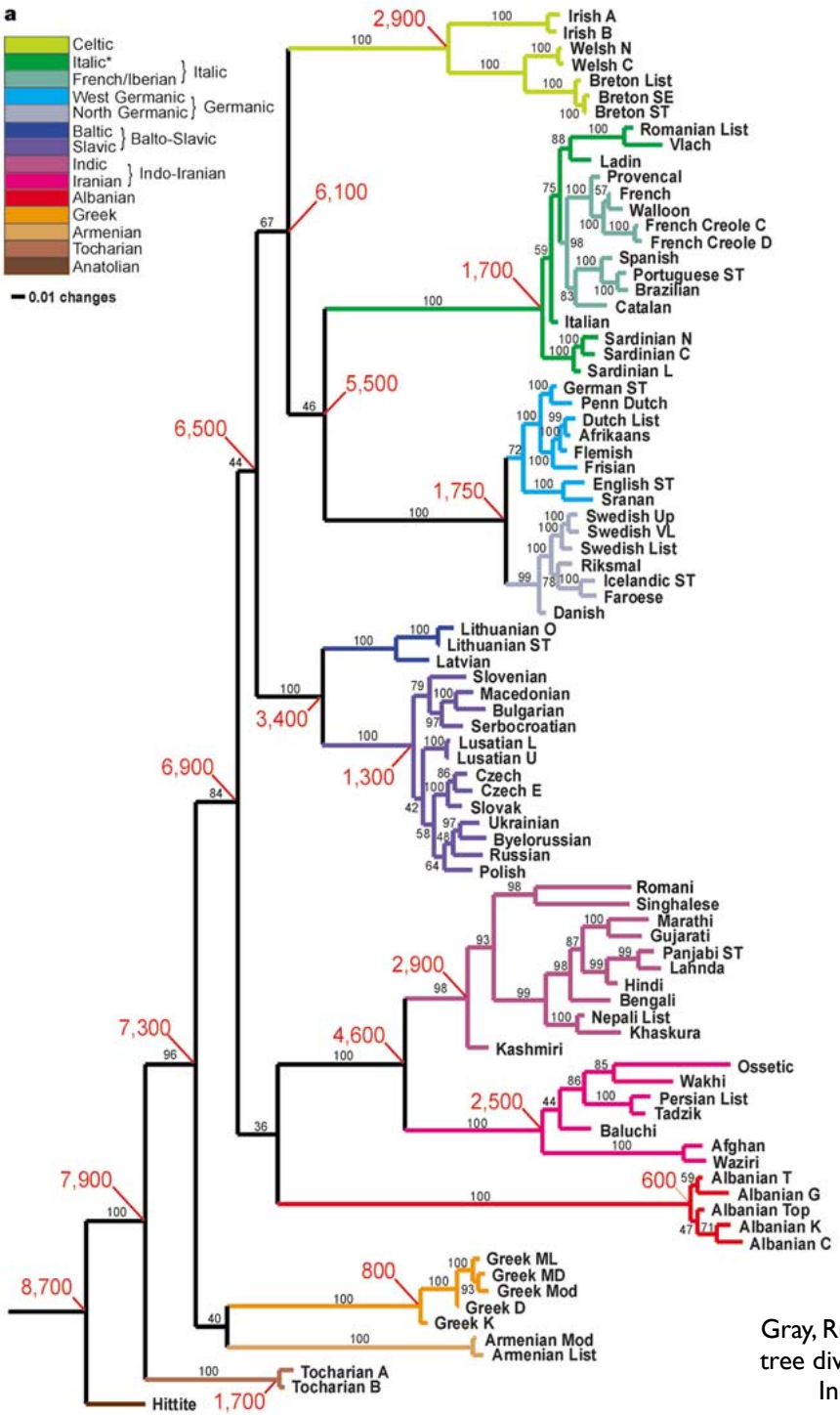
Towards phylogenetic analyses of linguistic data

Michael Cysouw

Max Planck Institute
for Evolutionary Anthropology

1. Phylogenetic Analyses in Linguistics

Gothic	***fotus***	'foot'
Old Icelandic	***fo:tr***	'foot'
Old High German	***fuoz***	'foot'
German	***Fuss***	'foot'
Old English	***fo:t***	'foot'
English	***foot***	'foot'
Hittite	***pata-***	'foot'
Luvian	***pati-***	'foot'
Latin	***pe:s, pedis***	'foot'
Tocharian A	***päts***	'foot'
Tocharian B	***ptsa***	'foot'
Greek	***poús, podós***	'foot'
Armenian	***ot-n, ot-k'***	'foot, feet'
Sanskrit	***pá:t, pá:dam***	'foot'
Avestan	***pad-***	'foot'
Lithuanian	***pa~das***	'shoe'
Latvian	***acu-pedius***	'swift-footed'
Old Church Slavic	***pods***	'ground'
Proto-Indo-European	****p^{h}et'-***	



Gray, Russel & Quentin D. Atkinson (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426: 435-439.

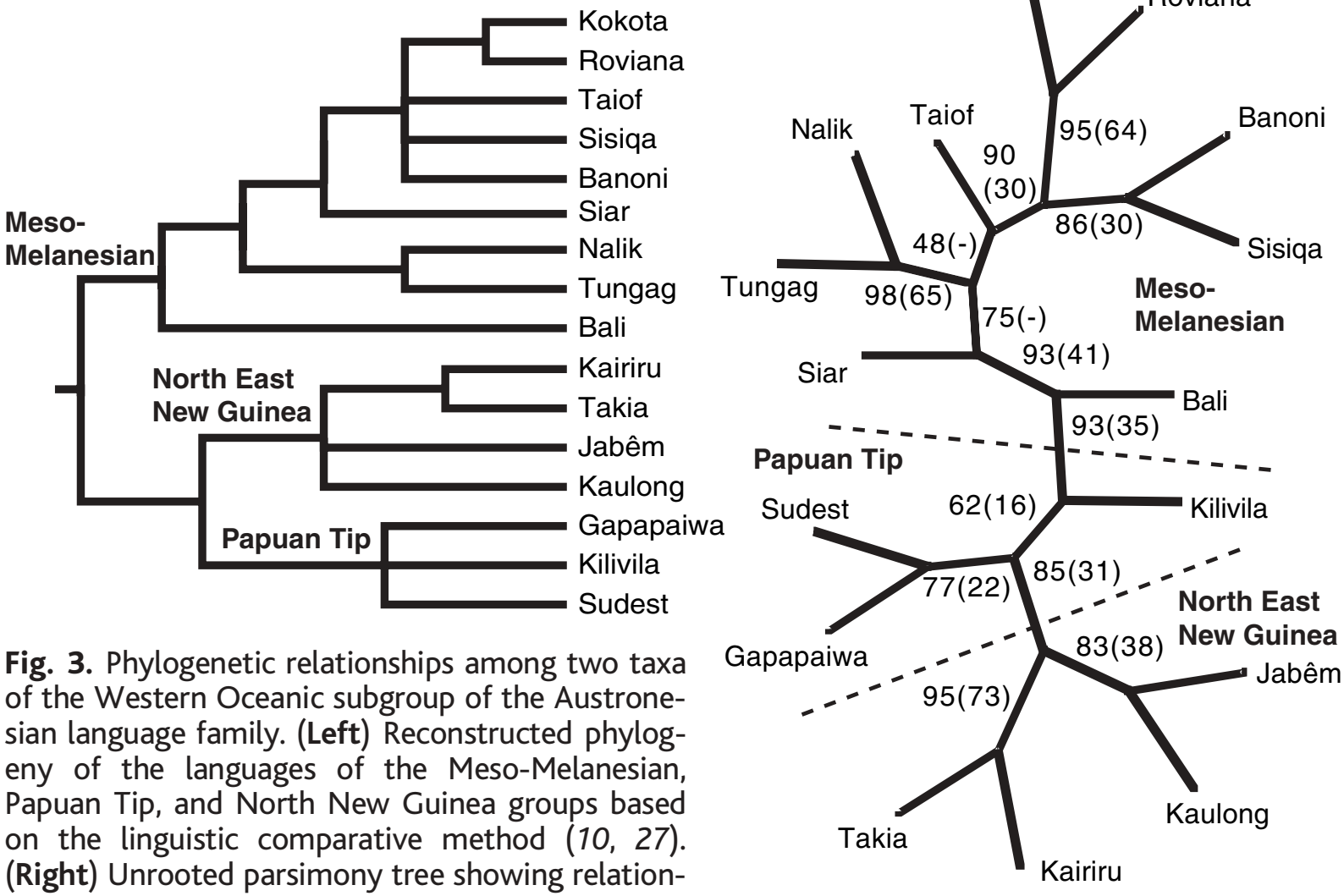
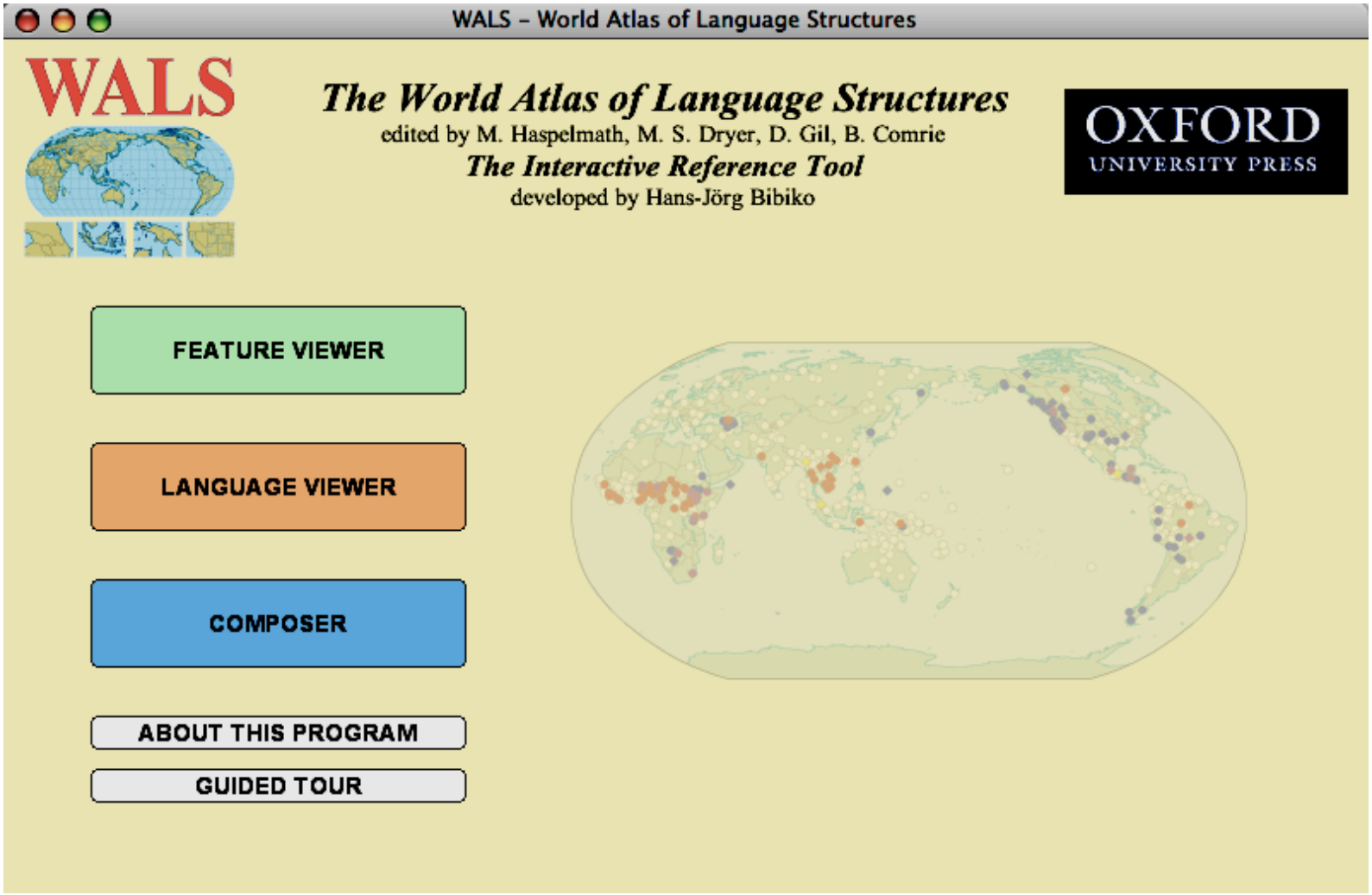


