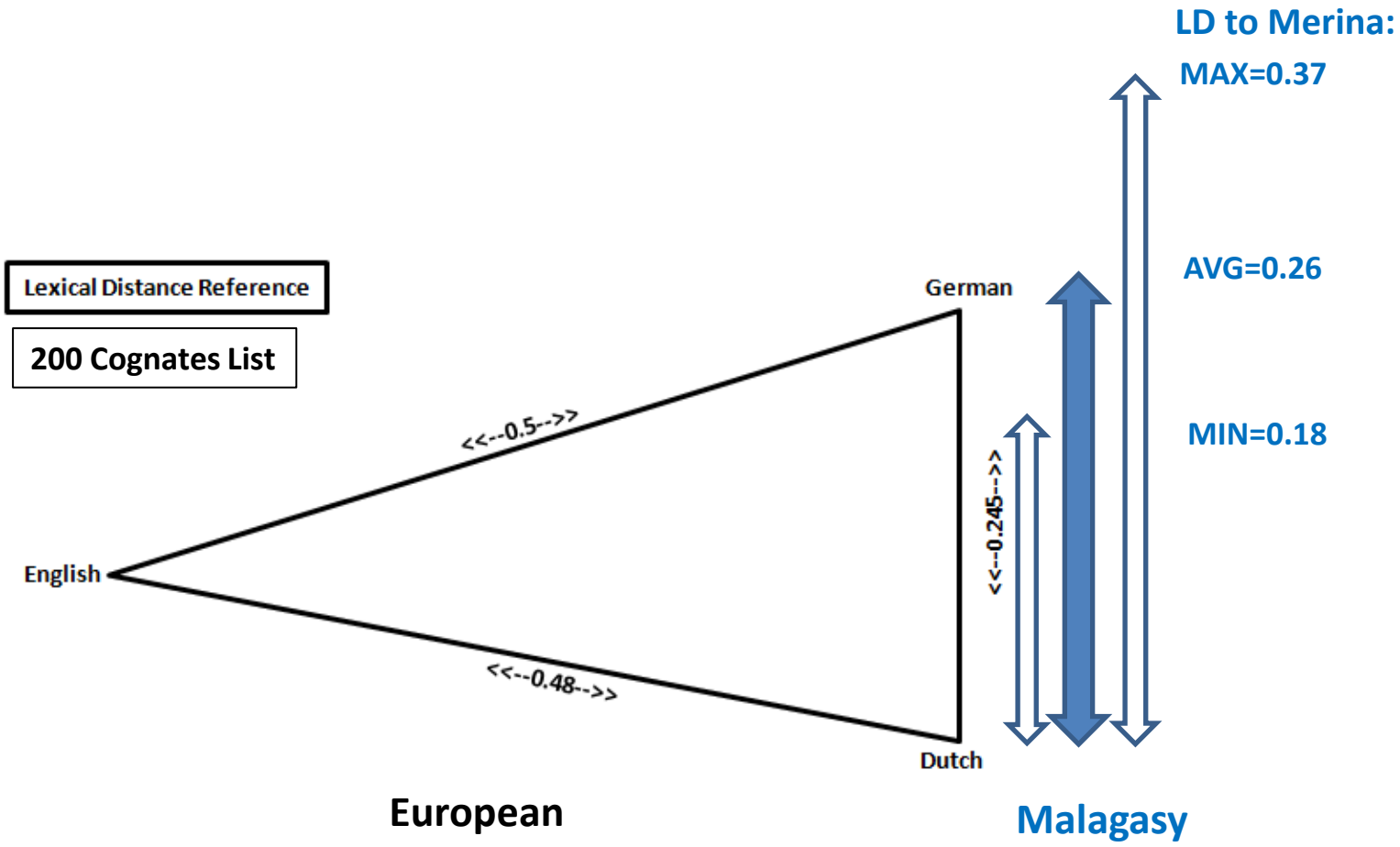
--Maurizio Serva, Universitá dell'Aquila, Italy

Moderate Malagasy Lexical Distances

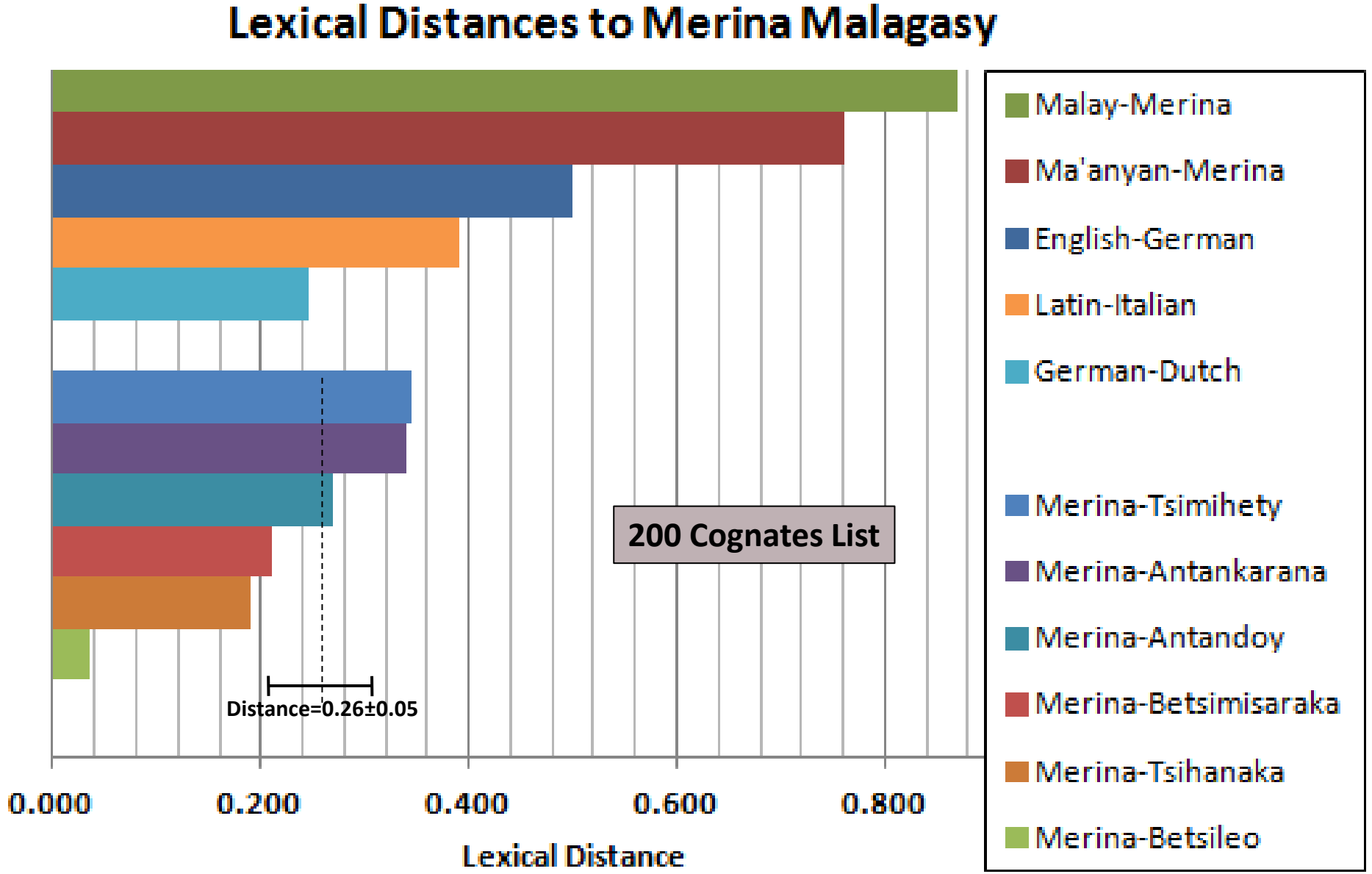


Here is the difference between three Malagasy dialects as compared to our English/German/Dutch reference. As we see, these inter-dialect distances are comparable to the German/Dutch distance.

Comparison of Lexical Distances: Malagasy –vs- European



The average lexical distance between Merina and the other Malagasy dialects is 0.26±0.05, slightly more than the distance from Dutch to German. This gives us some idea of how different these dialects are today. And this doesn't apply only to the Merina dialect. The average Lexical Distance between any two dialects is 0.28.



As we have seen, the distance between Merina and all other dialects is 0.26, similar to the distance from German to Dutch. But the Merina and Betsileo dialects are extremely close because the Betsileo, another plateau tribe, were conquered and subjugated by the Merina in the 18th century. Conquerors can impose speech on the conquered.

Lexical Distance Matrix for 23 Malagasy Dialects

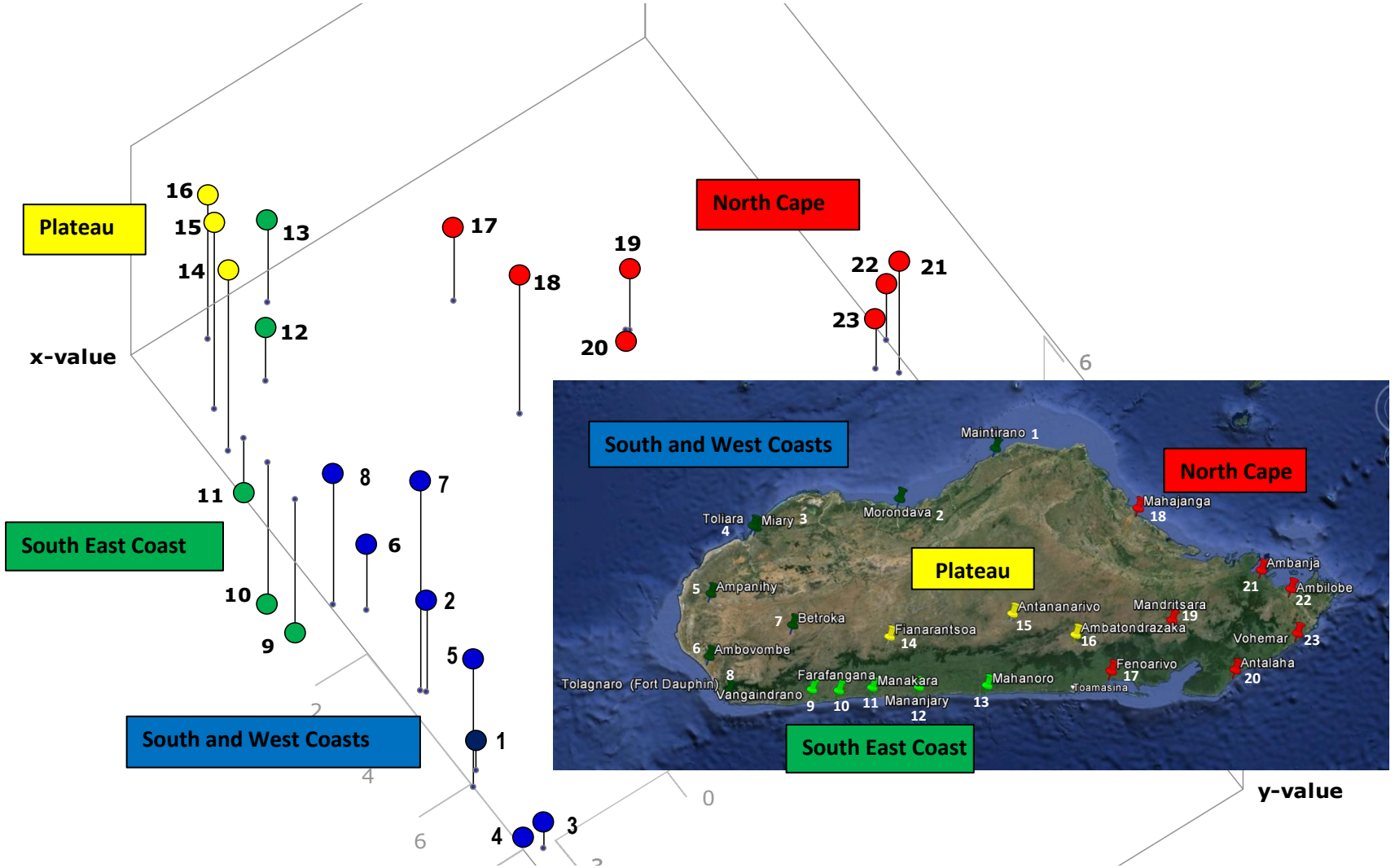
1000x Lexical Distance using 372 cognate sets denominator

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
MAINTIRANO	1		546	535	543	567	581	575	570	591	591	597	599	621	583	589	602	608	597	610	629	648	640	629
MORONDAVA	2	546		556	556	562	575	567	565	581	597	586	583	605	570	565	583	597	578	613	618	640	624	634
MIARY	3	535	556		508	556	567	578	556	597	602	605	613	634	594	599	621	618	597	621	618	640	634	634
TOLIARA	4	543	556	508		554	573	567	556	591	594	591	605	629	597	602	624	616	599	618	624	634	637	640
AMPANIHY	5	567	562	556	554		527	554	551	597	599	599	599	624	583	591	613	618	597	624	621	645	637	634
AMBOVOMBE	6	581	575	567	573	527		559	559	602	605	602	581	613	559	562	591	605	583	608	621	640	637	637
BETROKA	7	575	567	578	567	554	559		524	570	583	570	565	597	535	551	567	589	573	597	608	642	632	629
TOLAGNARO	8	570	565	556	556	551	559	524		556	565	562	556	591	546	554	565	583	562	597	591	624	624	618
VANGAINDRANO	9	591	581	597	591	597	602	570	556		524	548	575	605	562	567	581	586	594	616	597	640	651	626
FARAFANGANA	10	591	597	602	594	599	605	583	565	524		551	551	583	570	573	570	597	594	610	602	640	651	637
MANAKARA	11	597	586	605	591	599	602	570	562	548	551		556	575	556	551	565	578	586	602	594	642	645	648
MANANJARY	12	599	583	613	605	599	581	565	556	575	551	556		538	562	554	543	575	559	597	599	624	637	624
MAHANORO	13	621	605	634	629	624	613	597	591	605	583	575	538		575	562	551	562	586	608	613	637	648	629
FIANARANTSOA	14	583	570	594	597	583	559	535	546	562	570	556	562	575		484	527	567	562	594	599	637	624	640
ANTANANARIVO	15	589	565	599	602	591	562	551	554	567	573	551	554	562	484		516	559	556	589	599	634	626	640
AMBATONDRANZAKA	16	602	583	621	624	613	591	567	565	581	570	565	543	551	527	516		567	573	599	608	637	640	632
FENOARIVO-EST	17	608	597	618	616	618	605	589	583	586	597	578	575	562	567	559	567		567	535	559	602	602	602
MAJUNGA	18	597	578	597	599	597	583	573	562	594	594	586	559	586	562	556	573	567		573	578	578	581	586
MANDRITSARA	19	610	613	621	618	624	608	597	597	616	610	602	597	608	594	589	599	535	573		530	589	578	583
ANTALAHA	20	629	618	618	624	621	621	608	591	597	602	594	599	613	599	599	608	559	578	530		581	583	594
AMBANJA	21	648	640	640	634	645	640	642	624	640	640	642	624	637	637	634	637	602	578	589	581		527	527
AMBILOBE	22	640	624	634	637	637	637	632	624	651	651	645	637	648	624	626	640	602	581	578	583	527		530
VOHEMAR	23	629	634	634	640	634	637	629	618	626	637	648	624	629	640	640	632	602	586	583	594	527	530	

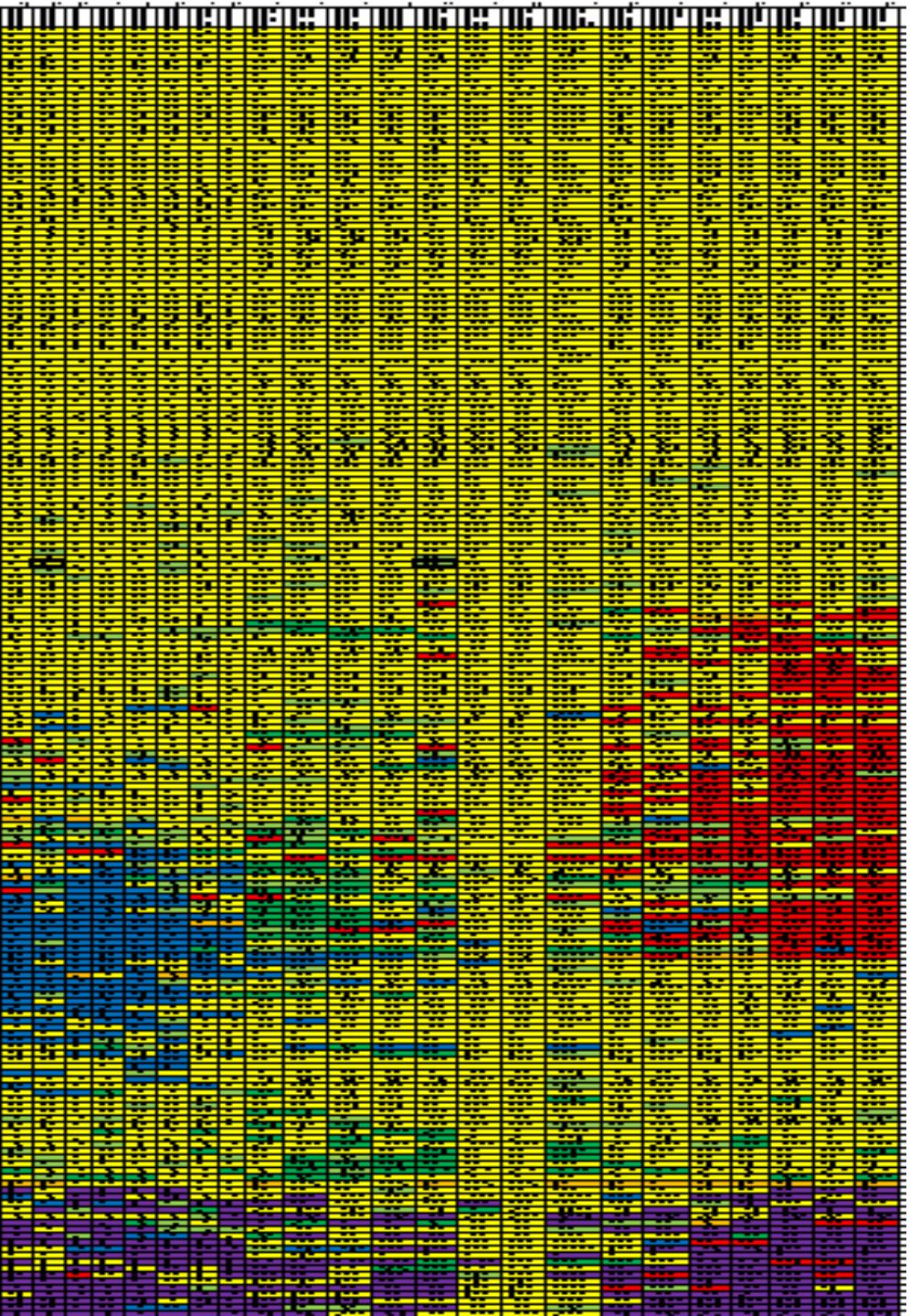
372 Cognate Sets List

We can further understand our Malagasy language data by carefully analyzing lexical distance data.
I generated this matrix using a table of 372 cognate sets derived from 198 word meanings for 23 dialects of the Malagasy language. Each entry is 1000x the lexical distance between the language in its row and the language in its column. The locations and number assigned to each dialect are shown on the left.