Fig. 3. Phylogenetic relationships among two taxa of the Western Oceanic subgroup of the Austronesian language family. (**Left**) Reconstructed phylogeny of the languages of the Meso-Melanesian, Papuan Tip, and North New Guinea groups based on the linguistic comparative method (*10*, *27*). (**Right**) Unrooted parsimony tree showing relationships among the Meso-Melanesian and Papuan Tip groups based on grammatical traits only (that is, discarding abundant lexical evidence) (the figure shows reweighted and raw bootstrap values). The two trees show a high degree of concordance, with monophyly in both major taxa and the similar geographical structuring of within-taxon diversity.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural Phylogenetics and the Reconstruction of Ancient Language History. Science 309: 2072-2075.

2. Introducing Linguistic Typology


WALS – World Atlas of Language Structures
WALS
The World Atlas of Language Structures
edited by M. Haspelmath, M. S. Dryer, D. Gil, B. Comrie
The Interactive Reference Tool
developed by Hans-Jörg Bibiko
OXFORD
UNIVERSITY PRESS
FEATURE VIEWER
LANGUAGE VIEWER
COMPOSER
ABOUT THIS PROGRAM
GUIDED TOUR

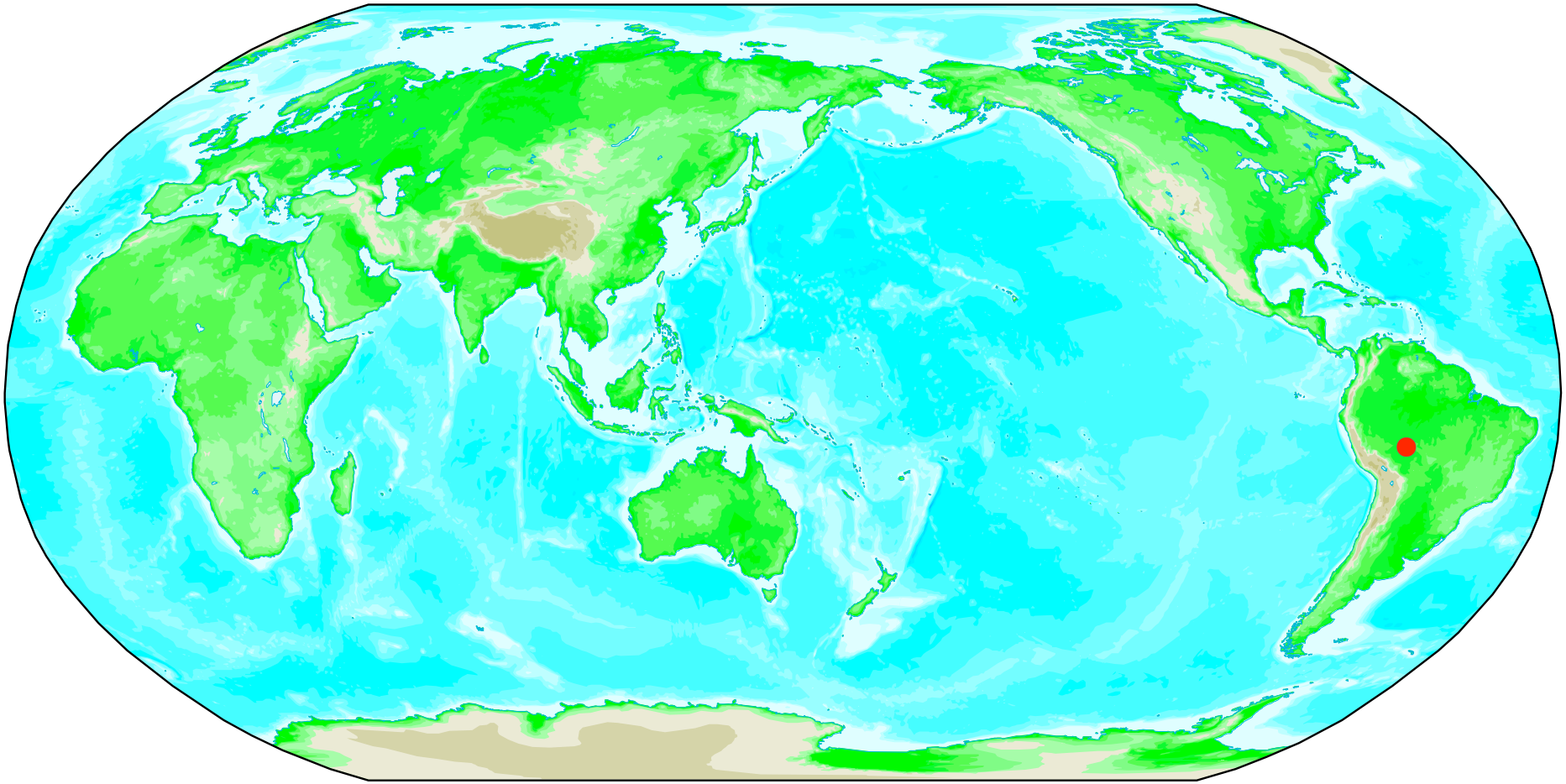
3. Investigating unusual characteristics

And the winners are:

In the category:

‘Most Unusual Individual Language’

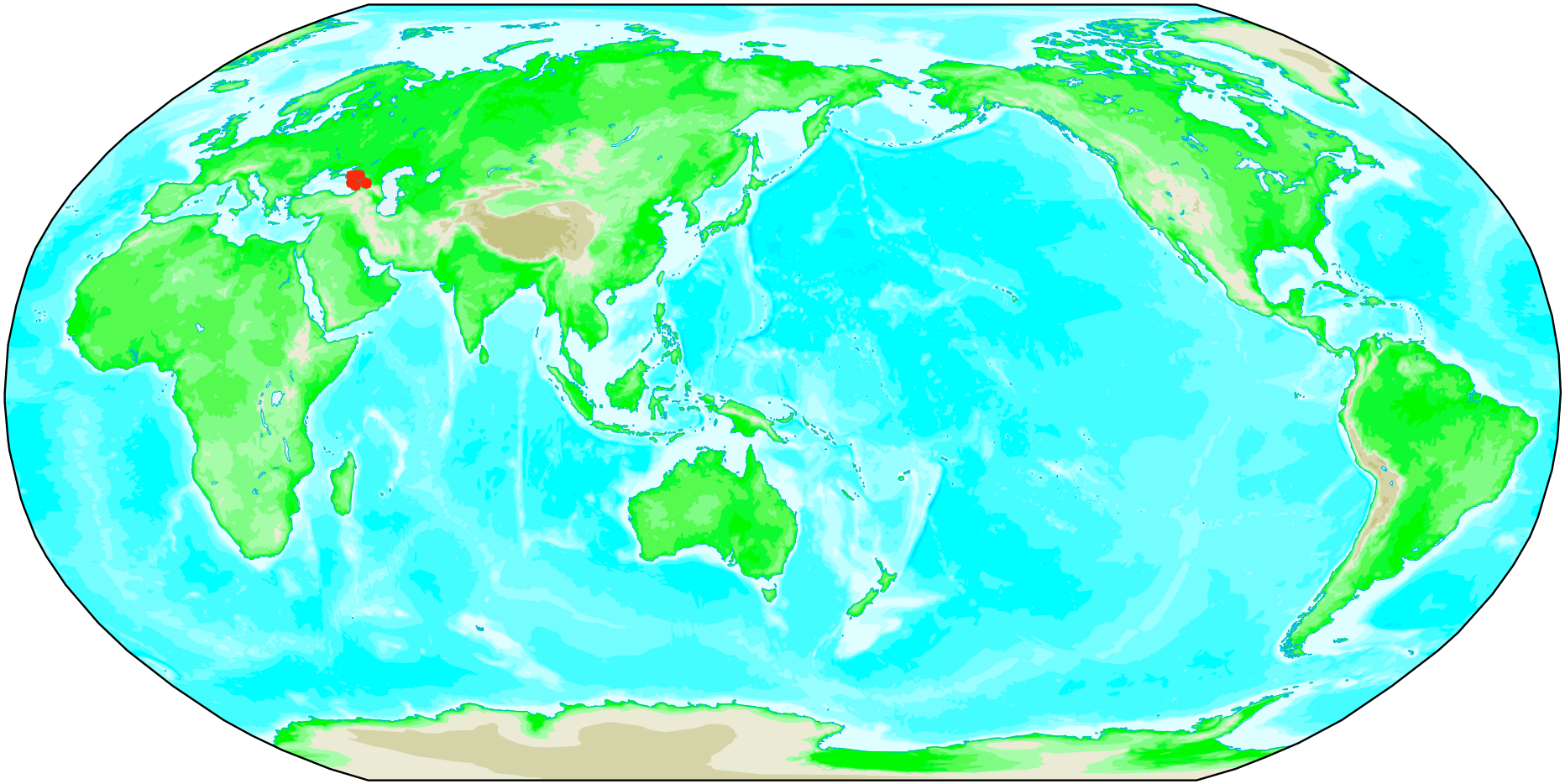
Wari’



In the category:

‘Most Unusual Genealogical Group’