Multi Dimensional Scaling Plot for Malagasy Dialects



As with the Polynesian languages, I converted the matrix into lexical distances and performed a 3D multidimensional scaling analysis of it. Plotting the first two principal components gives us an important insight: There appear to be 4 identifiable zones of similarity for the Malagasy dialects. And remarkably, this 3D map of lexical distances seems to somewhat replicate the numerical order in which the languages appear on the map of Madagascar.



Serva's Lists Analyzed for Clusters of Cognacy

76% Cognate to Merina

5% Individually non-Cognate toMerina

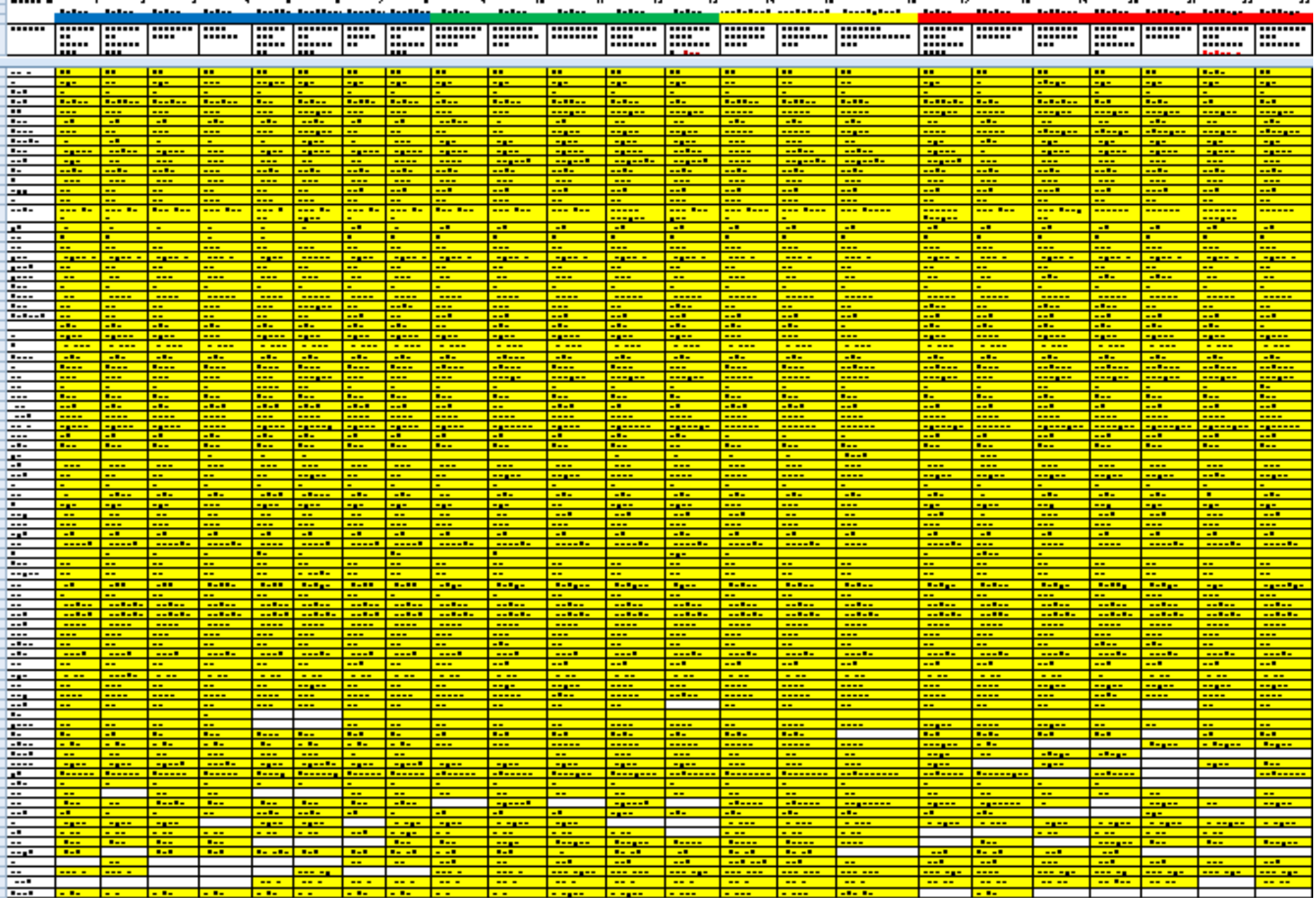
5% in Southern Clusters

5% in Northern Clusters

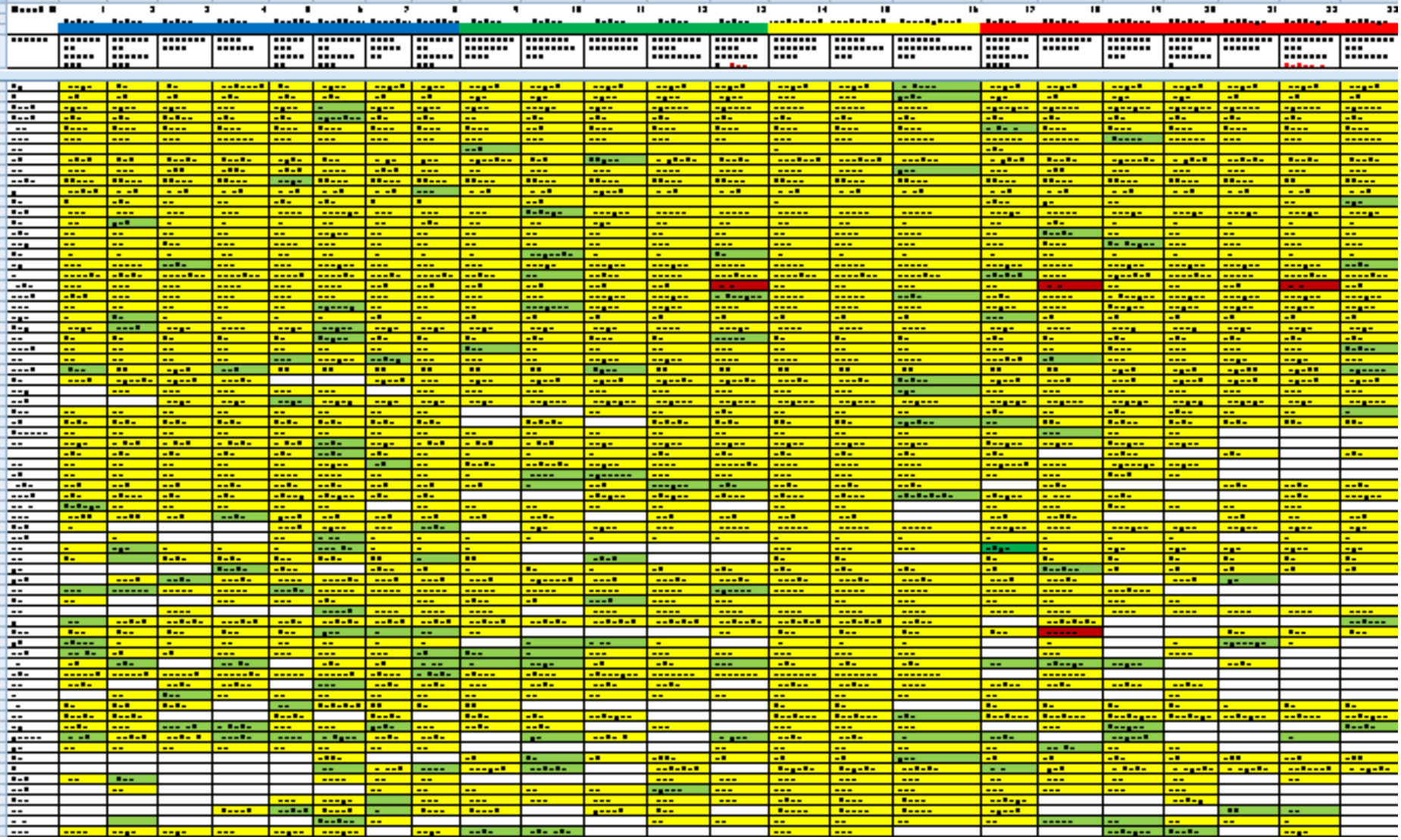
6% both North & South

3% in Central Clusters

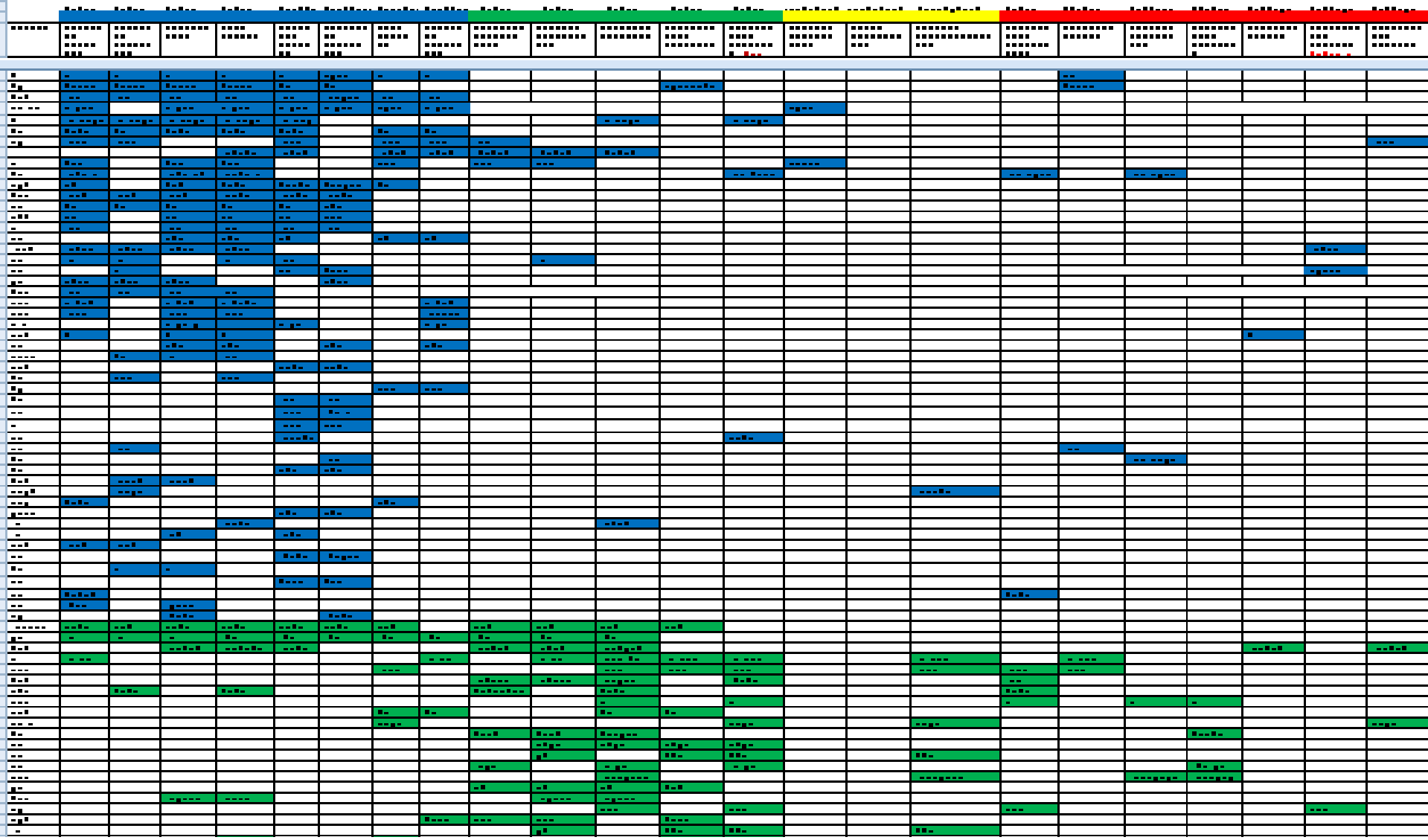
To really understand what's going on, we need to look at the words themselves, in gory detail.
Here, the 23 dialects are arrayed in columns and the 200 words in rows. I've sorted them into categories of cognacy and color-coded them as indicated.



An edited version of the table separates cognates into categories for easier interpretation.
At the top of this edited table we see a field of yellow words. These are words all cognate to the Merina dialect



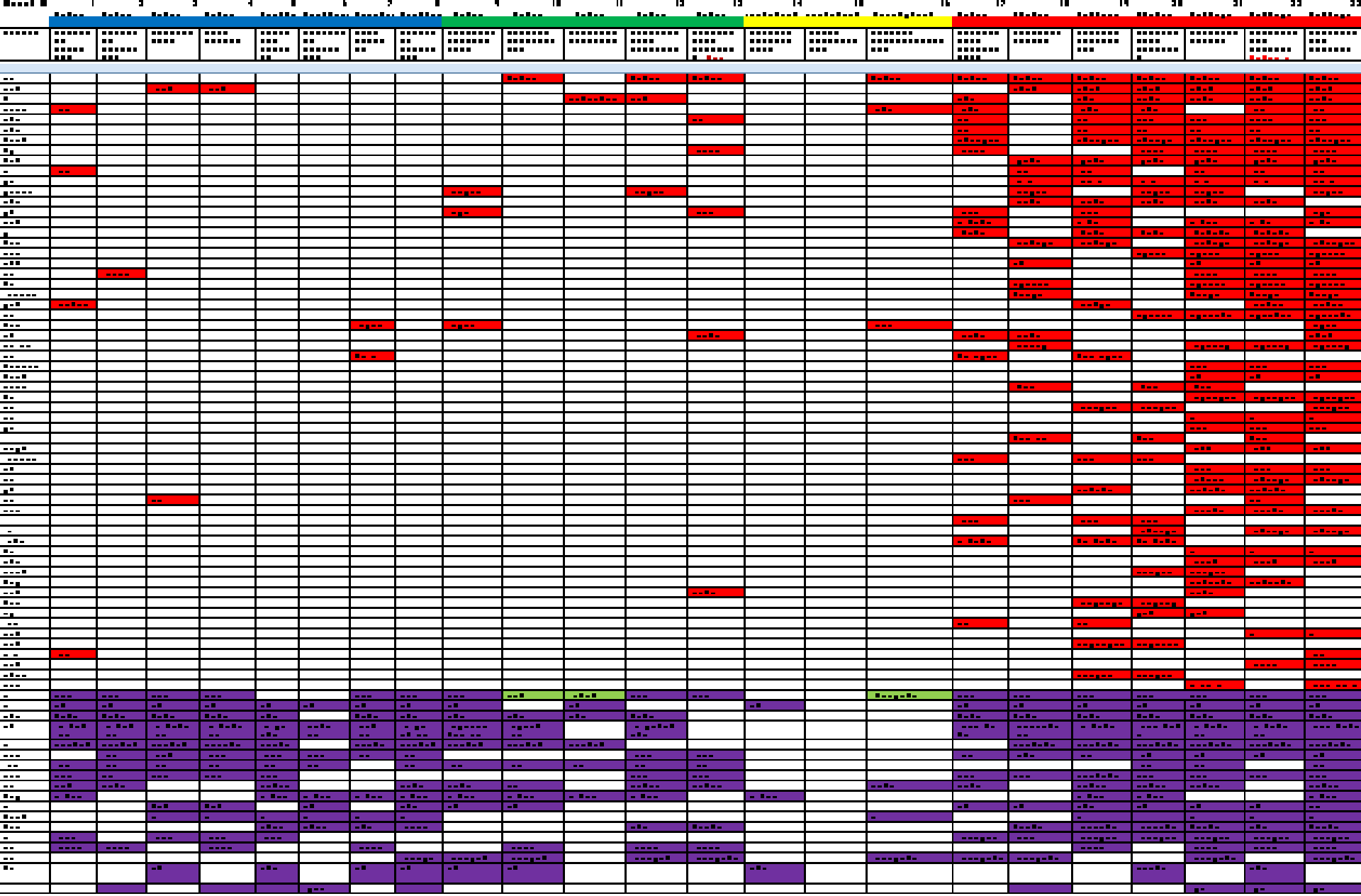
Further down the table we see the yellow words have been replaced in many locations by *Innovations* – words not cognate to Merina that occur in ONE DIALECT ONLY. I've shown these in light green. These innovations are not shared with surrounding dialects so they appear to have been spontaneously generated. There are all told 199 of these unique innovations. If we assume the Malagasy language has been on the island for 1200 years, this translates to one unique innovation per dialect every 139 years, on average. This is analogous to a genetic *mutation rate.*



Here we see the main reason for the regional clustering in the MDS visualization of the Malagasy Matrix: words not cognate to Merina and with high affinity to a geographic region.

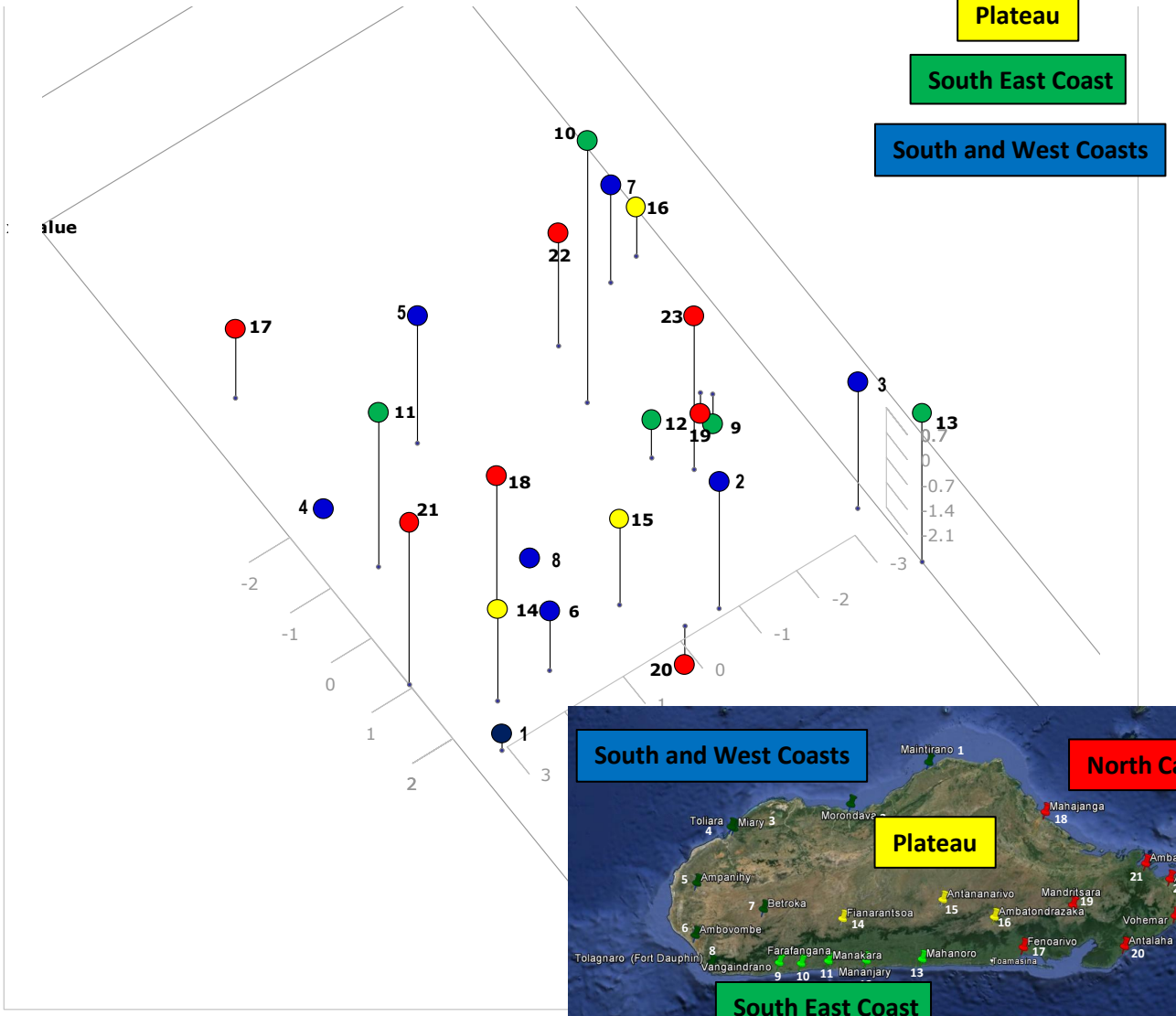
Here we see the blue words and green words. The blue words are non-cognate to Merina and are very popular in the south and west of the island, and much less so elsewhere.

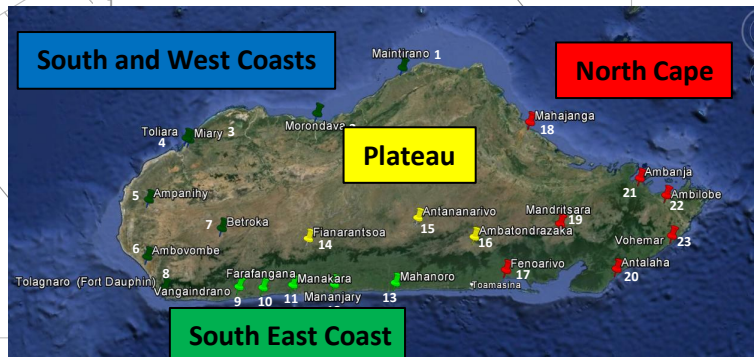
The green words concentrate mostly in the southeast coastal dialects, but some of them are also popular to the north and south.



The red words are heavily concentrated – but not exclusively - in the north of the island.
But the purple words are popular throughout the coastal region of Madagascar and have even crept into the non-Merina Plateau dialects.

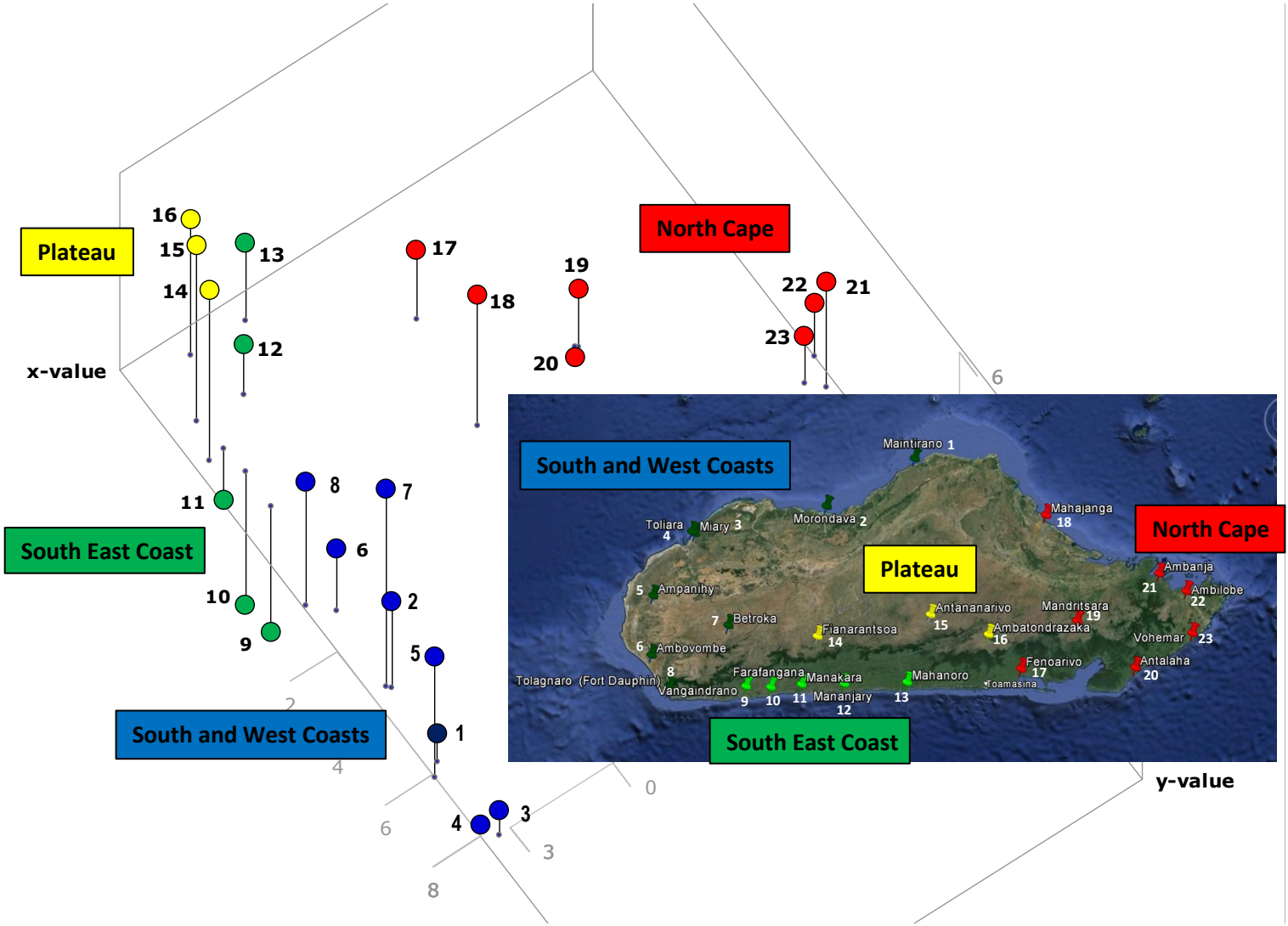
Multi Dimensional Scaling Plot for Malagasy Dialects With Regional Words Omitted



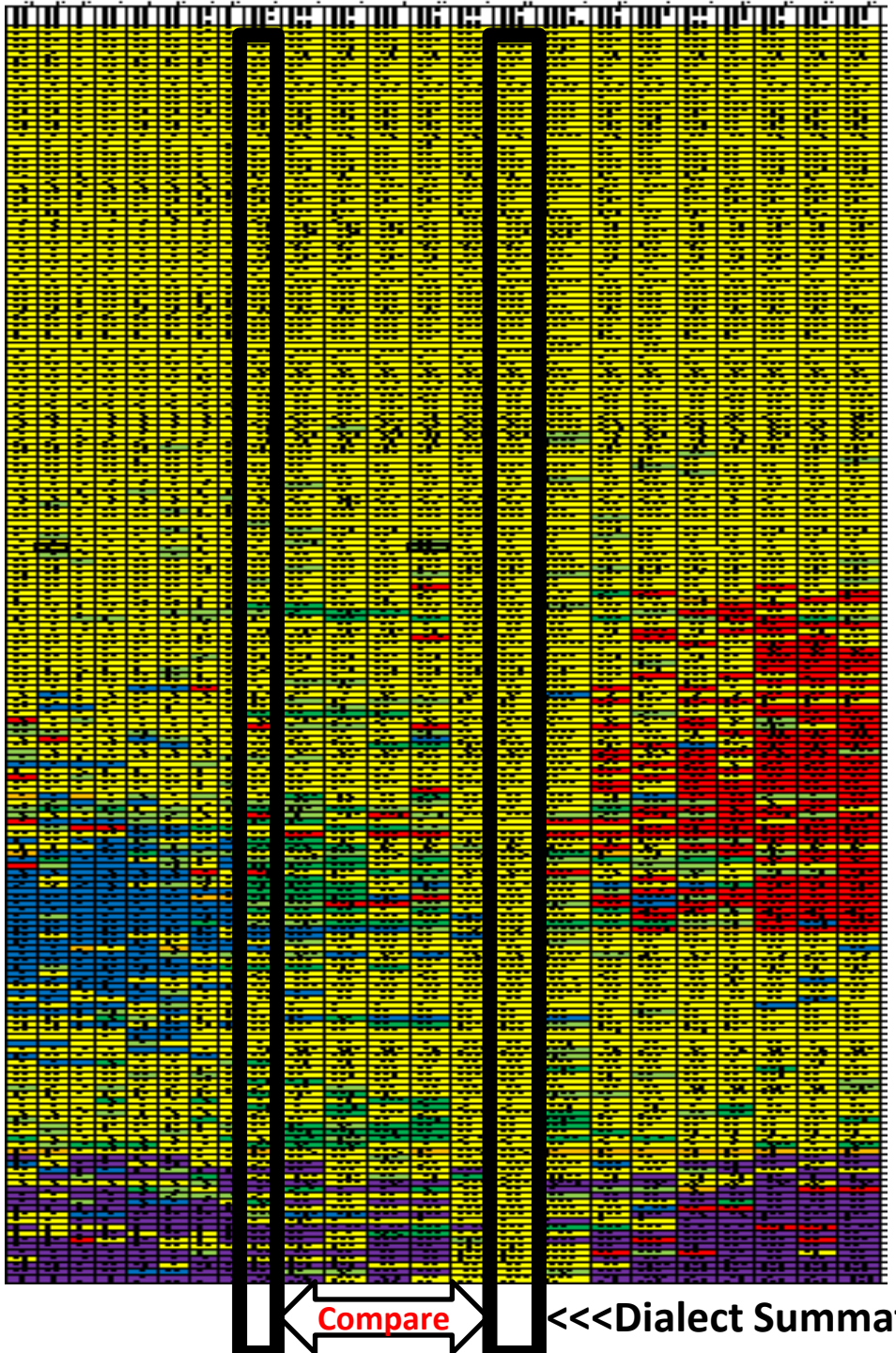


When we re-generate the Malagasy Matrix with the Regional Words Omitted, we can immediately see what was responsible for the order in the original MDS plot. It's those regional words! Without them, we have no information about the regional variation of the lexical distances.