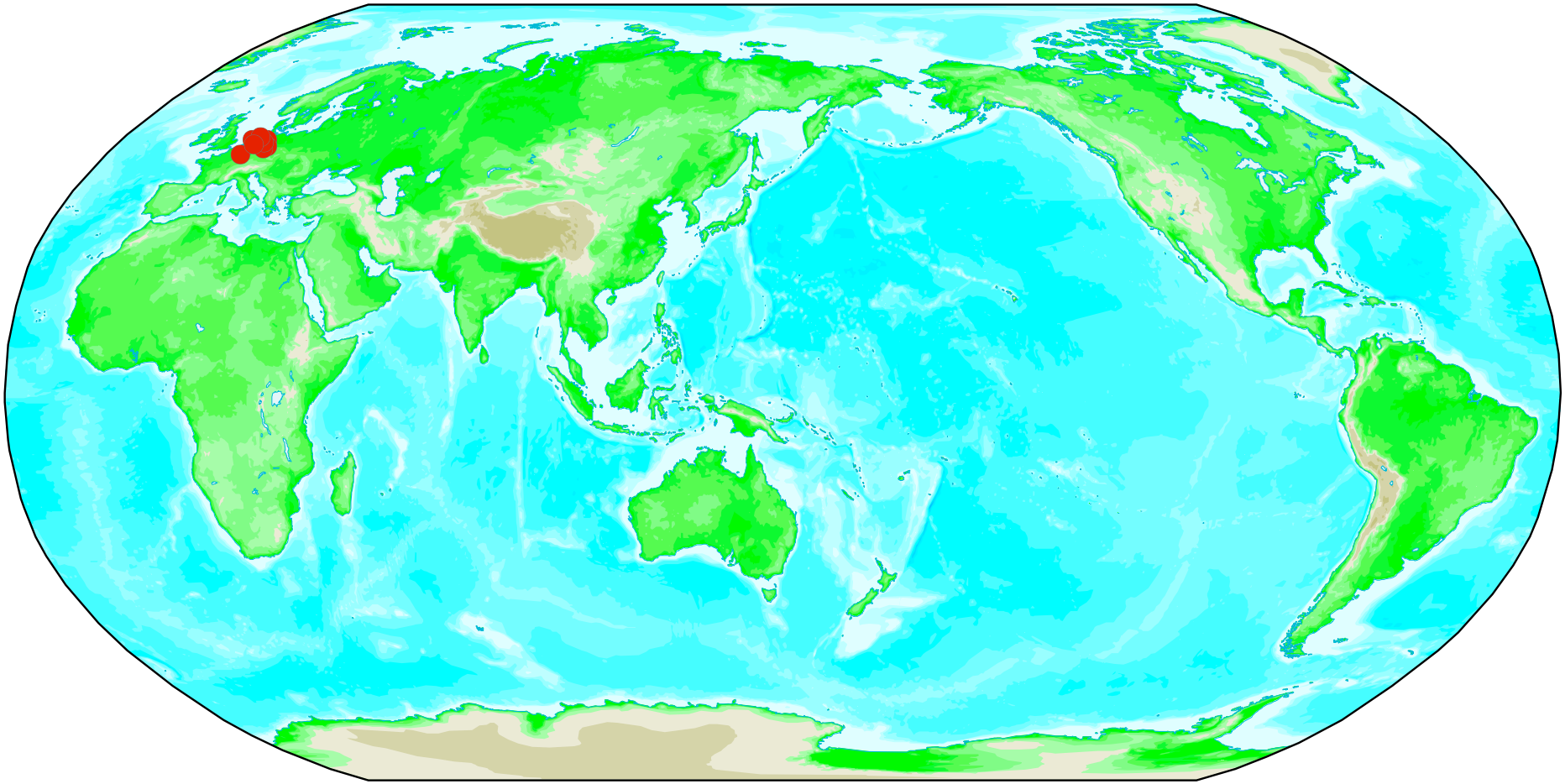
Northwest Caucasian



In the category:

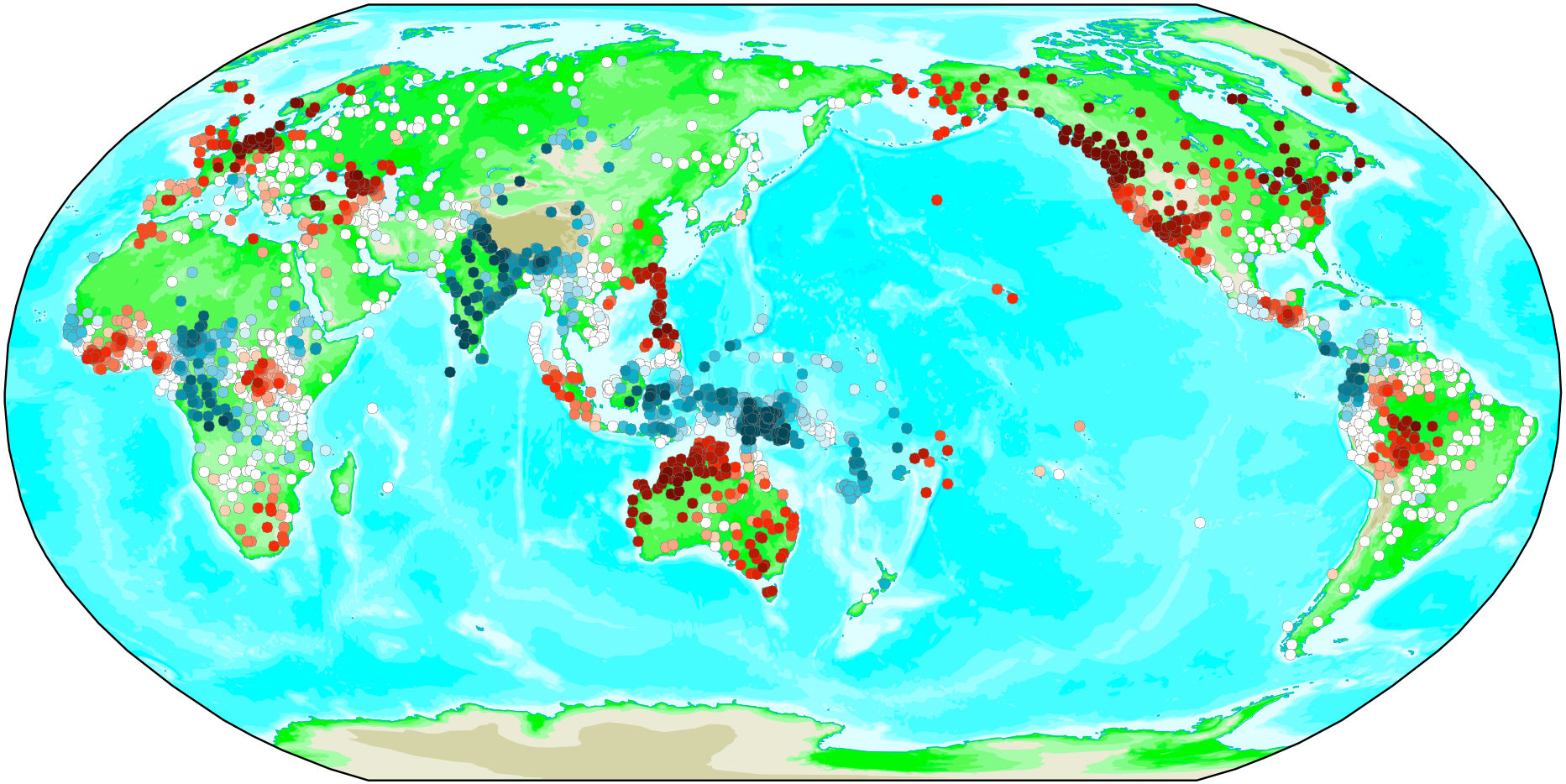
‘Most Unusual Geographical Area’

Northwest Continental Europe



All languages

(red = rare, blue = common)

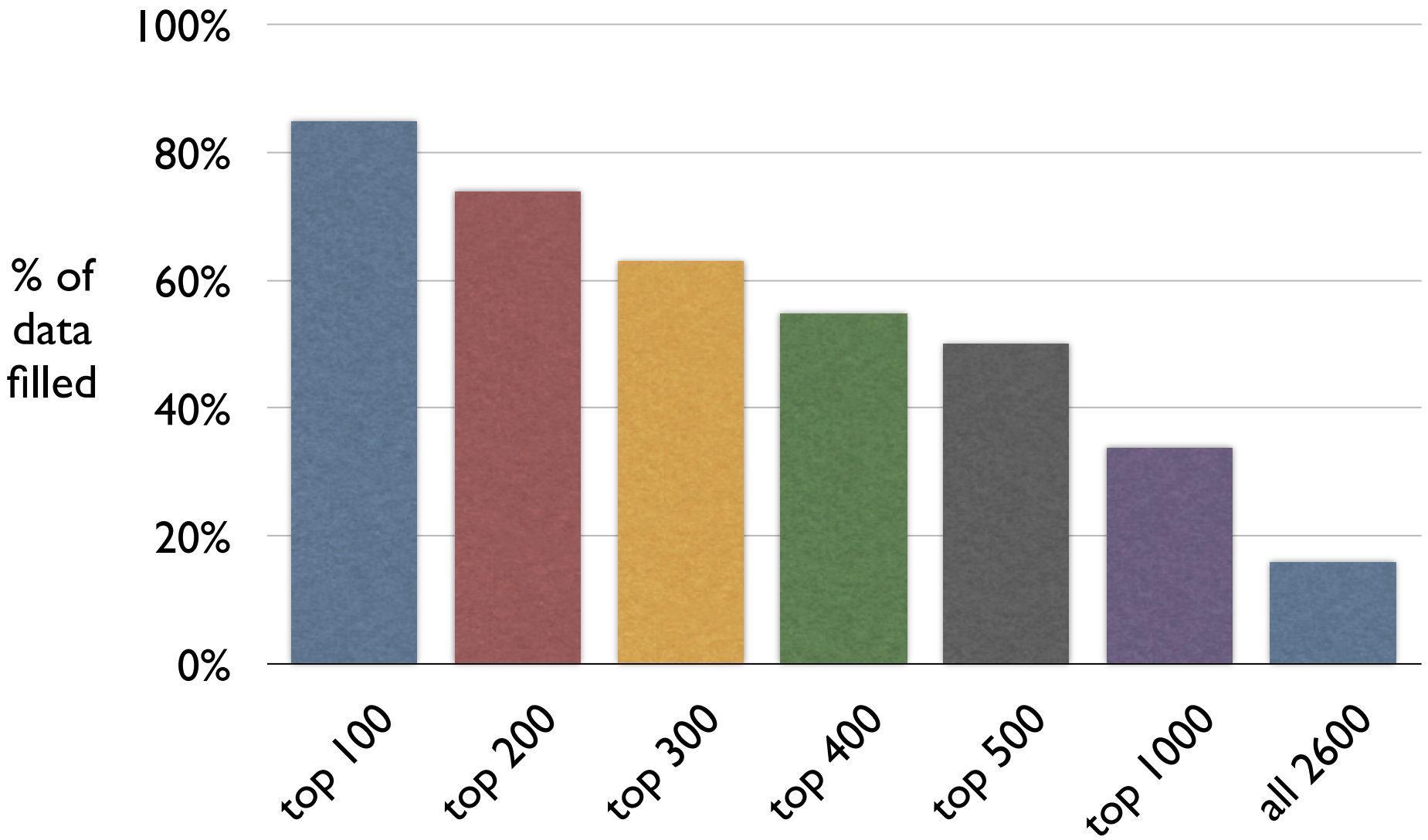


4. Towards Phylogenetic Analyses with WALS

A wealth of data (for linguistics)

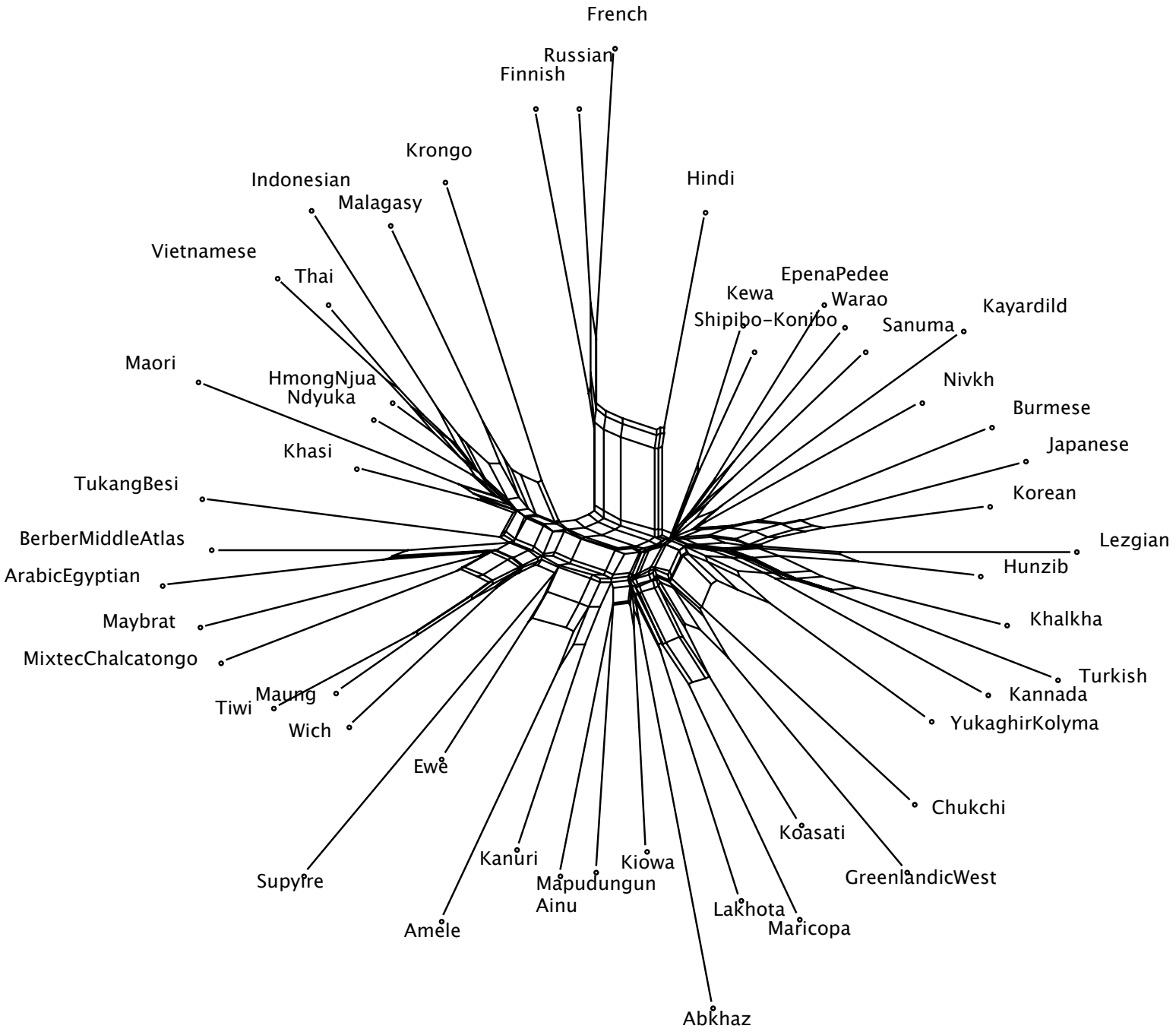
- 2,600 languages
- 140 characteristics
- almost 60,000 datapoints
- but: 60,000/2,600*140 = 0.165
- the datatable is only 16.5 % filled !

Choosing the best languages



Reliability

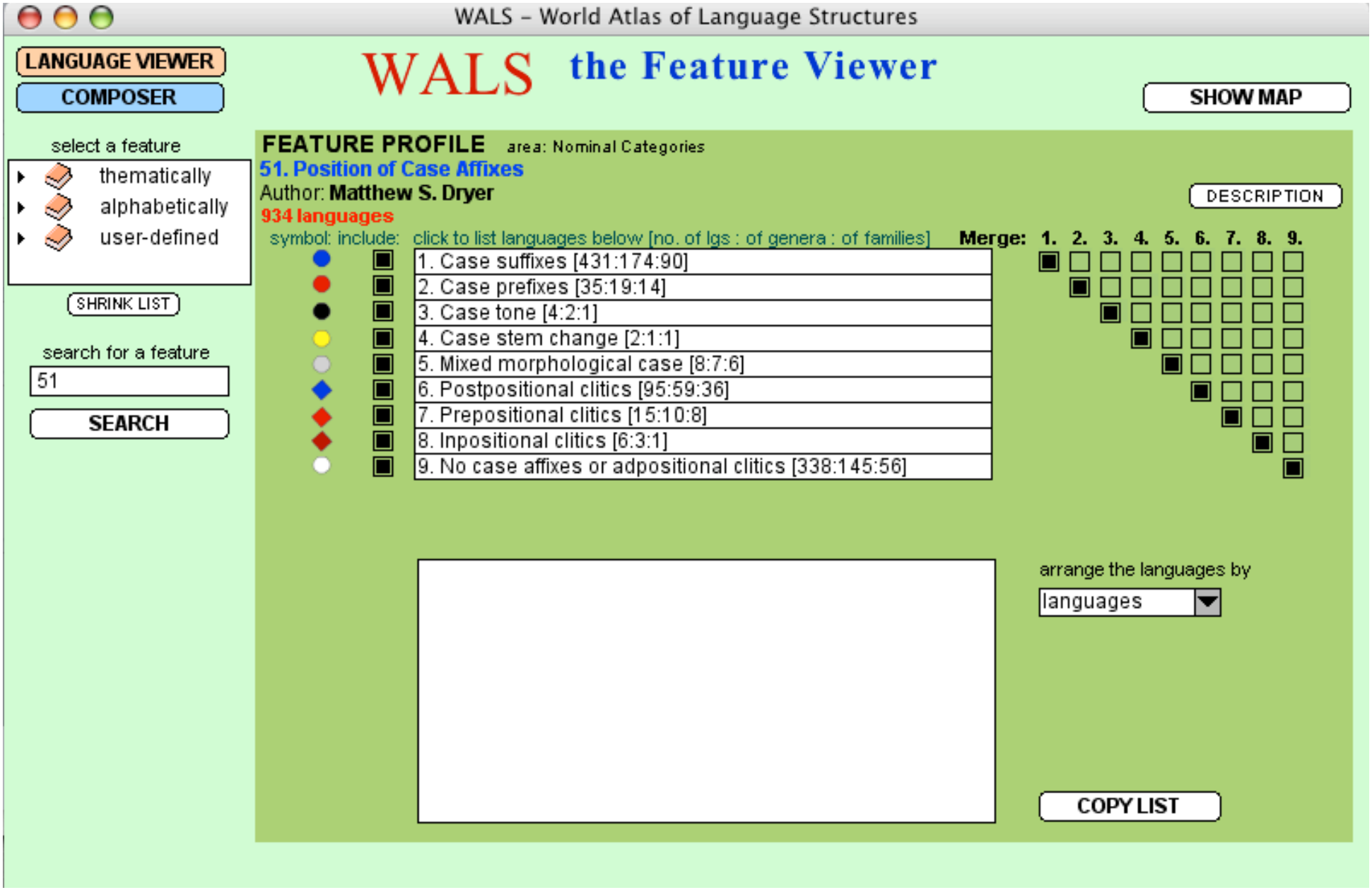
- Latvian was checked (by B. Wälchli)
- 109 coding point in WALS
- 2 'technical' errors (= 1.8 %)
- 5 'interpretative' errors (= 4.6 %)


French
Russian
Finnish
Krongo
Indonesian
Malagasy
Hindi
Vietnamese
Thai
EpenaPedee
Kewa
Warao
Kayardild
Shipibo-Konibo
Sanuma
Maori
HmongNjua
Ndyuka
Nivkh
Burmese
Khasi
Japanese
TukangBesi
Korean
BerberMiddleAtlas
Lezgian
ArabicEgyptian
Hunzib
Maybrat
Khalkha
MixtecChalcatongo
Turkish
Maung
Kannada
Tiwi
YukaghirKolyma
Wich
Ewe
Chukchi
Koasati
Kanuri
Kiowa
GreenlandicWest
Supyire
Mapudungun
Ainu
Lakhota
Amele
Maricopa
Abkhaz