Multi Dimensional Scaling Plot for Malagasy Dialects

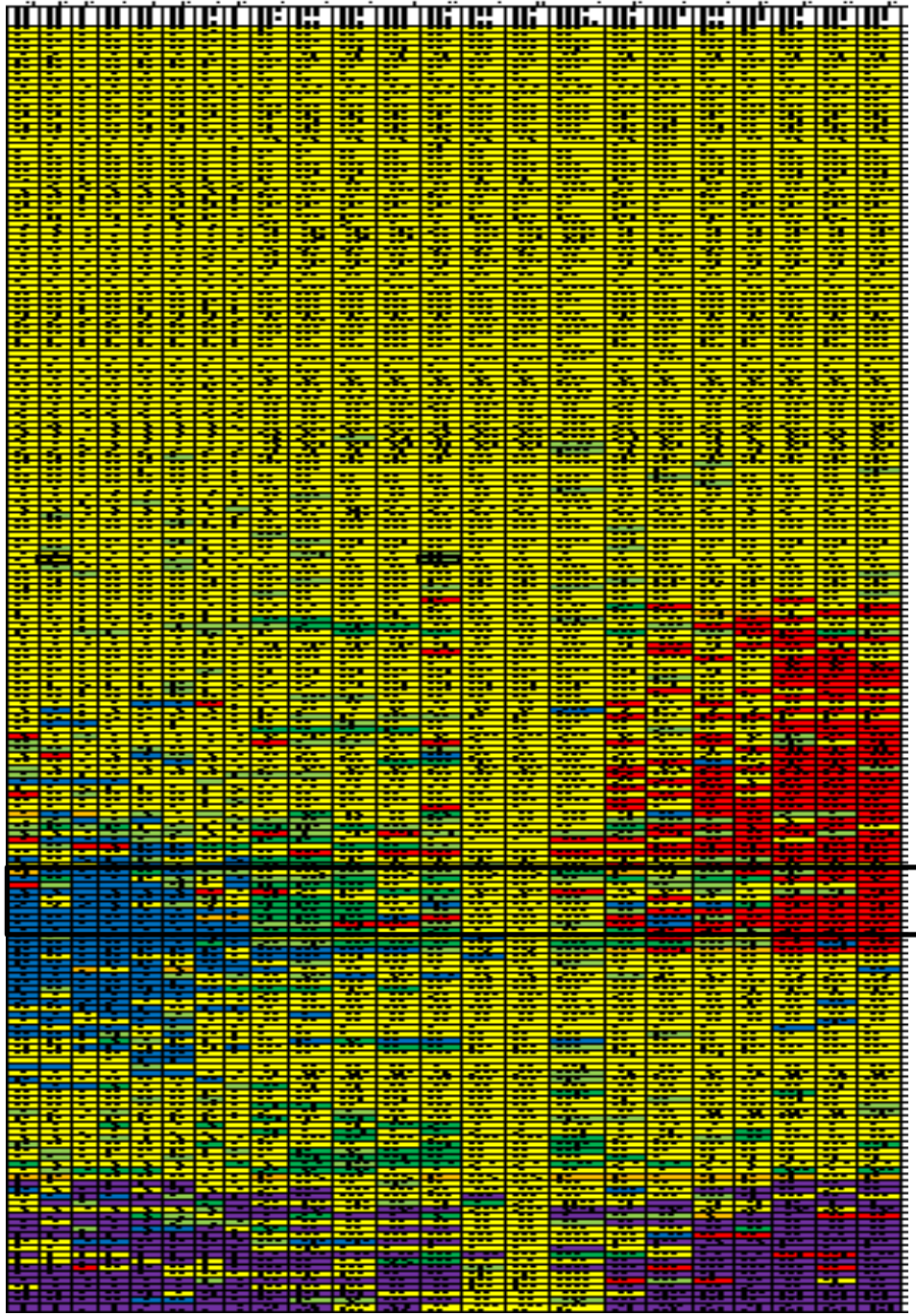


For comparison: here is the MDS plot with regional words included. Clearly, they are the determining factor in the shape of this MDS map.



Method 1: Lexical Distance Between Dialects

We have two basic ways to slice the data: Vertically and Horizontally.
The vertical method, the one we've discussed so far, counts the number of cognate differences between dialects to obtain a single number for each dialect pair: the Lexical distance.



Method 2: Compare Across All Dialects on a Word-by-Word Basis

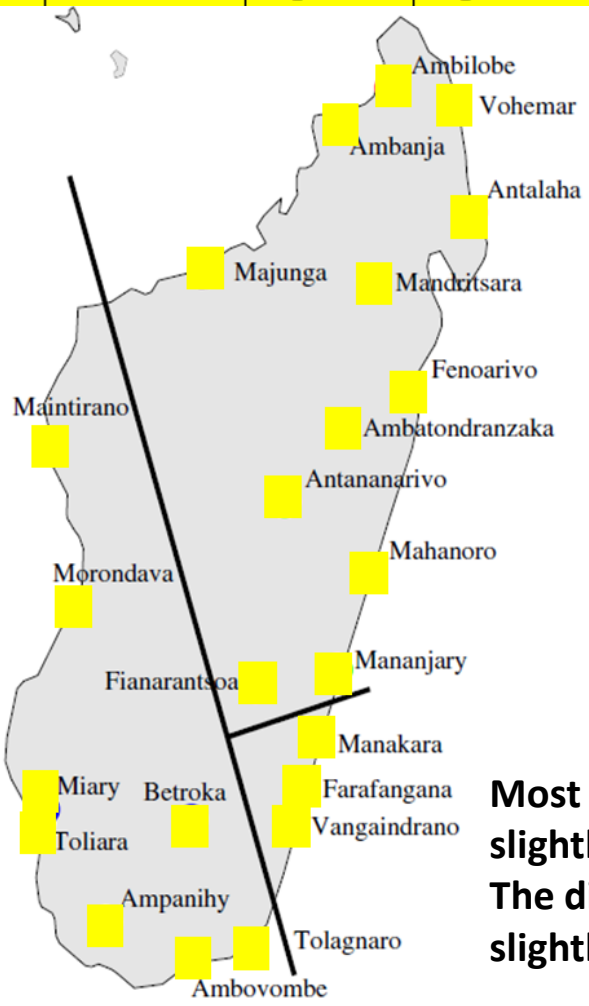
<<< Track Each Word Across All Dialects

The second method, the one we'll be discussing now, takes a horizontal slice through the data and studies the way each word changes from dialect to dialect across all 23 dialects.

Case 1: Minor Changes Between Dialects

Merina ⇩

mamono	mamono	mamono	mamono	mamono	mamono	mamono	mamono	mamono	mamono	mamono
velo	velogne	velo	velo	velo	velogny	velogna	velogna	velogno	velona	velona
atene	ate	aty	aty	aty	aty	aty	aty	aty	aty	aty
fitava	vava	vava	vava	vava	vava	vava	vava	vava	vava	vava
agnare	agnaragne	agnara	agnara	agnara	agnarana	agnara	agnarana	agnaragna	anarana	anarana
lala	lalagne	lala	lala	lala	lalagna	lalagna	lalana	lalana	lalana	lalana

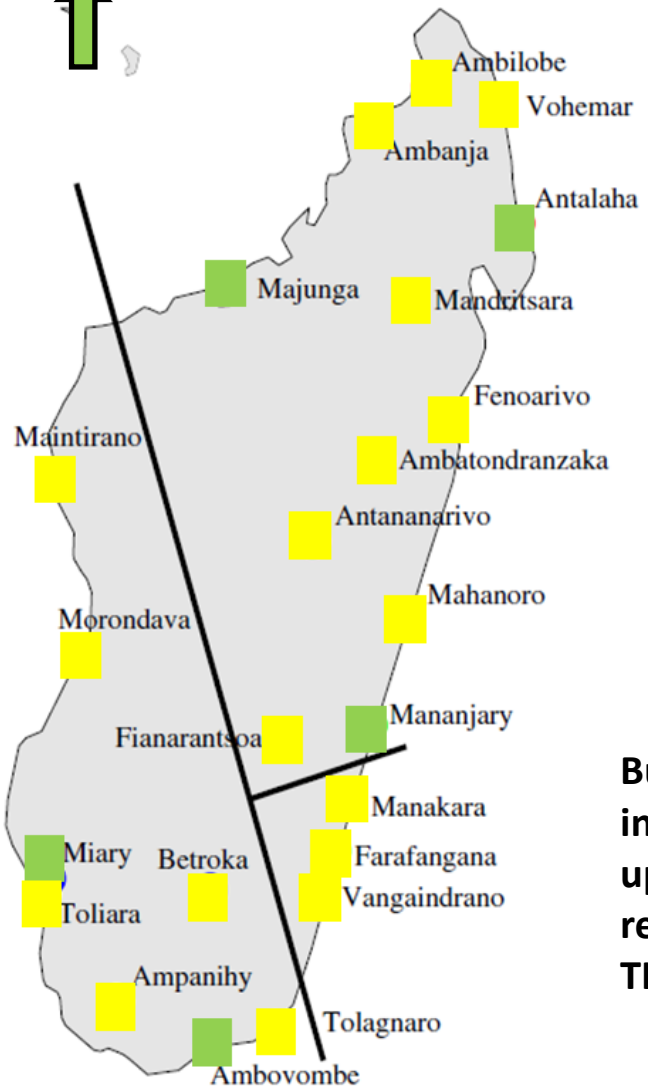


Most of the words vary not at all, or only slightly, across the island.
The differences that have evolved result in slightly different pronunciation, that's all.

Case 2: Independent Innovations…No Sharing

Merina ⇩

oro	fagnatsogne	oro	oro	oro	vatagnoro	orogna	orogna	orona	orona	orona
taly	taly	tady	taly	tady	tady	tady	tady	tady	tady	tady
mate	mate	maty	maty	maty	mangatsiaka	maty	maty	diso	maty	maty
ala	ala	ala	ala	hazo	ala	ala	ala	atiala	ala	ala

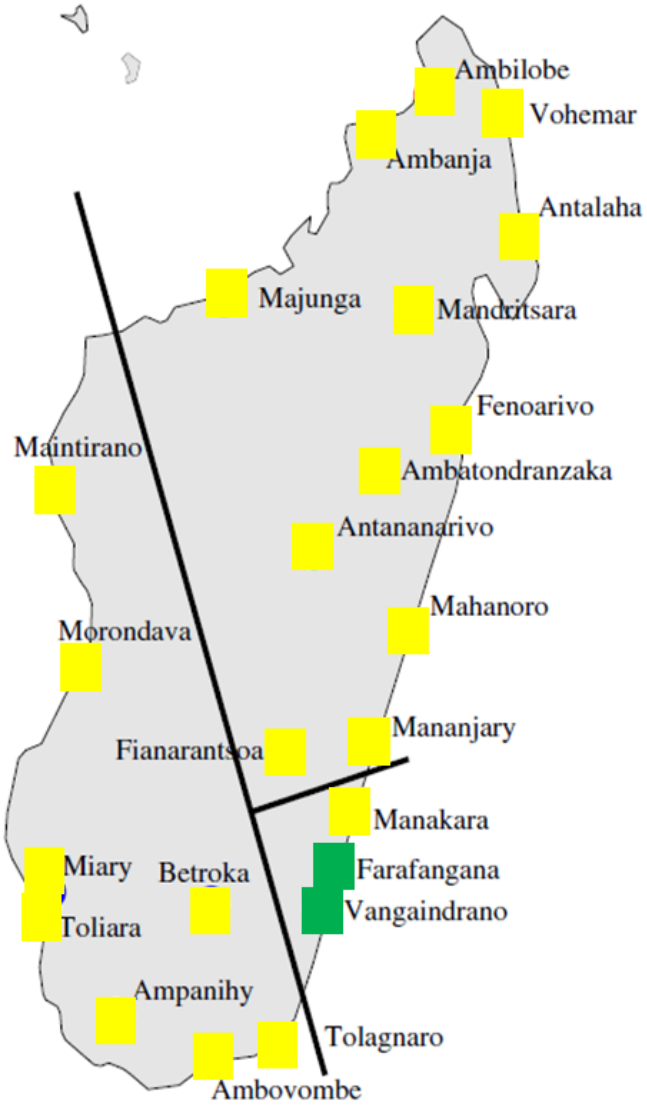


But in every dialect you find innovations, words that just pop up seemingly from nowhere, and replace the standard word IN THAT LANGUAGE ONLY.

Case 3: Independent Innovations with Local Sharing

Merina ⇩

rahongne	rahogne	raho	raho	mika	mikagny	rahogna	rahogna	rahona	rahona	rahona

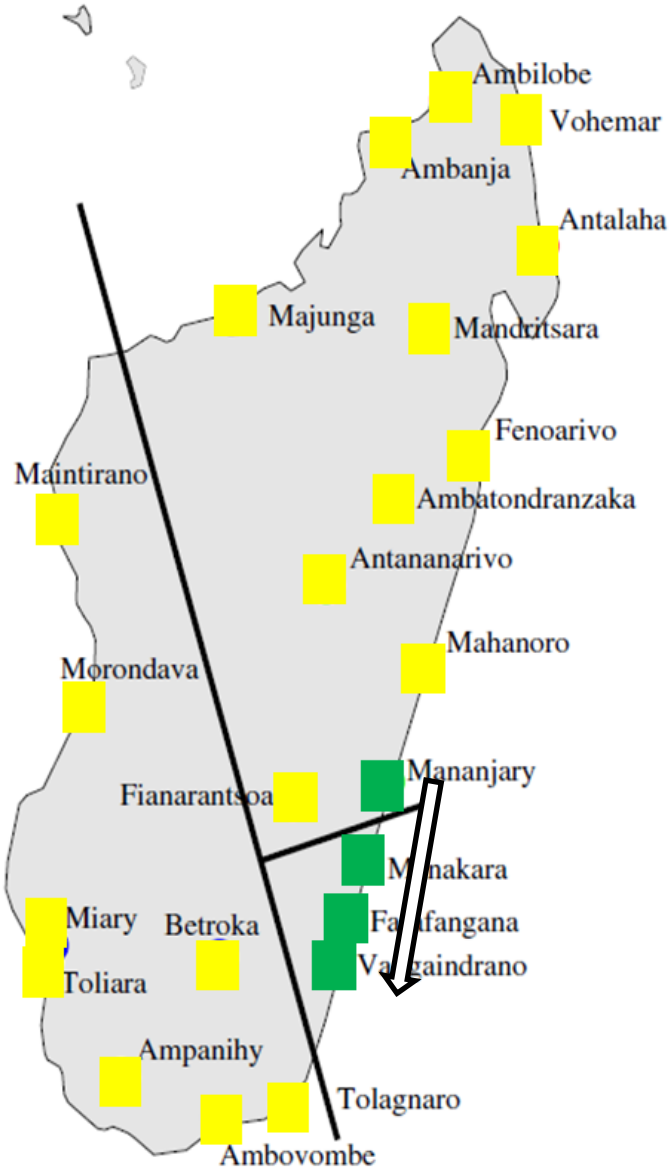


Sometimes, these local innovations are shared with a single neighbor language. But this is pretty rare, occurring less than 1% of the time.

Case 3: Independent Innovations with Local Sharing

Merina

tsinay	sinay	oliky	oliky	oliky	tsikoliky	tsine	tsinay	tsinay



And sometimes, these local innovations start a trend that continues on down the line.

First, TSINE gets elaborated into TSIKOLIKY. Only the first syllable is retained.

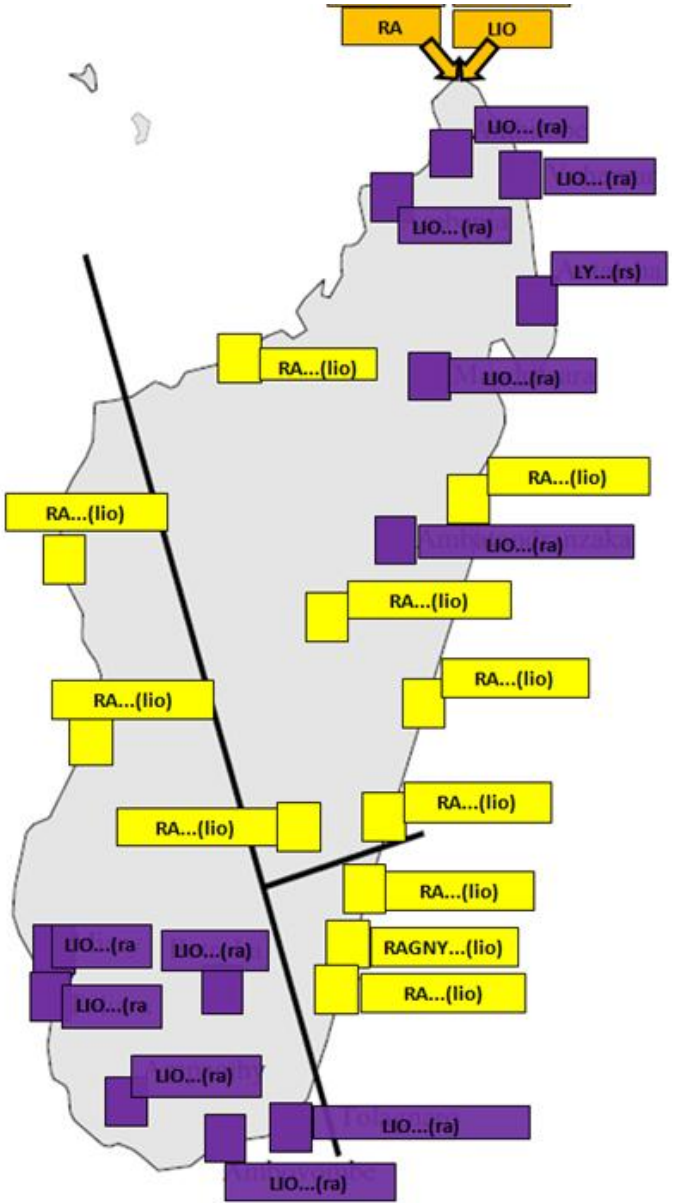
Then, as the word moves south, the first syllable is abandoned altogether, leaving just OLIKY.

But the original word TSINAY dominates outside the local area on the east coast where that first innovation occurred.