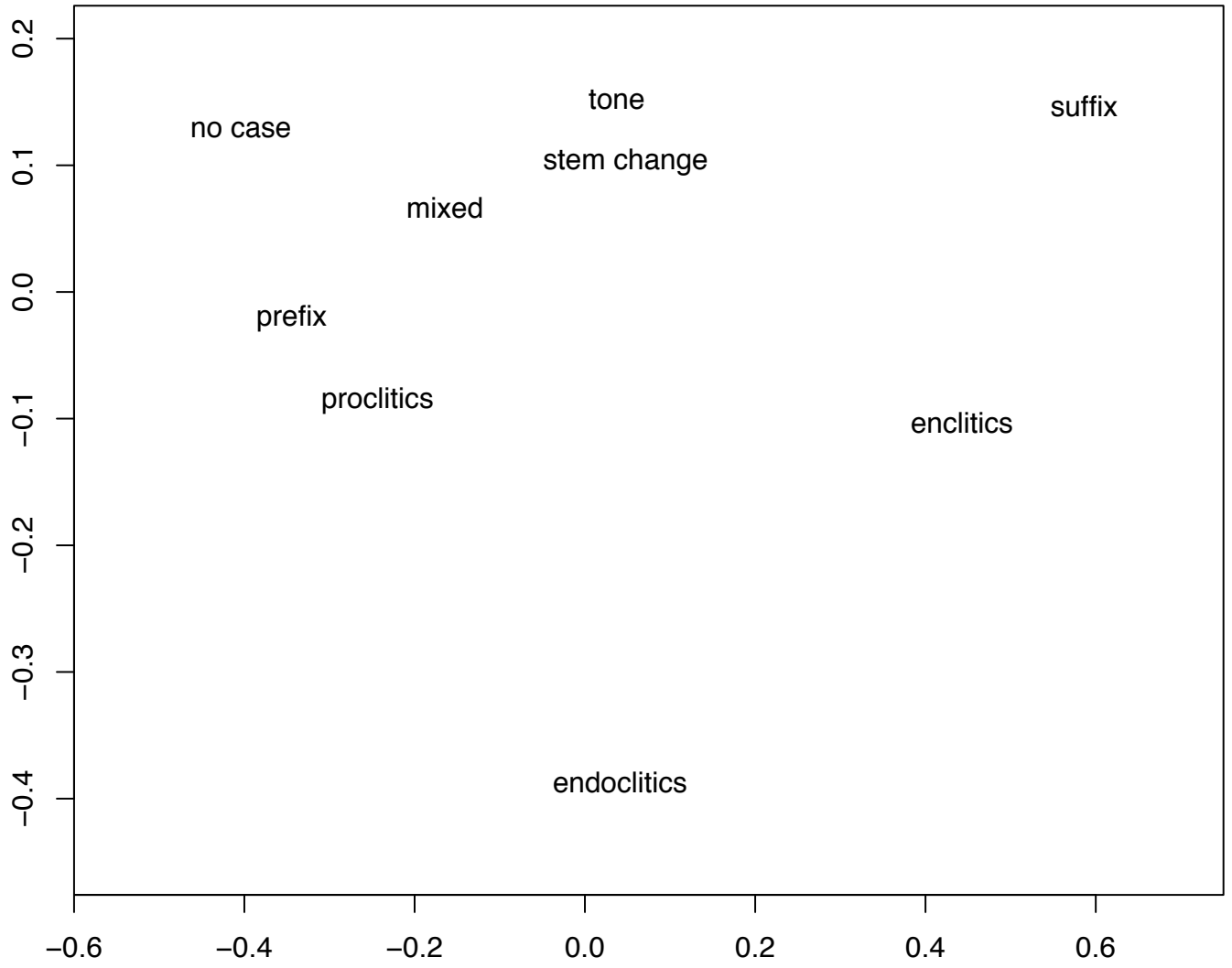
5. Improving Typology

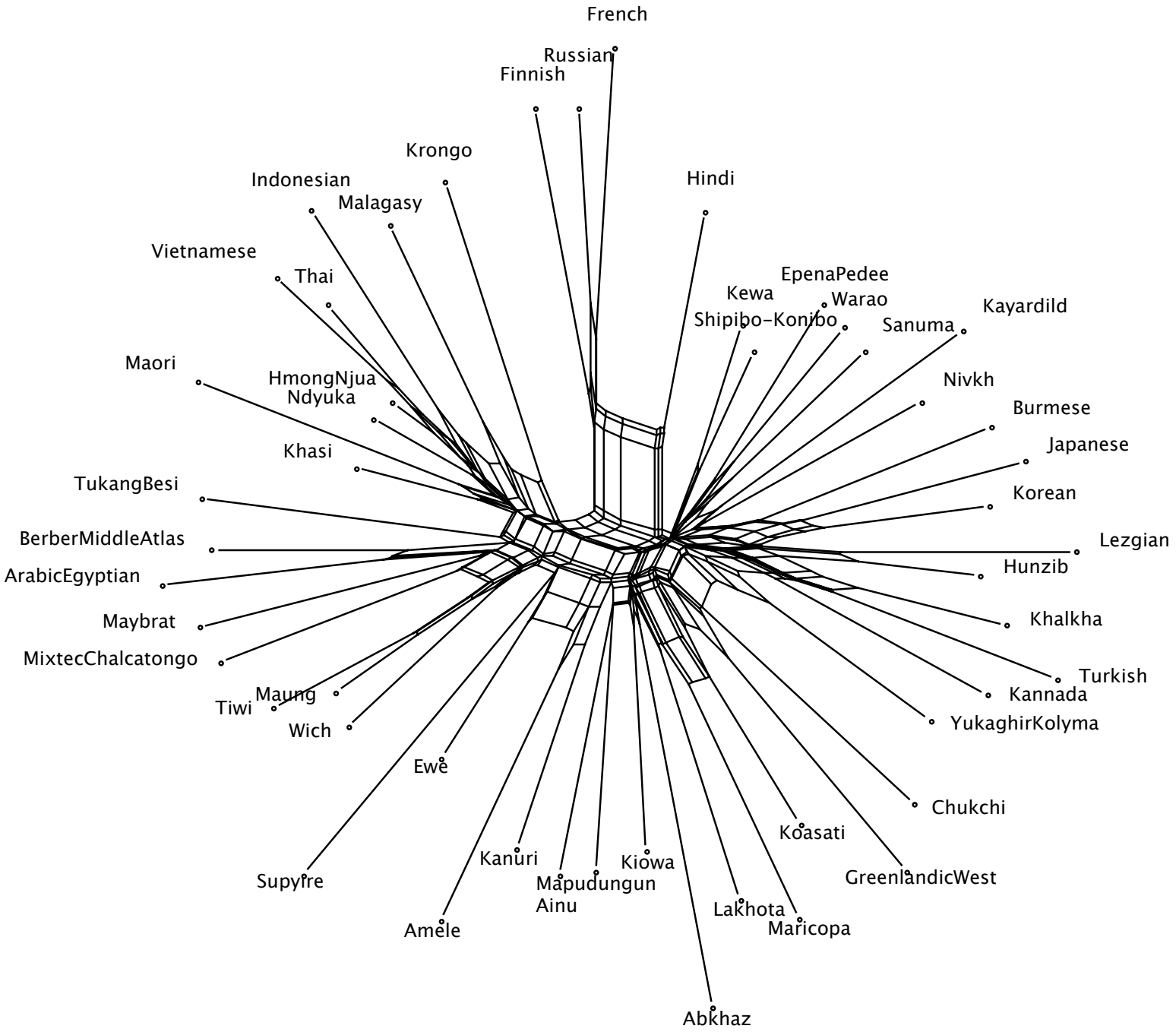
5. Improving Typology

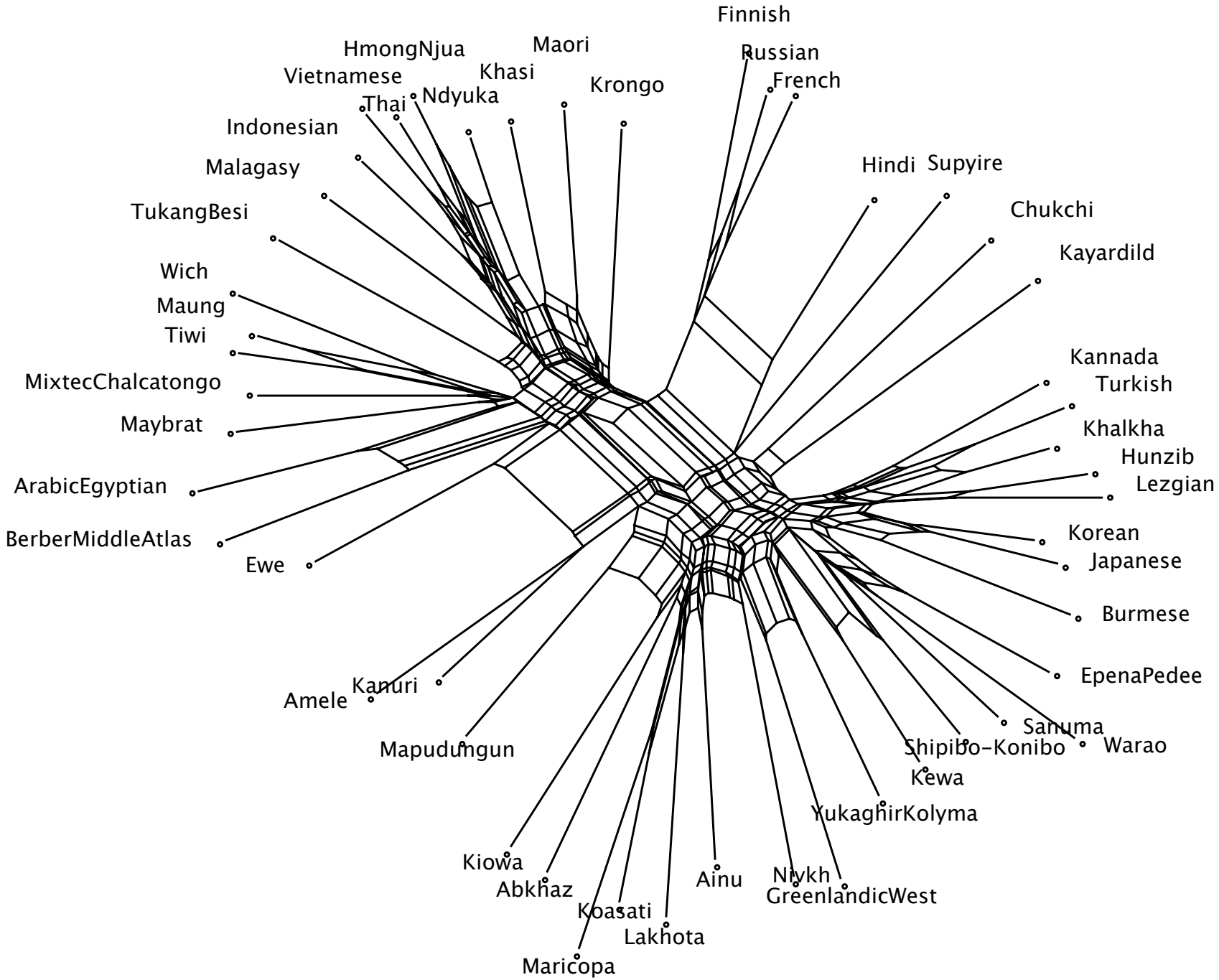
- Finer-graind Coding
- ‘Deconstructing’ Typology
- Select Suitable Characteristics


WALS – World Atlas of Language Structures
LANGUAGE VIEWER
COMPOSER
WALS the Feature Viewer
SHOW MAP
select a feature
thematically
alphabetically
user-defined
SHRINK LIST
search for a feature
51
SEARCH
FEATURE PROFILE area: Nominal Categories
51. Position of Case Affixes
Author: Matthew S. Dryer
934 languages
DESCRIPTION
symbol: include: click to list languages below [no. of lgs : of genera : of families]
Merge: 1. 2. 3. 4. 5. 6. 7. 8. 9.
1. Case suffixes [431:174:90]
2. Case prefixes [35:19:14]
3. Case tone [4:2:1]
4. Case stem change [2:1:1]
5. Mixed morphological case [8:7:6]
6. Postpositional clitics [95:59:36]
7. Prepositional clitics [15:10:8]
8. Inpositional clitics [6:3:1]
9. No case affixes or adpositional clitics [338:145:56]
arrange the languages by
languages
COPY LIST

Estimate **character-similarities** from cooccurrenc in low-level genetic groups




French
Russian
Finnish
Krongo
Indonesian
Malagasy
Hindi
Vietnamese
Thai
EpenaPedee
Kewa
Warao
Kayardild
Shipibo-Konibo
Sanuma
Maori
HmongNjua
Ndyuka
Nivkh
Burmese
Khasi
Japanese
TukangBesi
Korean
BerberMiddleAtlas
Lezgian
ArabicEgyptian
Hunzib
Maybrat
Khalkha
MixtecChalcatongo
Turkish
Maung
Kannada
Tiwi
YukaghirKolyma
Wich
Ewe
Chukchi
Koasati
Kanuri
Kiowa
GreenlandicWest
Supyire
Mapudungun
Ainu
Lakhota
Amele
Maricopa
Abkhaz


Finnish
Russian
French
Maori
HmongNjua
Khasi
Vietnamese
Krongo
Ndyuka
Thai
Indonesian
Malagasy
Hindi
Supyire
TukangBesi
Chukchi
Kayardild
Wich
Maung
Tiwi
Kannada
Turkish
MixtecChalcatongo
Khalkha
Maybrat
Hunzib
Lezgian
ArabicEgyptian
Korean
BerberMiddleAtlas
Japanese
Ewe
Burmese
EpenaPedee
Kanuri
Amele
Saruma
Mapudungun
Shipibo-Konibo
Warao
Kewa
YukaghirKolyma
Kiowa
Abkhaz
Ainu
Nivkh
GreenlandicWest
Koasati
Lakhota
Maricopa

5. Improving Typology

- Finer-graind Coding
- ‘Deconstructing’ Typology
- Select Suitable Characteristics

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1									
L_2									
L_3									
L_4									
L_5									
L_6									
L_7									
L_8									
…									

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1								
L_2		1							
L_3			1						
L_4				1					
L_5					1				
L_6						1			
L_7							1		
L_8								1	
…									

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	I								
L_2		I							
L_3			I						
L_4				I					
L_5					I				
L_6						I			
L_7							I		
L_8								I	
…									

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	I	I	I						
L_2	I	I	I						
L_3	I	I	I						
L_4				I	I				
L_5				I	I				
L_6						I	I	I	
L_7						I	I	I	
L_8						I	I	I	
…									

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1	1	1	0	0	0	0	0	
L_2	1	1	1	0	0	0	0	0	
L_3	1	1	1	0	0	0	0	0	
L_4	0	0	0	1	1	0	0	0	
L_5	0	0	0	1	1	0	0	0	
L_6	0	0	0	0	0	1	1	1	
L_7	0	0	0	0	0	1	1	1	
L_8	0	0	0	0	0	1	1	1	
…									

Undifferentiated Typology

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1	1	1	0.37	0.37	0.28	0.28	0.28	
L_2	1	1	1	0.37	0.37	0.28	0.28	0.28	
L_3	1	1	1	0.37	0.37	0.28	0.28	0.28	
L_4	0.37	0.37	0.37	1	1	0.58	0.58	0.58	
L_5	0.37	0.37	0.37	1	1	0.58	0.58	0.58	
L_6	0.28	0.28	0.28	0.58	0.58	1	1	1	
L_7	0.28	0.28	0.28	0.58	0.58	1	1	1	
L_8	0.28	0.28	0.28	0.58	0.58	1	1	1	
…									

Inter-type similarities

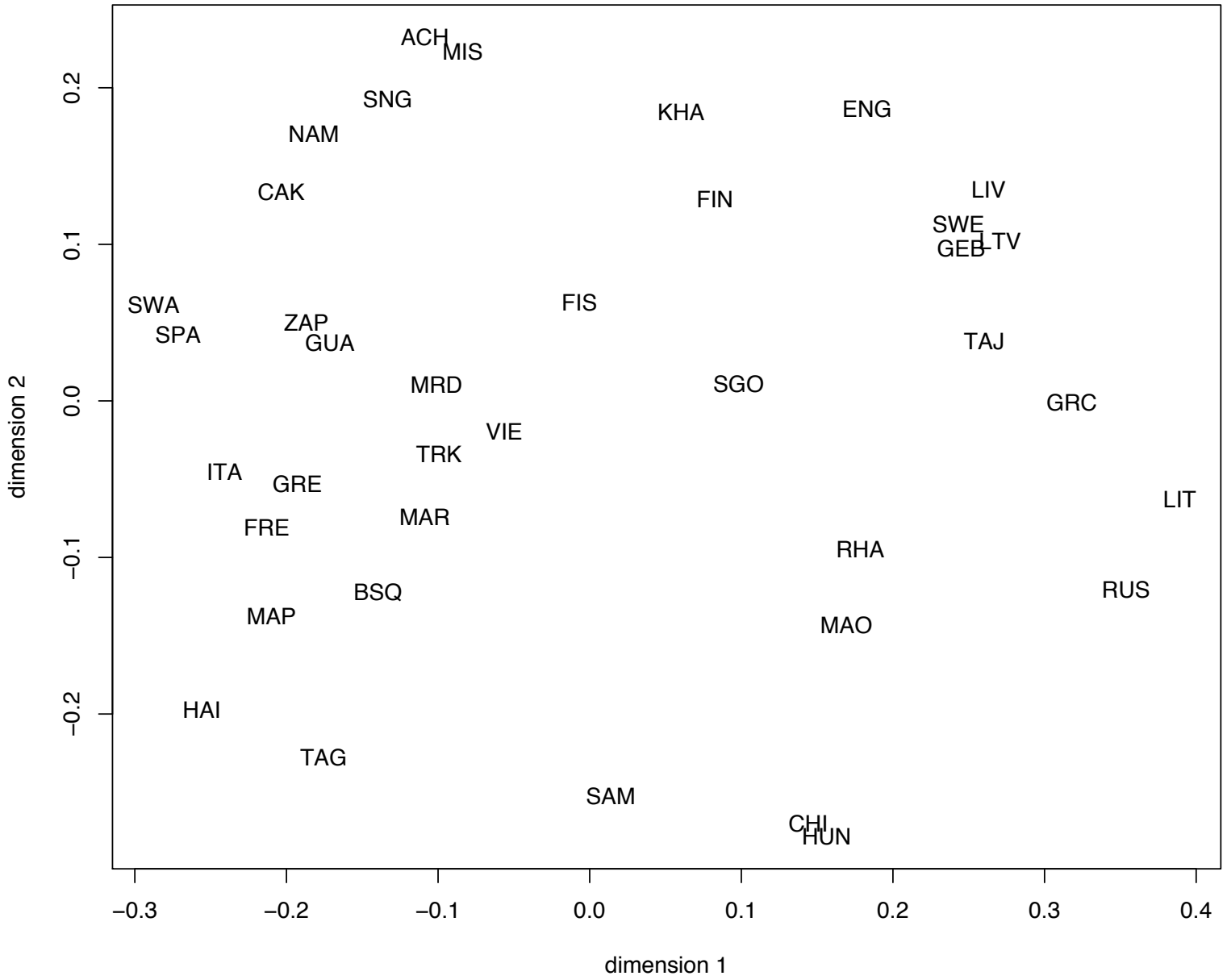
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1								
L_2		1							
L_3			1						
L_4				1					
L_5					1				
L_6						1			
L_7							1		
L_8								1	
…									

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1	0.55	0.72	0.31	0.70	0.61	0.50	0.58	
L_2	0.55	1	0.55	0.31	0.40	0.44	0.31	0.48	
L_3	0.72	0.55	1	0.29	0.53	0.51	0.48	0.60	
L_4	0.31	0.31	0.29	1	0.38	0.36	0.26	0.27	
L_5	0.70	0.40	0.53	0.38	1	0.64	0.51	0.46	
L_6	0.61	0.44	0.51	0.36	0.64	1	0.57	0.43	
L_7	0.50	0.31	0.48	0.26	0.51	0.57	1	0.47	
L_8	0.58	0.48	0.60	0.27	0.46	0.43	0.47	1	
…									