Case 4: Shared Substitutions

Merina ⇩

lio	lio	lio	lio	ra	ragny	ra	ra	ra	ra	ra

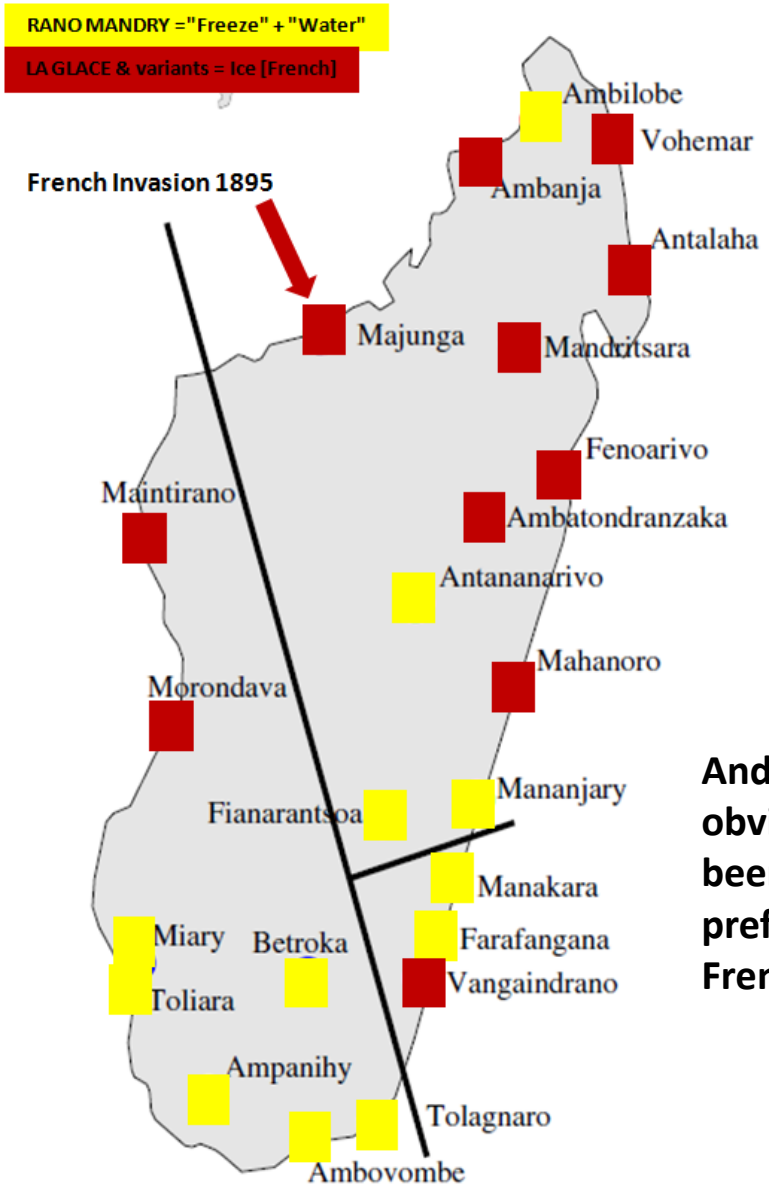


Frequently, it appears that some words travel together as synonyms, and change places in popularity along the way, as in the case of RA and LIO, competing words for "Blood".

Merina

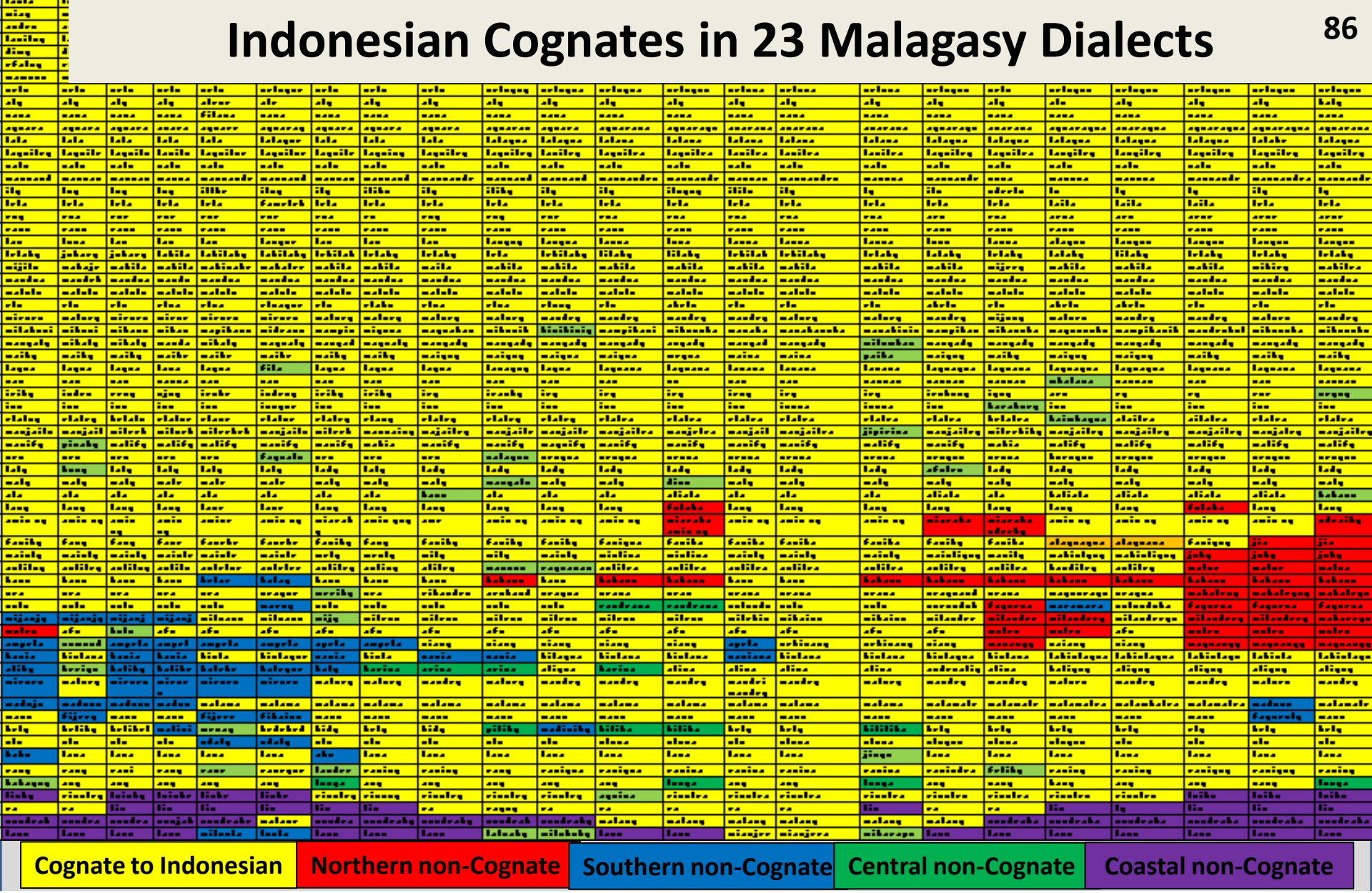
Case 5: Imported Words…"Borrowings"

rano mandry	glasy	gilasy	la glace	lagilasy	gilasy	lagilasy	ranomandry



And in a couple of cases, it's obvious that a foreign word has been taken up and used preferentially, as in the case of the French word for "Ice"

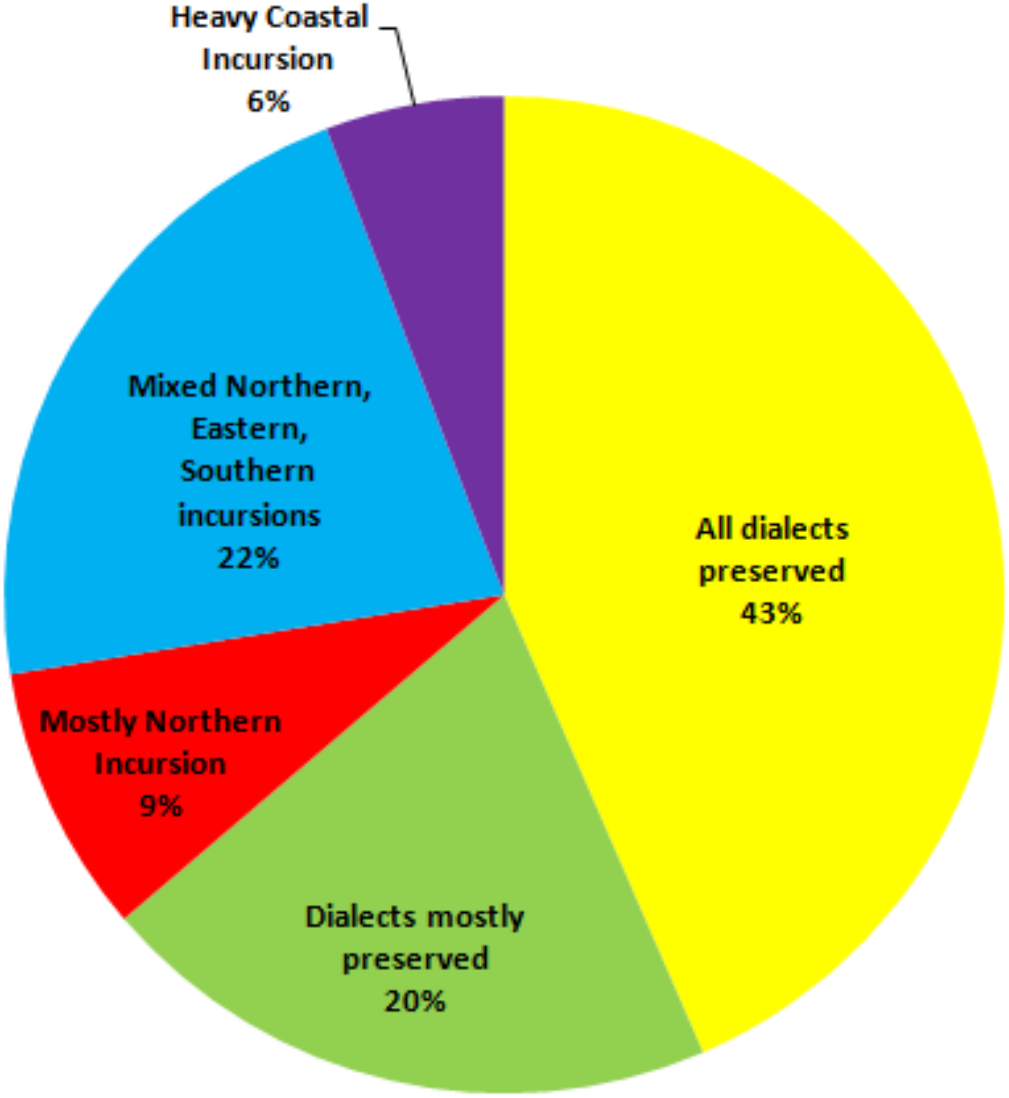
Indonesian Cognates in 23 Malagasy Dialects



Cognate to Indonesian | Northern non-Cognate | Southern non-Cognate | Central non-Cognate | Coastal non-Cognate

Here is our word table pruned down to contain only the 69 words which are known to be cognate to an Indonesian word. The colors other than yellow indicate these words are not cognate to the Merina dialect but are cognate to one of five Indonesian languages studied.

Distribution of Indonesian-Derived Words

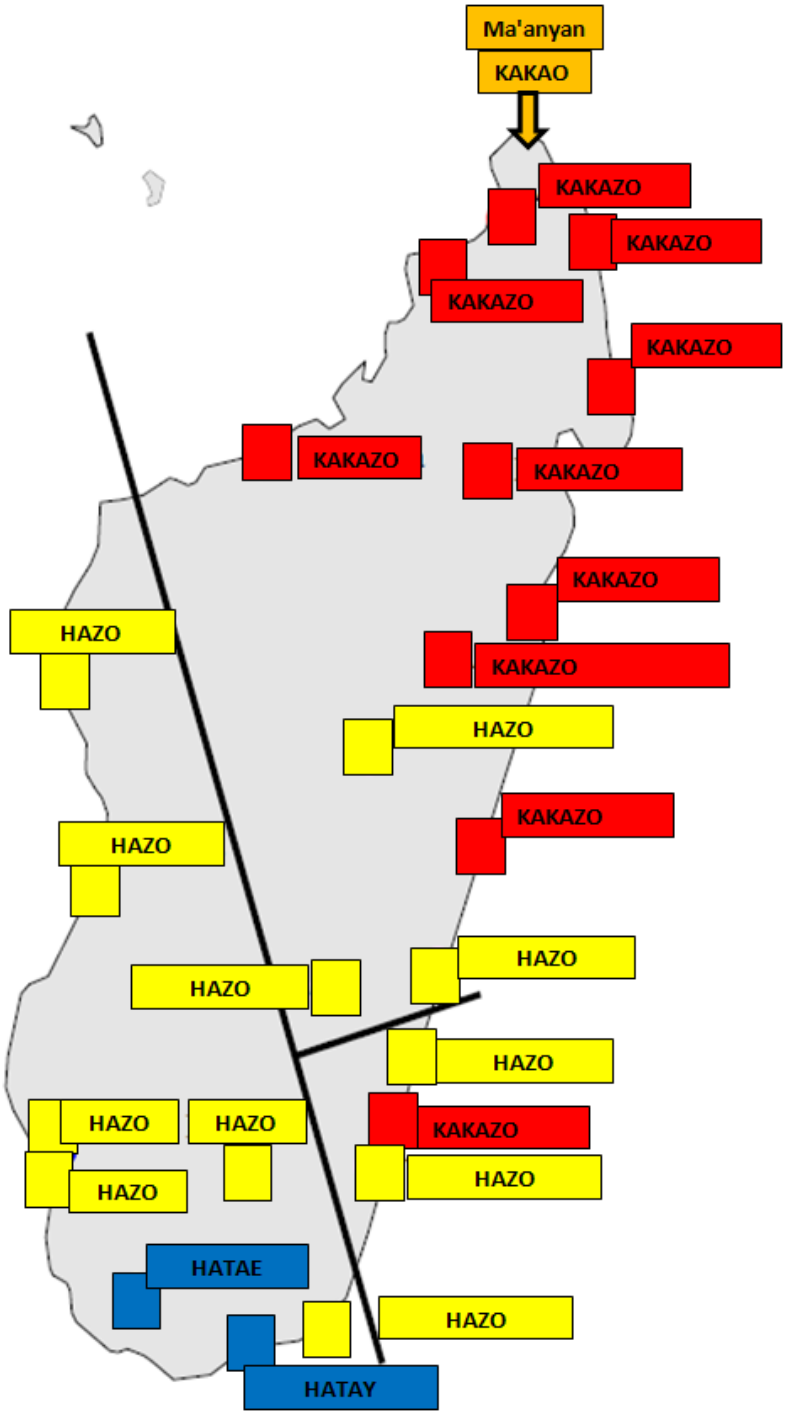


43% of these words survived with only minor variations in all dialects An additional 20% survived in most dialects, with only occasional local innovations. The remaining 37% were subjected to widespread substitutions, some from Bantu but for the most part from unknown sources.

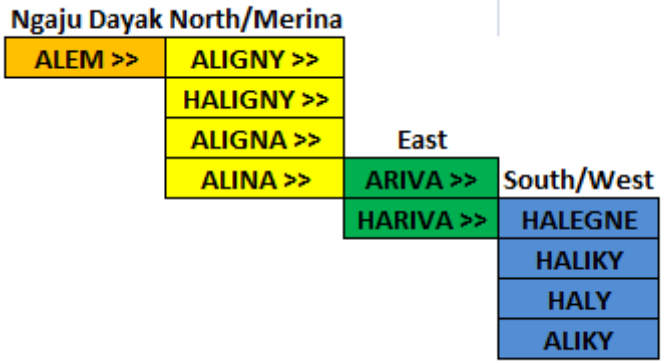
Evolution of a Word: KAKAZO (Tree)

Ma'anyan	North	Merina	South
KAKAO >>	KAKAZO >>	HAZO >>	HATAY

Consider the Ma’anyan word for “Tree”. In the north of the island it is rendered as KAKAZO. In the east coast and central highlands it was shortened to HAZO. In the far south it morphed in HATAY. Could this have happened in reverse order, with HAZ0 morphing into KAKAZO, a word so obviously cognate to the Ma’anyan KAKAO? Not likely! So support for a North-to-South migration hypothesis seems to come from the Ma’anyan word KAKAO, meaning “Tree”.