‘Deconstructed’ Typology

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher

	ACH	ALB	ARM	AVA	BSQ	CAK	CHA	CHI	CHO	EFI	ENG	EST	EWE	FIJ	FIN	FIS	FRE	GEB	GRC	GRE	GUA	HAI	HAU	HAW	HIN	HUN	ICE	ITA	JAM	KAT	KHA	LIT	LIV	LTV	MAO	MAP	MA
ACH	1	0.43	0.62	0.31	0.41	0.57	0.6	0.22	0.41	0.51	0.55	0.46	0.54	0.35	0.46	0.35	0.45	0.44	0.4	0.39	0.52	0.41	0.72	0.41	0.55	0.23	0.48	0.47	0.32	0.36	0.55	0.37	0.47	0.45	0.36	0.43	0.4
ALB	0.43	1	0.53	0.26	0.41	0.45	0.41	0.2	0.32	0.34	0.38	0.3	0.33	0.26	0.31	0.29	0.4	0.33	0.25	0.41	0.53	0.35	0.49	0.29	0.34	0.28	0.3	0.5	0.3	0.28	0.37	0.26	0.28	0.33	0.28	0.4	0.4
ARM	0.62	0.53	1	0.27	0.4	0.63	0.61	0.26	0.37	0.4	0.54	0.46	0.47	0.34	0.42	0.4	0.41	0.44	0.4	0.39	0.57	0.33	0.67	0.36	0.53	0.24	0.46	0.49	0.32	0.31	0.49	0.34	0.46	0.49	0.34	0.37	0.3
AVA	0.31	0.26	0.27	1	0.27	0.24	0.27	0.25	0.18	0.22	0.32	0.36	0.27	0.31	0.33	0.21	0.26	0.32	0.24	0.25	0.31	0.23	0.26	0.28	0.32	0.25	0.34	0.26	0.28	0.25	0.32	0.29	0.33	0.34	0.3	0.27	0.3
BSQ	0.41	0.41	0.4	0.27	1	0.41	0.34	0.22	0.25	0.28	0.29	0.28	0.32	0.29	0.32	0.21	0.36	0.26	0.23	0.37	0.44	0.33	0.38	0.3	0.32	0.26	0.28	0.42	0.25	0.32	0.34	0.25	0.27	0.3	0.34	0.36	0.3
CAK	0.57	0.45	0.63	0.24	0.41	1	0.55	0.21	0.39	0.4	0.4	0.38	0.42	0.3	0.36	0.34	0.44	0.31	0.32	0.37	0.56	0.35	0.61	0.31	0.43	0.21	0.37	0.48	0.27	0.26	0.41	0.29	0.38	0.39	0.29	0.41	0.3
CHA	0.6	0.41	0.61	0.27	0.34	0.55	1	0.24	0.41	0.41	0.6	0.41	0.46	0.33	0.43	0.39	0.43	0.43	0.39	0.43	0.47	0.33	0.59	0.39	0.45	0.24	0.5	0.46	0.3	0.35	0.57	0.33	0.43	0.46	0.33	0.37	0.3
CHI	0.22	0.2	0.26	0.25	0.22	0.21	0.24	1	0.16	0.2	0.27	0.37	0.22	0.28	0.26	0.19	0.19	0.31	0.28	0.18	0.25	0.18	0.23	0.23	0.25	0.29	0.3	0.2	0.33	0.25	0.3	0.29	0.33	0.33	0.25	0.25	0.2
CHO	0.41	0.32	0.37	0.18	0.25	0.39	0.41	0.16	1	0.29	0.28	0.24	0.31	0.21	0.26	0.22	0.27	0.24	0.22	0.27	0.33	0.27	0.45	0.23	0.27	0.17	0.26	0.32	0.21	0.26	0.31	0.21	0.24	0.27	0.2	0.29	0.3
EFI	0.51	0.34	0.4	0.22	0.28	0.4	0.41	0.2	0.29	1	0.32	0.27	0.36	0.23	0.31	0.27	0.32	0.26	0.23	0.32	0.33	0.32	0.51	0.26	0.35	0.18	0.29	0.37	0.22	0.27	0.41	0.2	0.29	0.27	0.25	0.32	0.3
ENG	0.55	0.38	0.54	0.32	0.29	0.4	0.6	0.27	0.28	0.32	1	0.58	0.47	0.47	0.48	0.37	0.32	0.52	0.5	0.34	0.39	0.25	0.5	0.46	0.54	0.33	0.6	0.37	0.44	0.35	0.58	0.5	0.56	0.6	0.41	0.31	0.3
EST	0.46	0.3	0.46	0.36	0.28	0.38	0.41	0.37	0.24	0.27	0.58	1	0.46	0.49	0.48	0.3	0.28	0.58	0.49	0.25	0.39	0.21	0.41	0.38	0.48	0.39	0.63	0.29	0.46	0.31	0.58	0.6	0.8	0.79	0.42	0.31	0.3
EWE	0.54	0.33	0.47	0.27	0.32	0.42	0.46	0.22	0.31	0.36	0.47	0.46	1	0.29	0.42	0.29	0.35	0.4	0.33	0.3	0.38	0.31	0.52	0.31	0.52	0.25	0.46	0.32	0.32	0.31	0.49	0.35	0.43	0.49	0.32	0.33	0.3
FIJ	0.35	0.26	0.34	0.31	0.29	0.3	0.33	0.28	0.21	0.23	0.47	0.49	0.29	1	0.32	0.19	0.22	0.36	0.47	0.22	0.34	0.21	0.36	0.73	0.34	0.36	0.43	0.28	0.35	0.32	0.38	0.66	0.45	0.48	0.67	0.33	0.
FIN	0.46	0.31	0.42	0.33	0.32	0.36	0.43	0.26	0.26	0.31	0.48	0.48	0.42	0.32	1	0.42	0.31	0.45	0.36	0.33	0.38	0.24	0.45	0.3	0.5	0.28	0.48	0.36	0.37	0.28	0.48	0.38	0.46	0.48	0.3	0.27	0.3
FIS	0.35	0.29	0.4	0.21	0.21	0.34	0.39	0.19	0.22	0.27	0.37	0.3	0.29	0.19	0.42	1	0.32	0.36	0.25	0.29	0.35	0.21	0.39	0.22	0.41	0.21	0.34	0.32	0.26	0.19	0.34	0.25	0.32	0.33	0.2	0.19	0.2
FRE	0.45	0.4	0.41	0.26	0.36	0.44	0.43	0.19	0.27	0.32	0.32	0.28	0.35	0.22	0.31	0.32	1	0.3	0.22	0.43	0.43	0.52	0.39	0.24	0.36	0.21	0.35	0.51	0.25	0.26	0.34	0.22	0.26	0.28	0.23	0.35	0.3
GEB	0.44	0.33	0.44	0.32	0.26	0.31	0.43	0.31	0.24	0.26	0.52	0.58	0.4	0.36	0.45	0.36	0.3	1	0.41	0.26	0.36	0.2	0.41	0.36	0.49	0.31	0.57	0.3	0.39	0.25	0.46	0.48	0.56	0.59	0.33	0.22	0
GRC	0.4	0.25	0.4	0.24	0.23	0.32	0.39	0.28	0.22	0.23	0.5	0.49	0.33	0.47	0.36	0.25	0.22	0.41	1	0.21	0.29	0.2	0.38	0.45	0.37	0.34	0.44	0.25	0.37	0.25	0.42	0.71	0.52	0.51	0.4	0.22	0.2
GRE	0.39	0.41	0.39	0.25	0.37	0.37	0.43	0.18	0.27	0.32	0.34	0.25	0.3	0.22	0.33	0.29	0.43	0.26	0.21	1	0.38	0.38	0.42	0.24	0.31	0.23	0.28	0.42	0.24	0.27	0.39	0.2	0.24	0.28	0.24	0.35	0.3
GUA	0.52	0.53	0.57	0.31	0.44	0.56	0.47	0.25	0.33	0.33	0.39	0.39	0.38	0.34	0.38	0.35	0.43	0.36	0.29	0.38	1	0.38	0.51	0.33	0.48	0.28	0.38	0.47	0.31	0.3	0.46	0.3	0.37	0.42	0.35	0.49	0.4
HAI	0.41	0.35	0.33	0.23	0.33	0.35	0.33	0.18	0.27	0.32	0.25	0.21	0.31	0.21	0.24	0.21	0.52	0.2	0.2	0.38	0.38	1	0.39	0.23	0.25	0.21	0.22	0.44	0.21	0.25	0.28	0.18	0.19	0.23	0.21	0.42	0.3
HAU	0.72	0.49	0.67	0.26	0.38	0.61	0.59	0.23	0.45	0.51	0.5	0.41	0.52	0.36	0.45	0.39	0.39	0.41	0.38	0.42	0.51	0.39	1	0.38	0.51	0.24	0.43	0.47	0.33	0.31	0.51	0.33	0.43	0.44	0.34	0.44	0.4
HAW	0.41	0.29	0.36	0.28	0.3	0.31	0.39	0.23	0.23	0.26	0.46	0.38	0.31	0.73	0.3	0.22	0.24	0.36	0.45	0.24	0.33	0.23	0.38	1	0.37	0.35	0.37	0.31	0.35	0.34	0.38	0.59	0.38	0.41	0.67	0.32	0.3
HIN	0.55	0.34	0.53	0.32	0.32	0.43	0.45	0.25	0.27	0.35	0.54	0.48	0.52	0.34	0.5	0.41	0.36	0.49	0.37	0.31	0.48	0.25	0.51	0.37	1	0.31	0.51	0.35	0.35	0.33	0.57	0.41	0.53	0.53	0.37	0.29	0.3
HUN	0.23	0.28	0.24	0.25	0.26	0.21	0.24	0.29	0.17	0.18	0.33	0.39	0.25	0.36	0.28	0.21	0.21	0.31	0.34	0.23	0.28	0.21	0.24	0.35	0.31	1	0.3	0.26	0.39	0.28	0.3	0.39	0.35	0.42	0.35	0.28	0.2
ICE	0.48	0.3	0.46	0.34	0.28	0.37	0.5	0.3	0.26	0.29	0.6	0.63	0.46	0.43	0.48	0.34	0.35	0.57	0.44	0.28	0.38	0.22	0.43	0.37	0.51	0.3	1	0.34	0.39	0.28	0.55	0.48	0.58	0.61	0.35	0.26	0.3
ITA	0.47	0.5	0.49	0.26	0.42	0.48	0.46	0.2	0.32	0.37	0.37	0.29	0.32	0.28	0.36	0.32	0.51	0.3	0.25	0.42	0.47	0.44	0.47	0.31	0.35	0.26	0.34	1	0.24	0.27	0.36	0.26	0.27	0.33	0.28	0.44	0.3
JAM	0.32	0.3	0.32	0.28	0.25	0.27	0.3	0.33	0.21	0.22	0.44	0.46	0.32	0.35	0.37	0.26	0.25	0.39	0.37	0.24	0.31	0.21	0.33	0.35	0.35	0.39	0.39	0.24	1	0.32	0.41	0.42	0.4	0.38	0.34	0.32	0.2
KAT	0.36	0.28	0.31	0.25	0.32	0.26	0.35	0.25	0.26	0.27	0.35	0.31	0.31	0.32	0.28	0.19	0.26	0.25	0.25	0.27	0.3	0.25	0.31	0.34	0.33	0.28	0.28	0.27	0.32	1	0.38	0.27	0.3	0.33	0.35	0.34	0.3
KHA	0.55	0.37	0.49	0.32	0.34	0.41	0.57	0.3	0.31	0.41	0.58	0.58	0.49	0.38	0.48	0.34	0.34	0.46	0.42	0.39	0.46	0.28	0.51	0.38	0.57	0.3	0.55	0.36	0.41	0.38	1	0.37	0.56	0.53	0.41	0.37	0.3
LIT	0.37	0.26	0.34	0.29	0.25	0.29	0.33	0.29	0.21	0.2	0.5	0.6	0.35	0.66	0.38	0.25	0.22	0.48	0.71	0.2	0.3	0.18	0.33	0.59	0.41	0.39	0.48	0.26	0.42	0.27	0.37	1	0.56	0.68	0.55	0.25	0.2
LIV	0.47	0.28	0.46	0.33	0.27	0.38	0.43	0.33	0.24	0.29	0.56	0.8	0.43	0.45	0.46	0.32	0.26	0.56	0.52	0.24	0.37	0.19	0.43	0.38	0.53	0.35	0.58	0.27	0.4	0.3	0.56	0.56	1	0.75	0.39	0.26	0.3
LTV	0.45	0.33	0.49	0.34	0.3	0.39	0.46	0.33	0.27	0.27	0.6	0.79	0.49	0.48	0.48	0.33	0.28	0.59	0.51	0.28	0.42	0.23	0.44	0.41	0.53	0.42	0.61	0.33	0.38	0.33	0.53	0.68	0.75	1	0.43	0.31	0.3