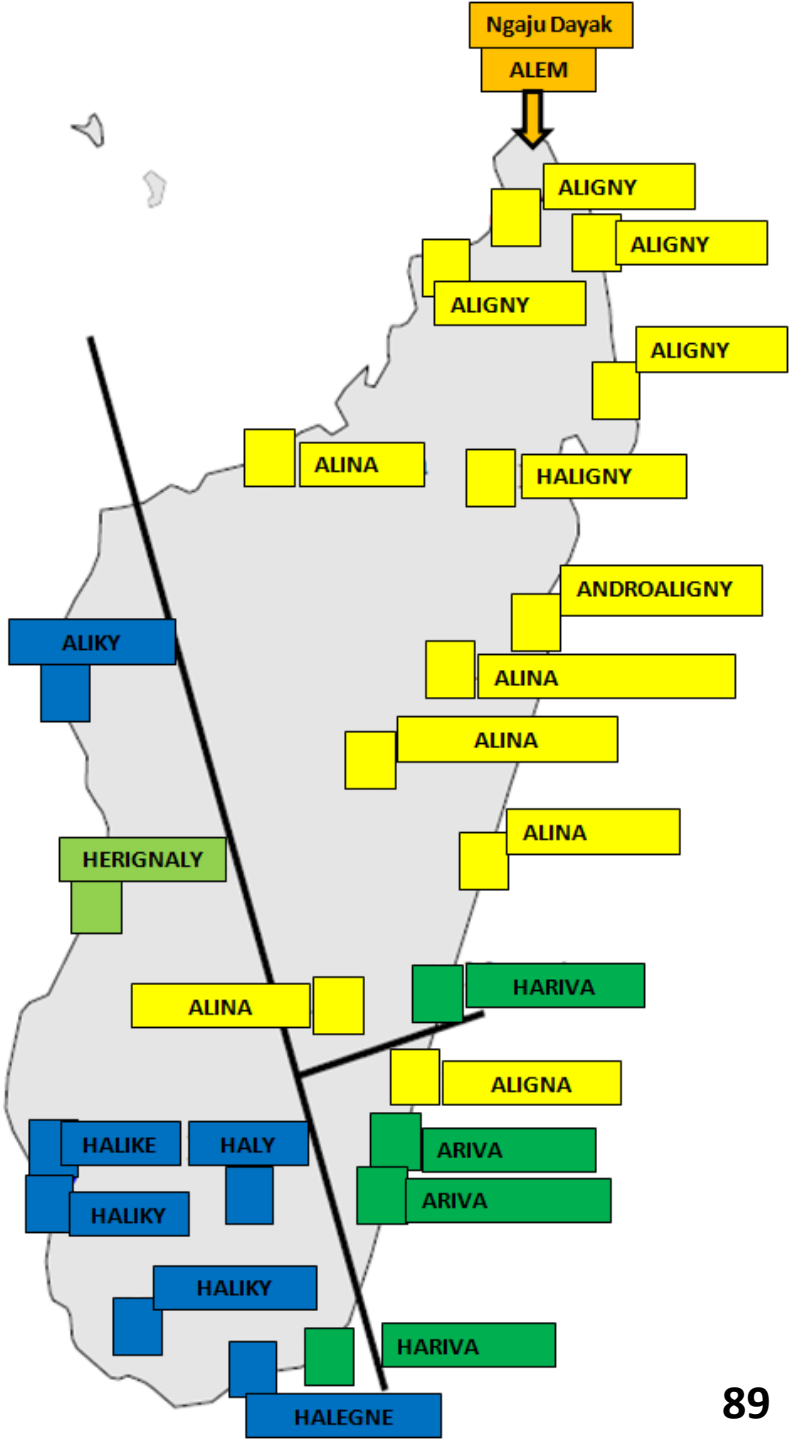


Evolution of a Word: ALINA (Night)



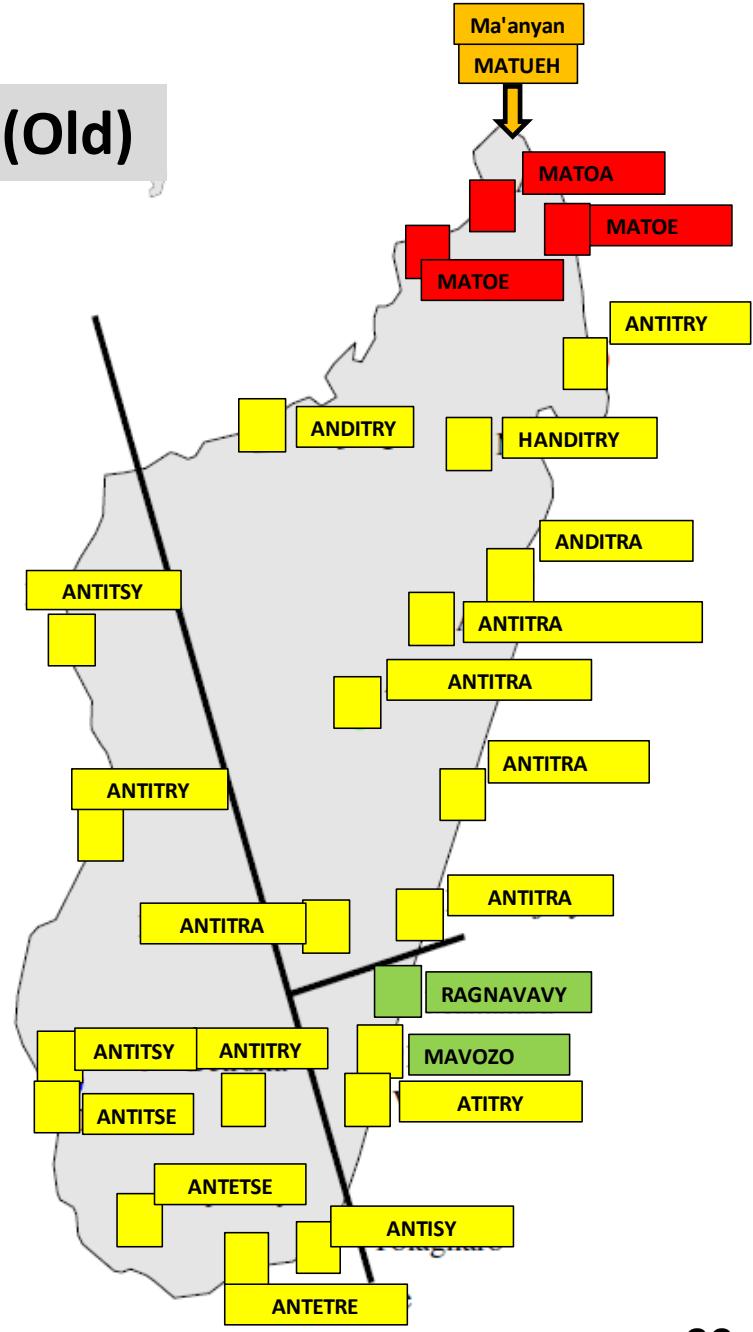
Another case is that of the word ALEM from Ngaju Dayak, meaning "Night". It continually evolved as it moved south, eventually becoming ALINA in Merina. It further evolved along the east coast into ARIVA, finally becoming ALIKY on the west coast after a series of stepwise changes.

However, once again we have an ambiguous piece of evidence. The word could have entered on the east coast and propagated unchanged northward and -in a changed form- southward.



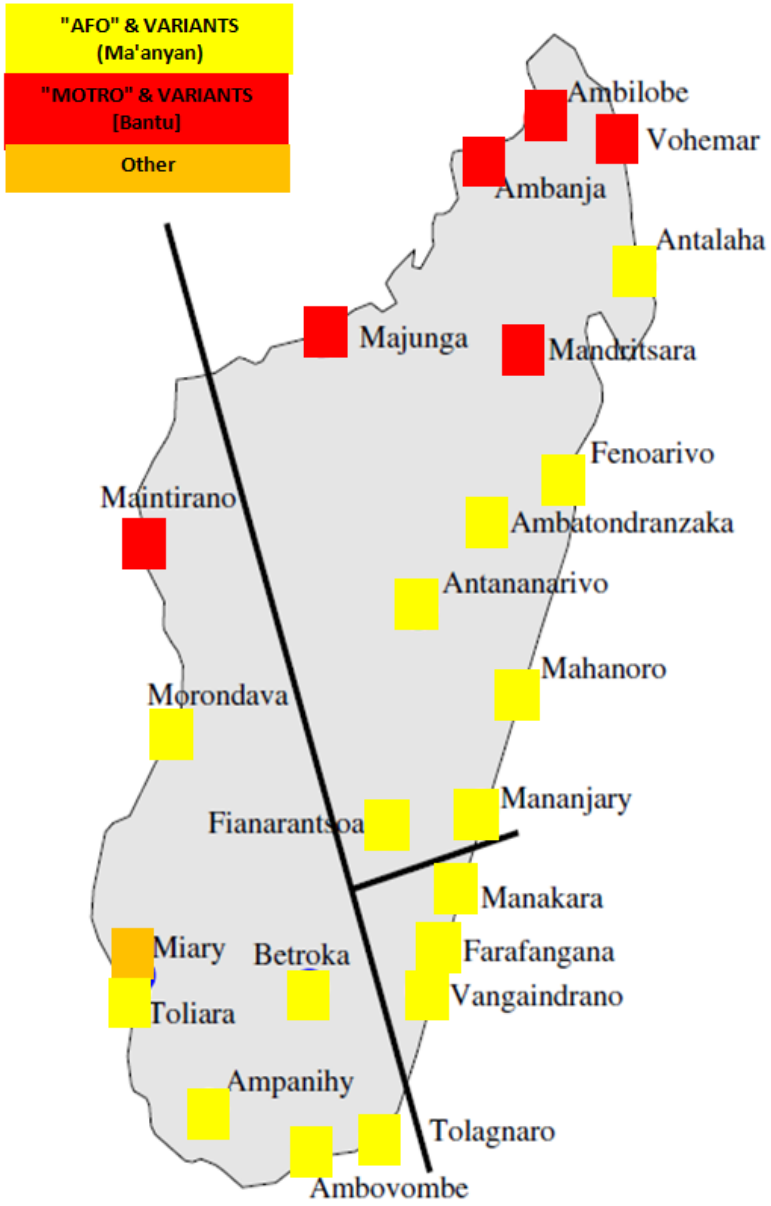
Ma’anyan Word Unique to North: MATEUH (Old)

Here is the strongest evidence I’ve found for immigration from the north. The Ma’anyan word for “Old” is MATUEH, and cognates to it are found only in the three northernmost dialects. All other dialects use ANTITRY, or something similar. So either there was a cognate to ANTITRY we were unaware of in the Proto language, or (more likely) there was an innovation that changed MATOE to ANTITRY, which then caught on and propagated throughout the rest of the island.

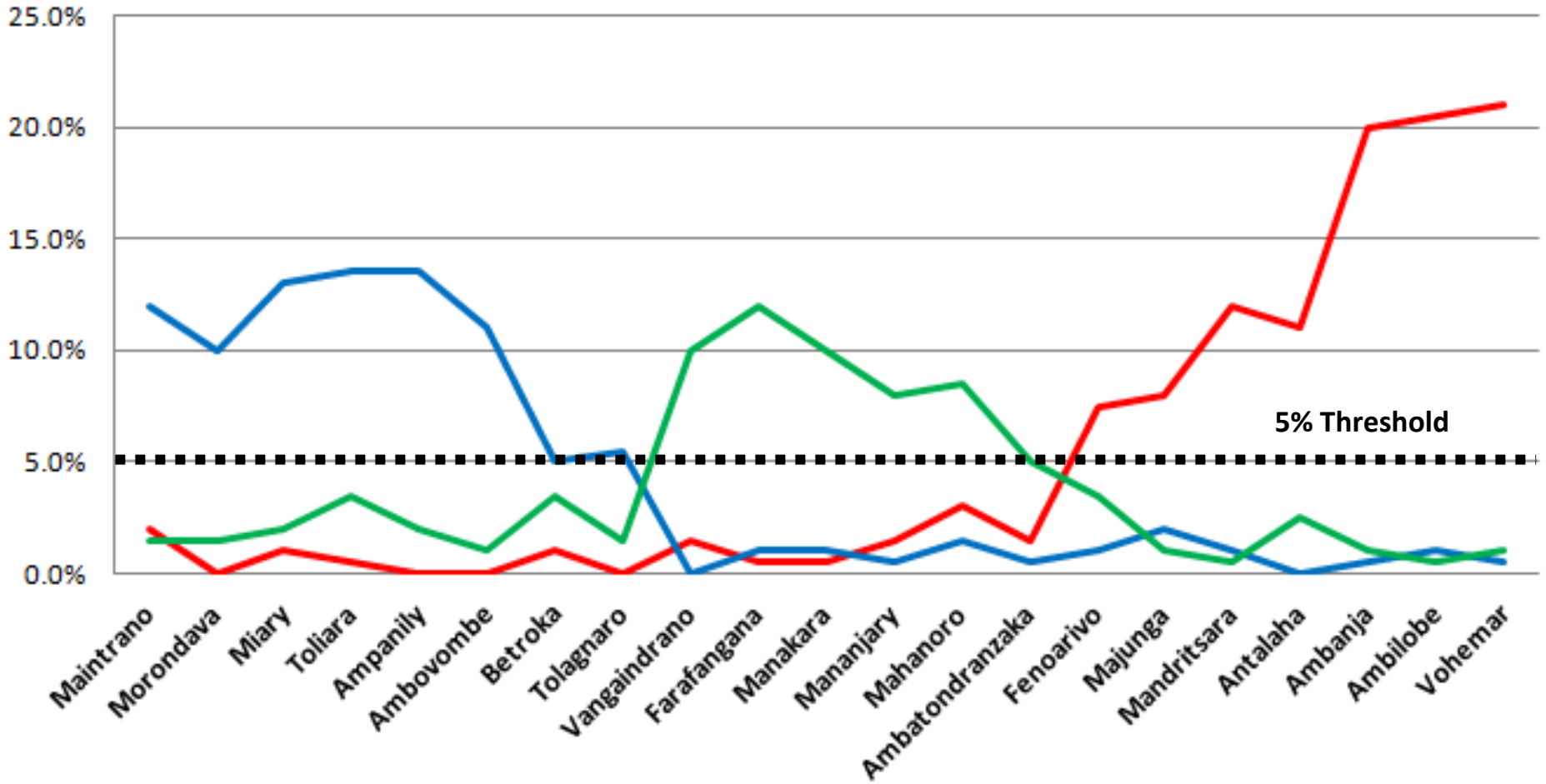


Bantu Influence: “Fire”

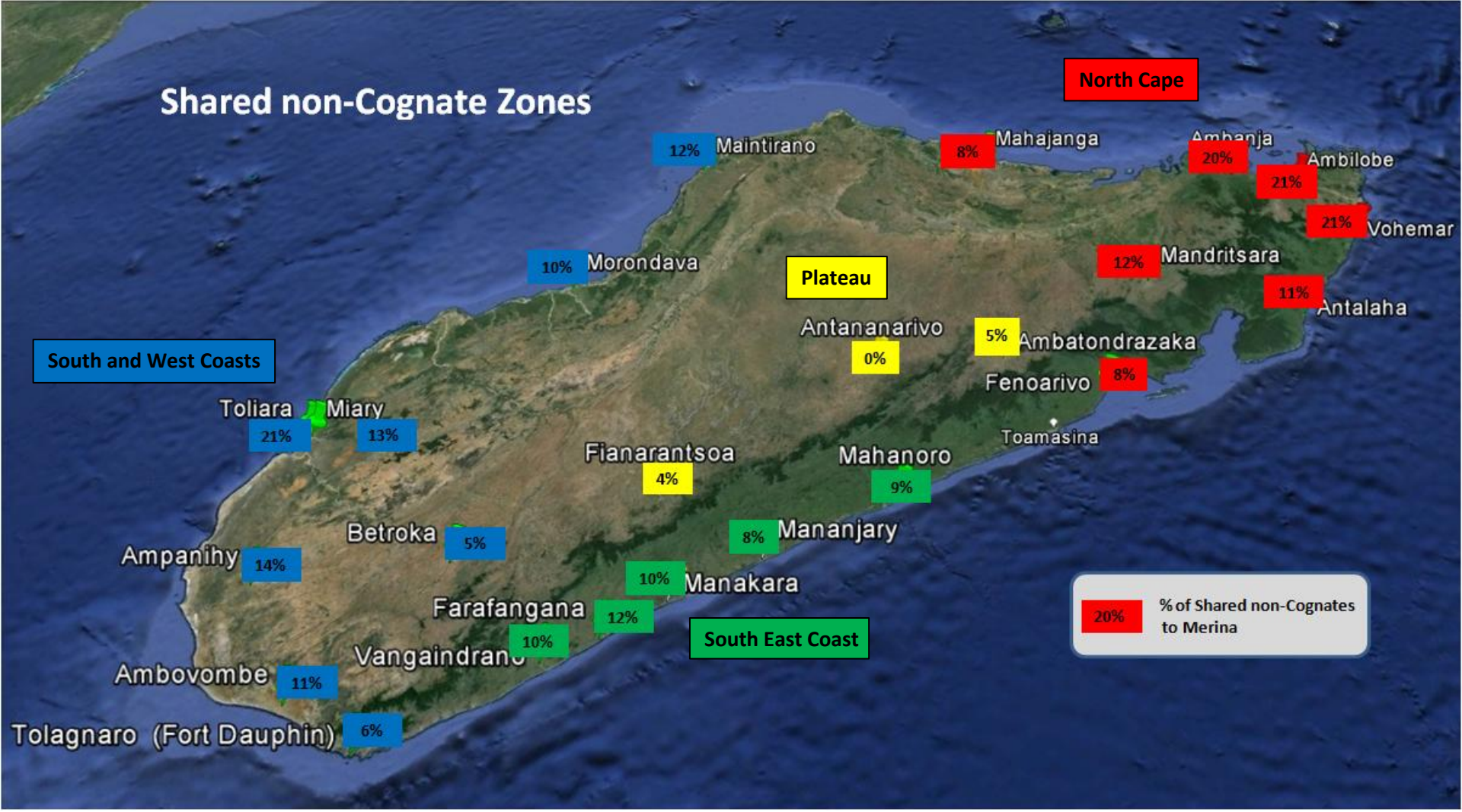
Here is another case of a word with cognates unique to the northern dialects.
The Bantu word for “Fire”, MOTO, became MOTRO in the northwest of the island.
AFO, is the Ma’anyan cognate for “Fire” is used in most other dialects.
It’s possible that the Bantu influence was a later introduction, due to later immigration or trading.
But it’s just as likely it was introduced by the original Bantu settlers, and if so argues for a northern point of entry.



Southwest, Eastern and Northern Clusters of Shared Words non-Cognate to Merina

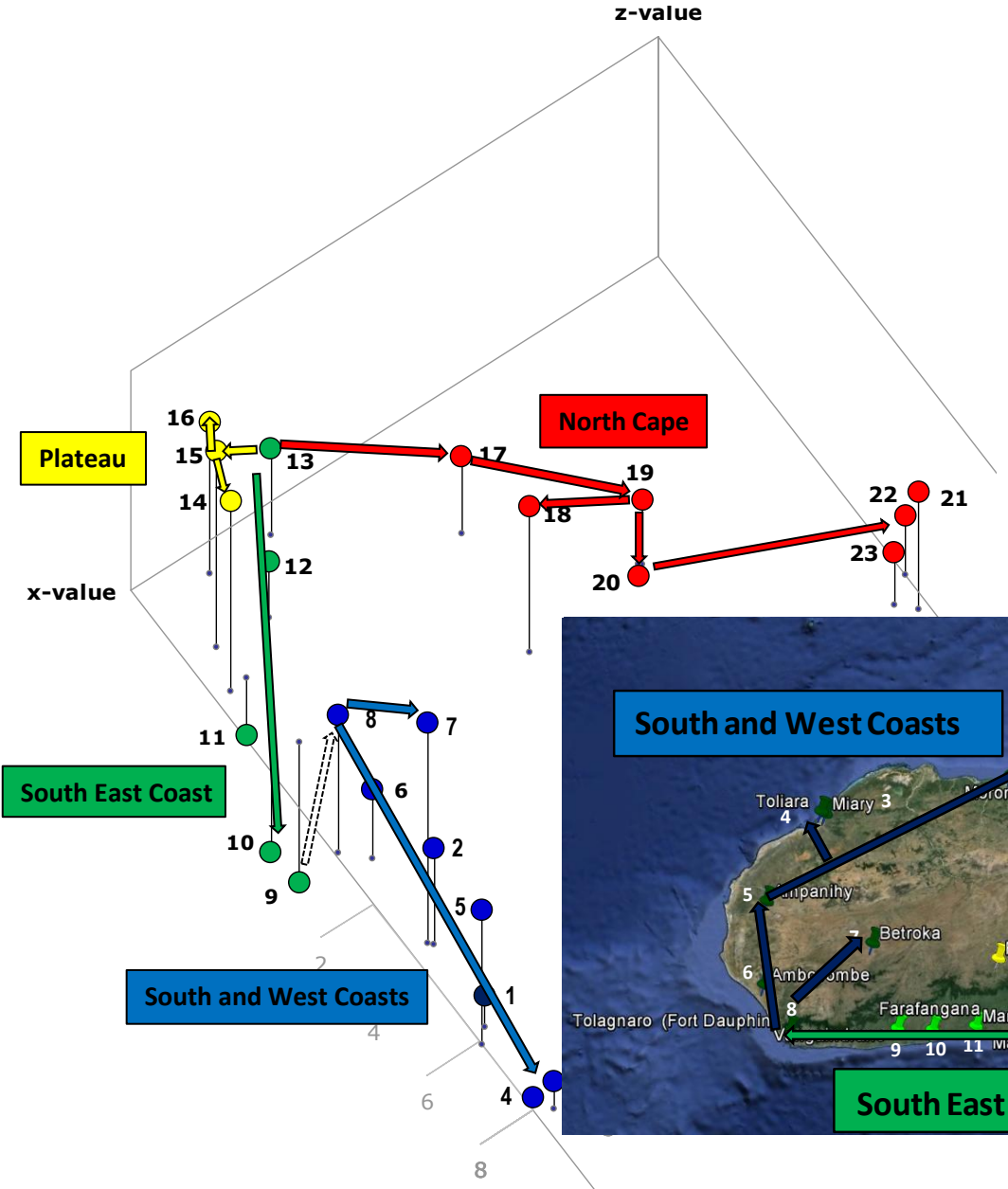


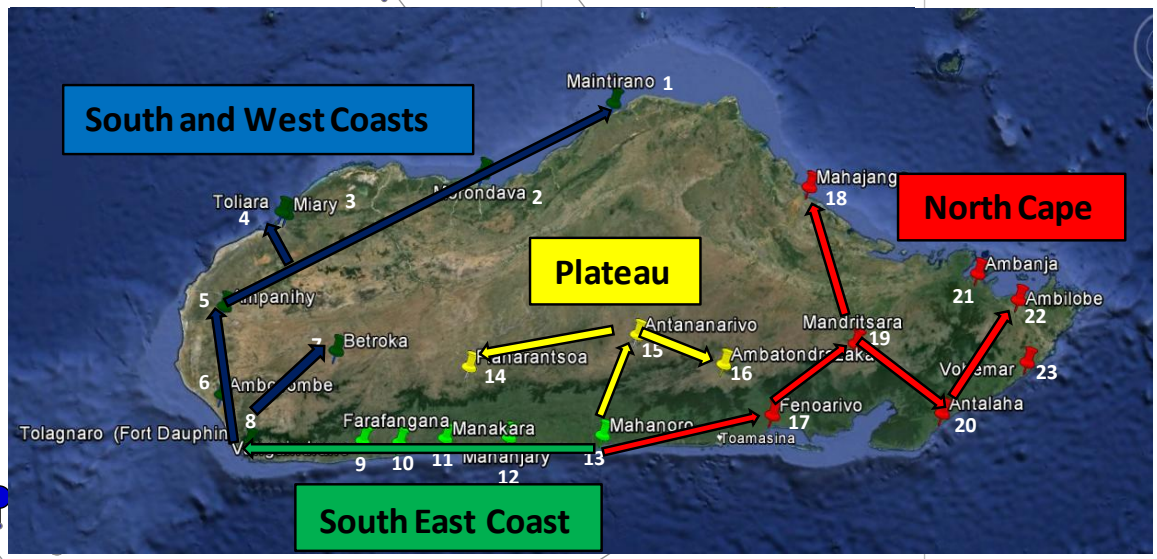
When we plot the frequency of these regionally-clustered words versus location, a pattern begins to emerge. If we set a threshold of 5% as a definition for the regional boundaries, we can justify the lumping of these words into regions.



So the four zones we first noticed in the MDS plot are reaffirmed by analysis of the frequency of occurrence for words not cognate to Merina.

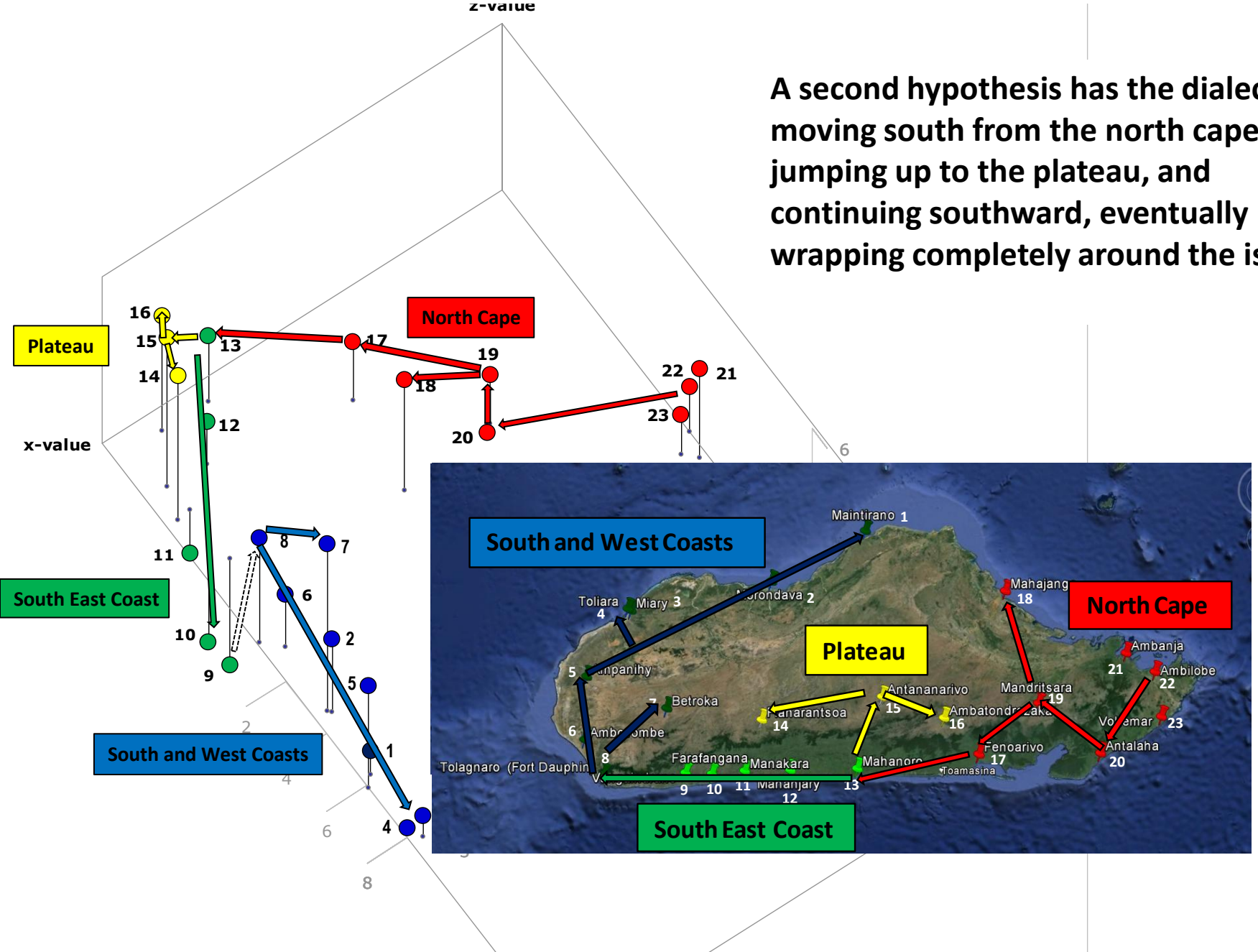
Language Dispersal Hypothesis 1: 3-Way Split





One hypothesis for language dispersal is this: both the northern and southern dialects started out in mid-island, along with the plateau dialects, then spread northward and southward as shown on the MDS plot and map.

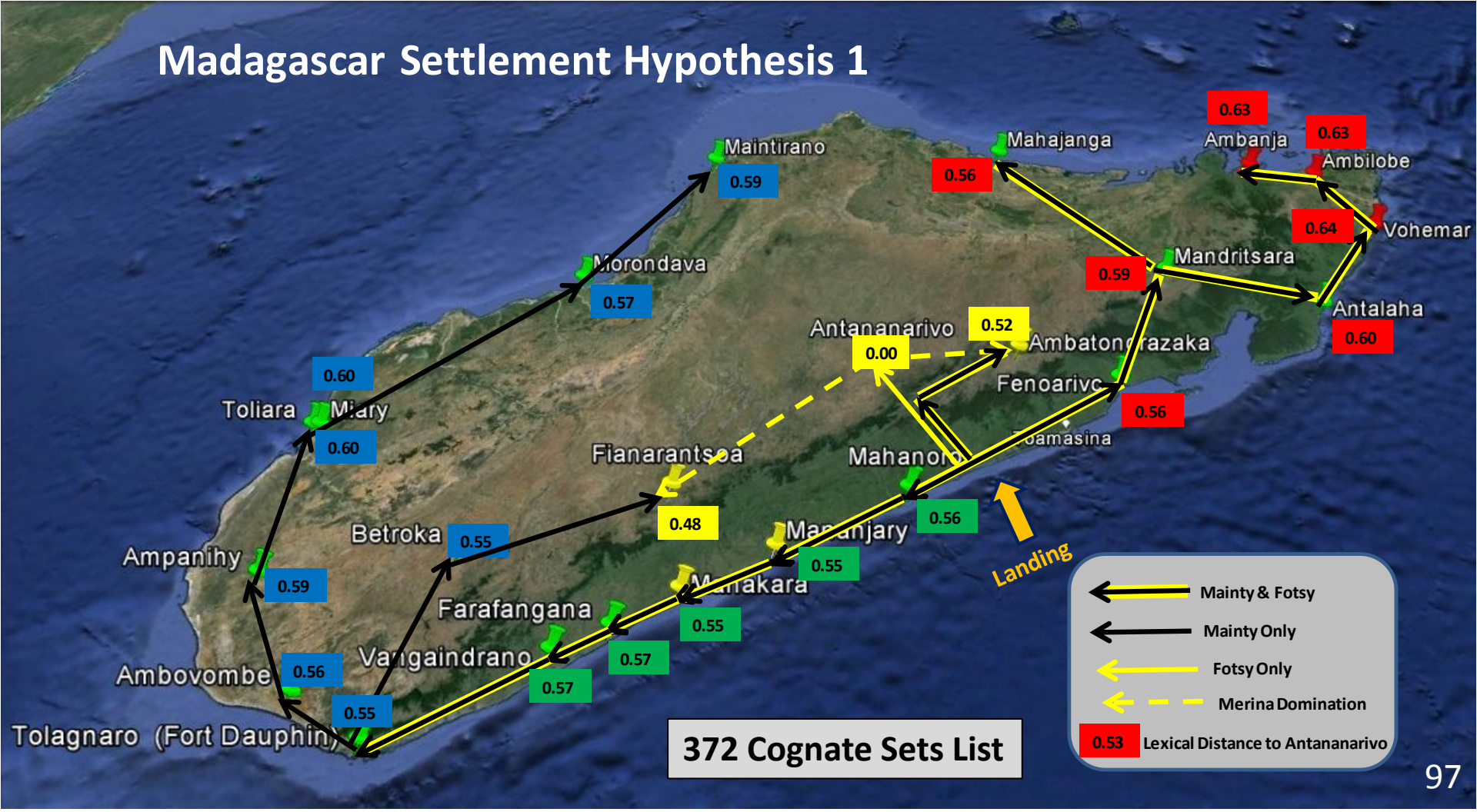
Language Dispersal Hypothesis 2: North to South



A second hypothesis has the dialects moving south from the north cape, jumping up to the plateau, and continuing southward, eventually wrapping completely around the island.

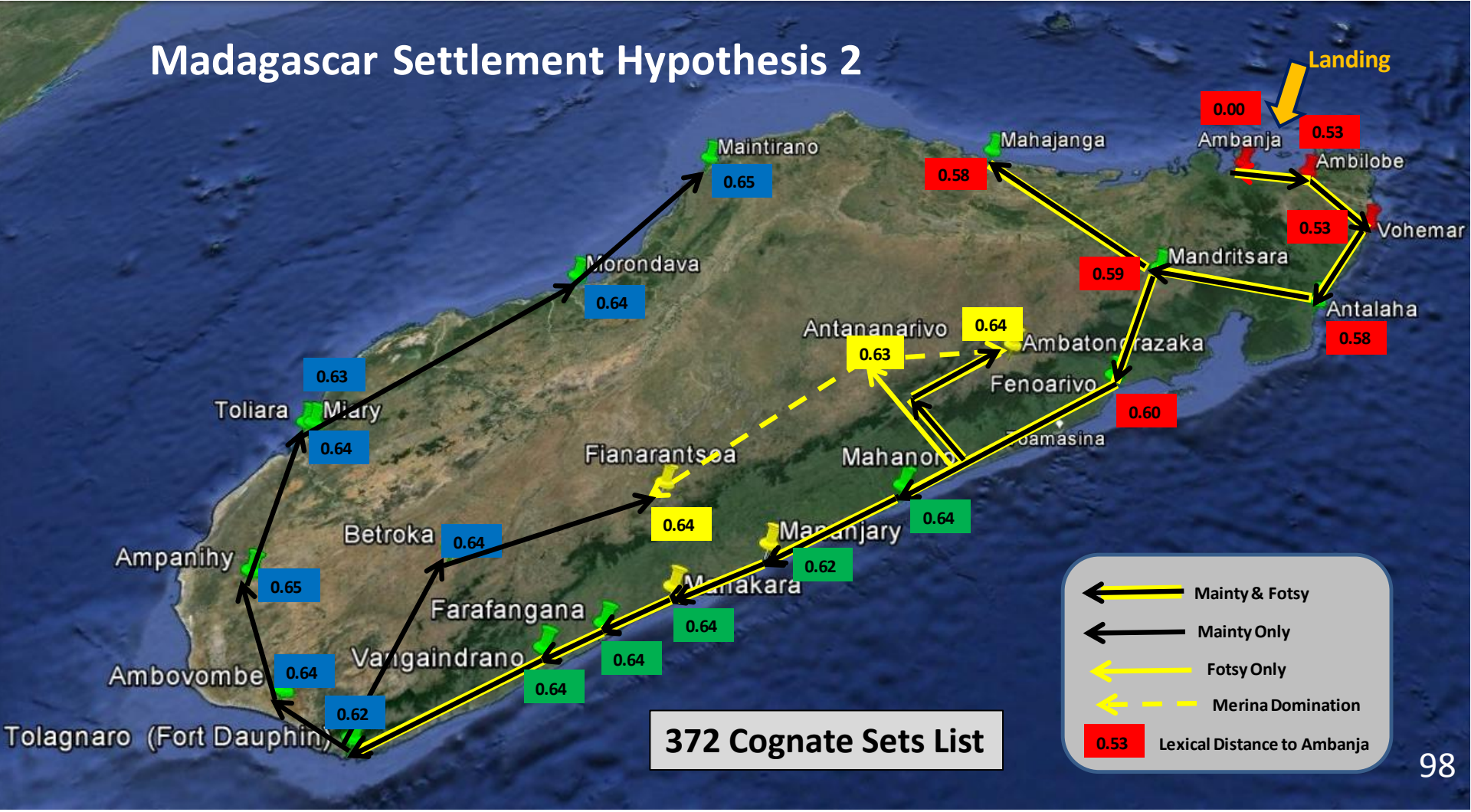
Etienne de Flacourt wrote in 1658 that the area surrounding Fort Dauphin was divided into two hierarchies, one classified as 'white' (fotsy), the other 'black' (mainty). The royal family, at the top of the 'white' hierarchy, was descended from a group of immigrants known as *Zafiraminina* who had reached Madagascar some seventeen generations previously and migrated gradually down the east coast before arriving in the far southeast. The system of double authority reflected an accommodation between the *Zafiraminina* immigrants and the local peoples.

This might explain the pattern of word transmission we observe in the Malagasy dialects, where the plateau dialects and the coastal dialects could coexist in two separate groups, the Mainty and the Fotsy, as they gradually moved southward from their point of landing.



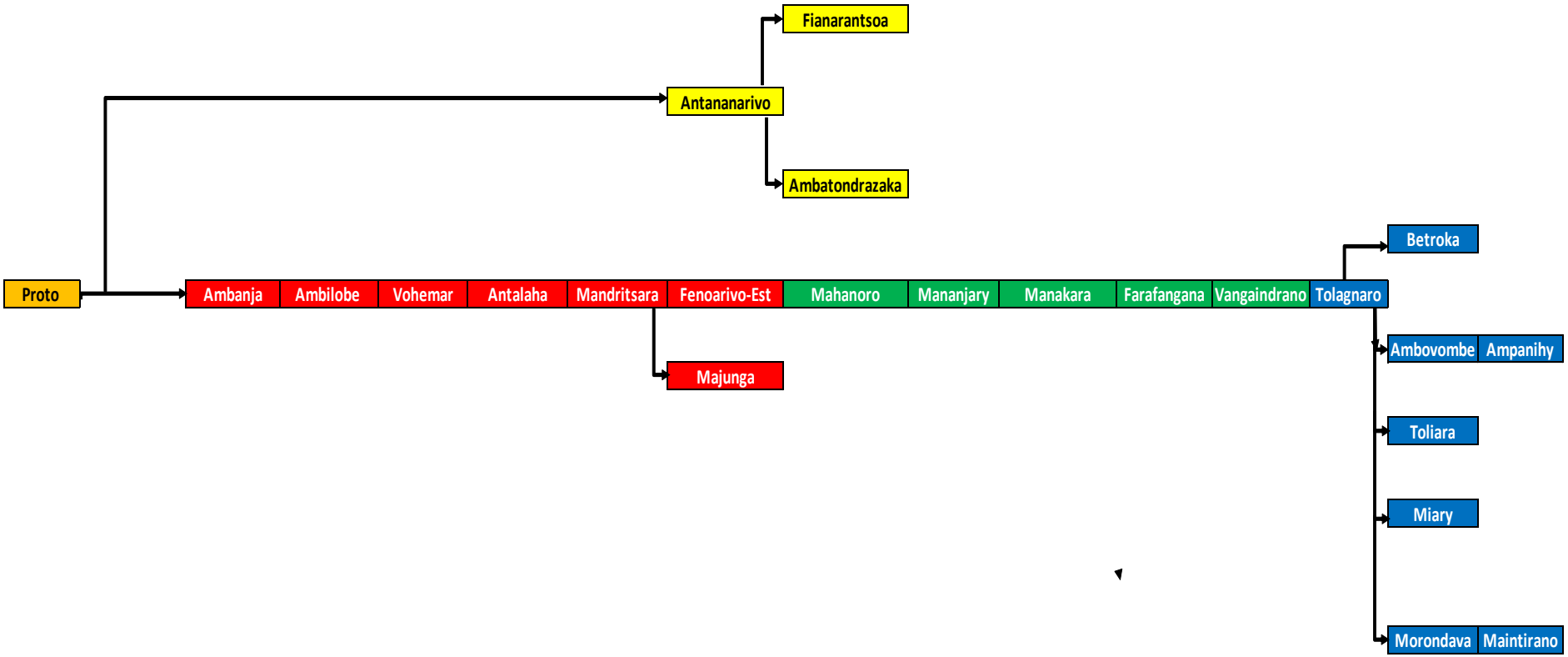
97

Here is a map of Madagascar settlement according to Hypothesis 1. Indonesians and Africans landed on the east coast and spread together, as Flacourt recorded, but both northward and southward. The south and west was settled mostly by Africans as Indonesians dominated the other areas and the Merina conquered and developed the high plains. The biggest problem with this hypothesis is the unlikelihood of an east coast landing by a mixed African/Indonesian group. The possibility of Africans and Indonesians did not meet until the Indonesians landed alone is remote. How could a small band of seafarers just happen to meet a similarly small group of African settlers on the east coast?



Here is a map of Madagascar settlement consistent with Hypothesis 2. It is driven by historical and archaeological findings as well as by common sense, and it is supported by the linguistic results. It assumes that the Fotsy and Mainty people had migrated together from the North Cape all the way to the south. At some point before the 18th century, the Merina moved en masse to the highlands and conquered or dominated other tribes in the area. Meanwhile, the mostly-African tribes, perhaps augmented by newcomers from Africa to the south cape, continued their settlement of the island south cape and south-west coast.

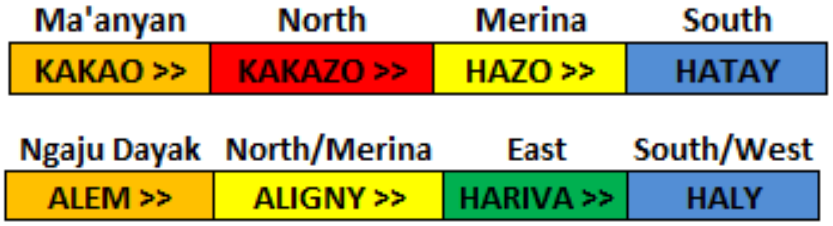
Phylogenic Tree Derived from Settlement Hypothesis 2



This tree diagrams the Northern settlement hypothesis. It is not based on Lexical Distance, but rather on a hypothetical settlement history supported by lexical data, not driven by it.
From the northern landing, the Fotsy and Mainty migrated southward while keeping their linguistic preferences. The Mainty became the coastal dwellers shown in Red, Green and Blue variations, while the Fotsy maintained the purest Indonesian dialect, eventually becoming the Merina and exerting their lingustic domination over the Betsileo and Tsimehety highland tribes.

Support for Northern Immigration Hypothesis

- Early Archaeological Remains (685-745 CE) in Caves in the north, the most likely landing spot for a joint expedition from Africa.
- An East Coast Landing is Logically Unlikely
- Flacourt's story about Mainty/Fotsy N>>S Migration
- North-to-South Word Evolution:

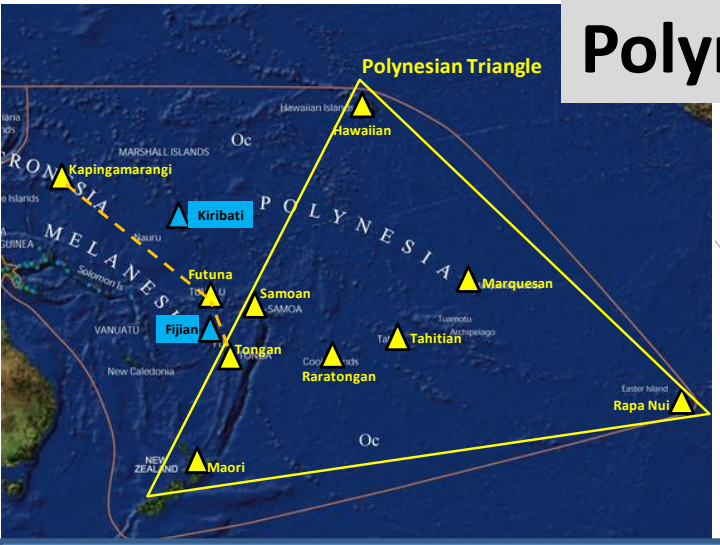


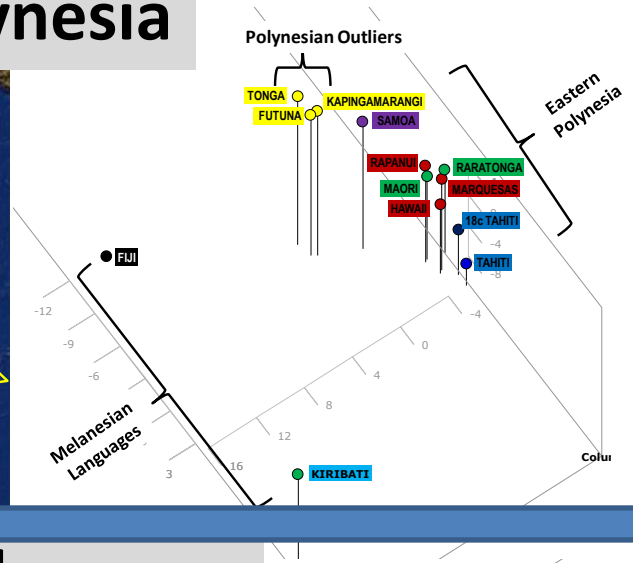
- The presence of imported words *Motro* and *Matoe* in the North Only

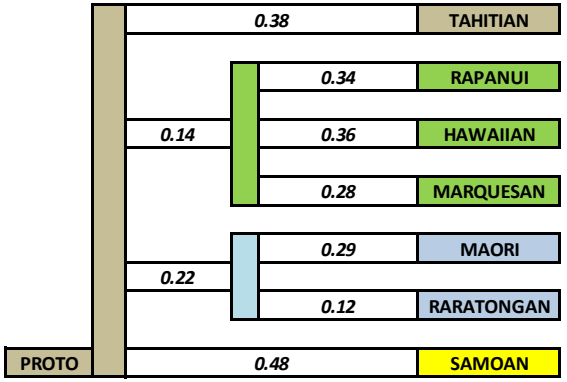
Conclusions

- Lexicography and DNA indicate Borneo and Mozambique are Origins of Malagasy People
- Malagasy Languages Evolved from a Prototype Austronesian Language no Longer in Existence
- Lexical Distance Analysis Alone Does Not Accurately Predict Settlement Patterns
- But Analysis of Shared Cognates and Lexical Distances Do Inform Understanding of Settlement
- North-to-South Mainty/Fotsy Settlement Hypothesis is Supported, but not Proved by Lexical Analysis

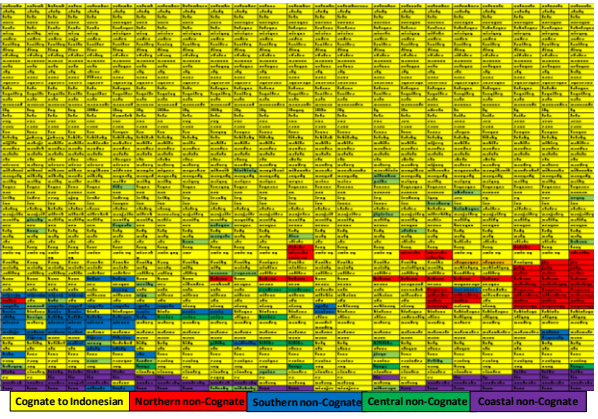
Polynesia

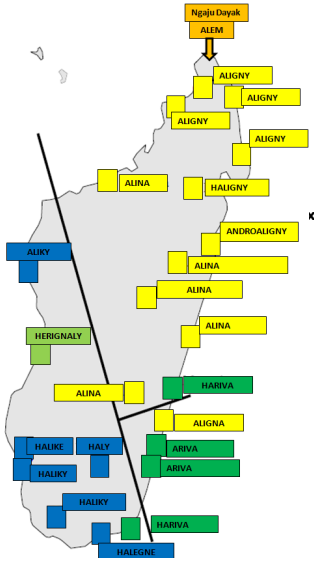


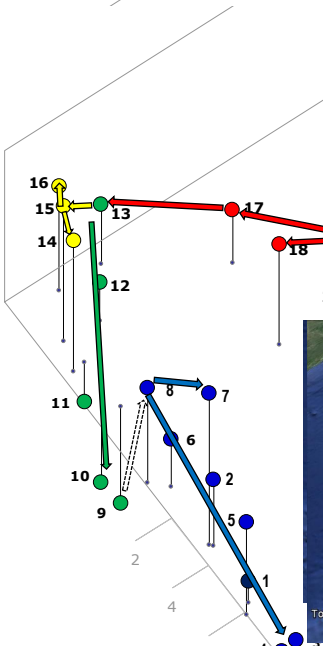


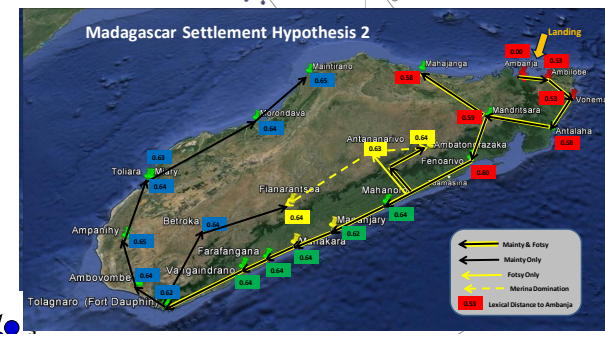


Madagascar









We've seen how Polynesia could be differentiated from Melanesia based on Lexical Analysis and Multidimensional Scaling. And how we could develop a Phylogenic Tree for the languages of eastern Polynesia. For Madagascar, we saw how regional word patterns dominated the lexical distances between Malagasy dialects, and how word-by-word analysis along with Lexical Distance and MDS support the Northern Settlement Hypothesis for the world's 4th largest island.

References:

Serva, M. et al., "Malagasy Dialects and the Peopling of Madagascar," in J. R. Soc. Interface 7 January 2012 vol. 9 no. 66 54-67

Tofanelli, S., et al., "On the Origins and Admixture of Malagasy: New Evidence from High-Resolution Analyses of Paternal and Maternal Lineages," in Mol Biol Evol *(2009) 26 (9): 2109-2124.*

Cox, M.P. et al. "A small cohort of Island Southeast Asian women founded Madagascar," in Proc. R. Soc. B 22 July 2012 vol. 279 no. 1739 2761-2768

Kayser, M. et al., "Y Chromosome STR Haplotypes and the Genetic Structure of U.S. Populations of African, European, and Hispanic Ancestry," in Genome Res. 2003 13: 624-634

Kikusawa, R., "Standardization as Language Loss: Potentially Endangered Malagasy Languages and Their Linguistic Features" in *People and Culture in Oceania*, 28: 23–44, 2012

Randrianja, S. and Ellis, S., "Madagascar, a Short History," University of Chicago Press, 2009.

Dahl, O. C., "Migration from Kalimantan to Madagascar," 1991. Oslo: Egede Institute, 1991. ISBN 82-00-21140-1

Brown, M., "Madagascar Rediscovered," Archon Books, 1979.

Verin, J. and Wright, H., "Madagascar and Indonesia: new evidence from archaeology and linguistics," Bulletin of the Indo-Pacific Prehistory Association, vol. 18 (1999).

YouTube Videos:

Tsimihety Girls (Music Video):
https://www.youtube.com/watch?v=bvoFt3UvO3w

How to Speak Malagasy (taught by a cool instructor):
https://www.youtube.com/watch?v=Dl3oMPLUNwY

Trials and Tribulations of Road Transport in Northern Madagascar:
https://www.youtube.com/watch?v=UOjxSNbuTqM

Vezo Fishing People at Toliara, West Coast:
https://www.youtube.com/watch?v=xEmPJC6soAA

Antananarivo City:
https://www.youtube.com/watch?v=oyi_dUQVi-I

This Presentation is viewable at:

http://www.roryvantuyl.com/Linguistics.html