MAO	0.36	0.28	0.34	0.3	0.34	0.29	0.33	0.25	0.2	0.25	0.41	0.42	0.32	0.67	0.3	0.2	0.23	0.33	0.4	0.24	0.35	0.21	0.34	0.67	0.37	0.35	0.35	0.28	0.34	0.35	0.41	0.55	0.39	0.43	1	0.3	0.3
MAP	0.43	0.4	0.37	0.27	0.36	0.41	0.37	0.25	0.29	0.32	0.31	0.31	0.33	0.33	0.27	0.19	0.35	0.22	0.22	0.35	0.49	0.42	0.44	0.32	0.29	0.28	0.26	0.44	0.32	0.34	0.37	0.25	0.26	0.31	0.3	1	0.4
MAR	0.42	0.44	0.39	0.32	0.36	0.39	0.38	0.24	0.32	0.32	0.35	0.33	0.34	0.3	0.35	0.26	0.35	0.3	0.25	0.31	0.46	0.33	0.44	0.31	0.38	0.28	0.32	0.38	0.29	0.31	0.37	0.26	0.31	0.35	0.32	0.46	
MIS	0.64	0.39	0.56	0.29	0.36	0.5	0.6	0.23	0.36	0.43	0.6	0.41	0.5	0.36	0.44	0.34	0.39	0.41	0.38	0.39	0.48	0.36	0.67	0.38	0.5	0.25	0.46	0.45	0.3	0.32	0.54	0.33	0.44	0.47	0.34	0.43	0.4
MIZ	0.48	0.34	0.42	0.27	0.34	0.4	0.39	0.22	0.29	0.31	0.37	0.35	0.33	0.57	0.28	0.2	0.3	0.3	0.37	0.3	0.4	0.29	0.47	0.6	0.35	0.3	0.32	0.34	0.3	0.32	0.46	0.43	0.33	0.34	0.59	0.36	0.3
MLT	0.64	0.58	0.72	0.29	0.42	0.57	0.54	0.24	0.33	0.43	0.52	0.41	0.46	0.32	0.42	0.35	0.43	0.41	0.34	0.37	0.62	0.36	0.61	0.34	0.48	0.25	0.45	0.52	0.31	0.28	0.52	0.31	0.42	0.48	0.33	0.39	0.4
MRD	0.46	0.4	0.49	0.35	0.36	0.46	0.5	0.23	0.33	0.33	0.38	0.32	0.37	0.28	0.39	0.35	0.4	0.36	0.31	0.32	0.45	0.35	0.45	0.31	0.42	0.22	0.39	0.41	0.27	0.28	0.4	0.28	0.36	0.37	0.29	0.33	0
MRS	0.51	0.32	0.45	0.3	0.33	0.38	0.5	0.27	0.29	0.34	0.52	0.44	0.36	0.67	0.37	0.27	0.29	0.41	0.55	0.27	0.38	0.26	0.5	0.64	0.41	0.31	0.44	0.36	0.4	0.36	0.46	0.61	0.44	0.46	0.58	0.37	0.3
NAM	0.62	0.46	0.59	0.29	0.37	0.52	0.56	0.21	0.39	0.46	0.47	0.38	0.49	0.36	0.41	0.29	0.42	0.33	0.33	0.44	0.5	0.37	0.68	0.37	0.49	0.26	0.39	0.47	0.29	0.35	0.54	0.3	0.37	0.42	0.34	0.45	0.4
NIC	0.48	0.33	0.41	0.24	0.28	0.39	0.52	0.21	0.29	0.33	0.41	0.35	0.36	0.28	0.32	0.3	0.37	0.31	0.33	0.33	0.41	0.26	0.44	0.34	0.4	0.24	0.34	0.35	0.27	0.3	0.42	0.29	0.34	0.36	0.3	0.3	0.3
OSS	0.29	0.23	0.34	0.3	0.22	0.25	0.28	0.33	0.17	0.19	0.46	0.58	0.31	0.64	0.31	0.21	0.17	0.46	0.62	0.17	0.27	0.16	0.3	0.5	0.34	0.31	0.42	0.2	0.4	0.25	0.33	0.8	0.5	0.57	0.5	0.23	0.2
PAP	0.49	0.41	0.45	0.35	0.4	0.43	0.53	0.28	0.33	0.33	0.51	0.49	0.4	0.41	0.43	0.29	0.38	0.42	0.35	0.42	0.51	0.3	0.5	0.41	0.46	0.33	0.48	0.43	0.4	0.35	0.57	0.37	0.44	0.47	0.39	0.48	0.4
PIR	0.63	0.47	0.59	0.31	0.44	0.58	0.52	0.22	0.36	0.46	0.41	0.4	0.45	0.34	0.41	0.31	0.45	0.34	0.33	0.42	0.59	0.41	0.61	0.35	0.43	0.26	0.39	0.54	0.32	0.32	0.53	0.31	0.37	0.42	0.34	0.52	0.4
PIT	0.55	0.4	0.48	0.29	0.38	0.48	0.53	0.25	0.31	0.39	0.47	0.41	0.4	0.52	0.35	0.29	0.4	0.36	0.4	0.37	0.47	0.3	0.58	0.49	0.45	0.28	0.41	0.42	0.35	0.32	0.48	0.44	0.39	0.41	0.49	0.36	0.3
RHA	0.33	0.3	0.35	0.37	0.3	0.27	0.3	0.31	0.16	0.21	0.44	0.48	0.32	0.41	0.35	0.24	0.29	0.45	0.3	0.25	0.35	0.21	0.3	0.38	0.41	0.37	0.46	0.3	0.43	0.29	0.41	0.39	0.43	0.46	0.38	0.26	0.2
RUM	0.64	0.52	0.63	0.3	0.43	0.58	0.56	0.22	0.39	0.44	0.53	0.4	0.5	0.32	0.44	0.37	0.48	0.38	0.34	0.46	0.55	0.41	0.62	0.37	0.5	0.23	0.42	0.58	0.32	0.31	0.53	0.3	0.38	0.43	0.33	0.42	0.4
RUS	0.29	0.23	0.33	0.25	0.21	0.27	0.3	0.3	0.18	0.18	0.46	0.51	0.3	0.54	0.31	0.2	0.18	0.4	0.65	0.18	0.26	0.18	0.31	0.51	0.34	0.36	0.39	0.22	0.39	0.25	0.35	0.77	0.43	0.54	0.45	0.25	0.2
SAA	0.52	0.32	0.51	0.31	0.28	0.42	0.54	0.31	0.26	0.32	0.59	0.65	0.47	0.4	0.55	0.4	0.33	0.55	0.48	0.27	0.4	0.23	0.48	0.37	0.55	0.32	0.65	0.36	0.41	0.29	0.56	0.5	0.62	0.63	0.34	0.28	0.3
SAM	0.26	0.22	0.24	0.22	0.2	0.22	0.29	0.24	0.19	0.23	0.28	0.28	0.23	0.31	0.24	0.18	0.23	0.21	0.24	0.2	0.24	0.2	0.27	0.28	0.25	0.25	0.27	0.24	0.24	0.24	0.28	0.24	0.24	0.25	0.26	0.27	0.2
SGO	0.42	0.37	0.4	0.31	0.31	0.3	0.33	0.27	0.29	0.28	0.45	0.45	0.39	0.31	0.34	0.22	0.26	0.38	0.36	0.27	0.39	0.26	0.4	0.3	0.43	0.29	0.39	0.28	0.41	0.25	0.47	0.4	0.41	0.44	0.32	0.3	0.3
SNG	0.66	0.48	0.62	0.27	0.36	0.55	0.61	0.22	0.44	0.45	0.51	0.41	0.49	0.32	0.39	0.32	0.4	0.36	0.36	0.45	0.5	0.36	0.64	0.36	0.48	0.25	0.46	0.46	0.33	0.33	0.53	0.33	0.4	0.43	0.32	0.43	0.4
SPA	0.53	0.52	0.61	0.24	0.45	0.62	0.48	0.21	0.38	0.37	0.38	0.34	0.37	0.29	0.35	0.28	0.47	0.27	0.28	0.43	0.56	0.42	0.62	0.32	0.37	0.27	0.31	0.6	0.28	0.26	0.4	0.28	0.32	0.35	0.3	0.51	0.3
SRA	0.42	0.31	0.41	0.29	0.3	0.3	0.39	0.34	0.25	0.28	0.44	0.49	0.38	0.3	0.39	0.28	0.29	0.52	0.42	0.3	0.34	0.23	0.37	0.32	0.42	0.29	0.48	0.29	0.38	0.28	0.46	0.4	0.46	0.43	0.32	0.25	0.2
SWA	0.57	0.44	0.49	0.22	0.36	0.54	0.43	0.17	0.37	0.44	0.31	0.28	0.36	0.27	0.34	0.28	0.39	0.24	0.26	0.38	0.5	0.47	0.59	0.27	0.33	0.22	0.29	0.48	0.24	0.25	0.38	0.23	0.28	0.31	0.25	0.48	0.3
SWE	0.45	0.25	0.4	0.32	0.27	0.31	0.43	0.28	0.26	0.25	0.52	0.51	0.41	0.41	0.5	0.3	0.29	0.6	0.43	0.24	0.33	0.19	0.39	0.39	0.52	0.33	0.62	0.29	0.4	0.3	0.47	0.53	0.49	0.54	0.36	0.21	0.2
TAG	0.27	0.23	0.22	0.15	0.18	0.22	0.29	0.14	0.22	0.3	0.19	0.14	0.21	0.12	0.2	0.19	0.29	0.17	0.13	0.28	0.21	0.3	0.26	0.15	0.18	0.13	0.18	0.28	0.12	0.21	0.22	0.11	0.16	0.17	0.13	0.26	0.2
TAJ	0.39	0.29	0.44	0.29	0.25	0.3	0.41	0.3	0.23	0.24	0.58	0.52	0.34	0.4	0.4	0.28	0.24	0.47	0.51	0.23	0.38	0.19	0.38	0.38	0.44	0.3	0.49	0.28	0.39	0.28	0.44	0.53	0.49	0.49	0.37	0.24	0.3
TKP	0.36	0.32	0.34	0.34	0.35	0.3	0.33	0.33	0.23	0.25	0.43	0.53	0.4	0.37	0.41	0.22	0.28	0.41	0.4	0.29	0.39	0.26	0.33	0.35	0.39	0.38	0.41	0.3	0.5	0.31	0.46	0.44	0.47	0.49	0.35	0.32	0.3
TOA	0.45	0.37	0.37	0.34	0.38	0.36	0.42	0.26	0.26	0.31	0.45	0.41	0.35	0.37	0.32	0.24	0.35	0.35	0.28	0.34	0.42	0.35	0.4	0.39	0.36	0.33	0.38	0.39	0.37	0.38	0.41	0.32	0.36	0.44	0.4	0.4	0.3
TOB	0.41	0.36	0.37	0.26	0.29	0.36	0.39	0.16	0.27	0.32	0.32	0.28	0.34	0.24	0.36	0.29	0.41	0.28	0.23	0.35	0.39	0.27	0.38	0.24	0.36	0.22	0.34	0.39	0.26	0.23	0.38	0.23	0.24	0.28	0.24	0.27	0.2
TON	0.48	0.29	0.4	0.3	0.27	0.35	0.47	0.24	0.29	0.3	0.6	0.48	0.38	0.46	0.38	0.32	0.31	0.36	0.37	0.31	0.37	0.23	0.48	0.4	0.43	0.28	0.44	0.33	0.35	0.33	0.5	0.37	0.42	0.45	0.35	0.34	0.3
TRK	0.48	0.5	0.46	0.34	0.38	0.42	0.43	0.24	0.28	0.34	0.42	0.37	0.41	0.29	0.4	0.32	0.41	0.4	0.28	0.37	0.5	0.35	0.48	0.31	0.42	0.29	0.38	0.43	0.31	0.28	0.39	0.29	0.35	0.4	0.31	0.37	0.4
UDM	0.57	0.46	0.54	0.3	0.34	0.49	0.5	0.24	0.35	0.39	0.42	0.35	0.41	0.31	0.37	0.33	0.4	0.33	0.32	0.38	0.5	0.37	0.57	0.34	0.46	0.23	0.39	0.47	0.27	0.28	0.43	0.31	0.35	0.38	0.3	0.38	0.4
ULA	0.46	0.33	0.4	0.27	0.32	0.37	0.37	0.24	0.28	0.34	0.36	0.38	0.34	0.49	0.3	0.22	0.27	0.31	0.42	0.29	0.41	0.3	0.46	0.5	0.35	0.3	0.34	0.37	0.29	0.33	0.4	0.47	0.39	0.44	0.5	0.38	0.3
VEP	0.44	0.32	0.45	0.35	0.32	0.38	0.47	0.29	0.27	0.3	0.49	0.49	0.42	0.34	0.76	0.42	0.31	0.47	0.39	0.32	0.38	0.25	0.43	0.34	0.48	0.27	0.46	0.32	0.35	0.26	0.48	0.41	0.48	0.51	0.3	0.24	0
VIE	0.46	0.36	0.42	0.26	0.34	0.37	0.45	0.22	0.32	0.29	0.41	0.34	0.31	0.36	0.33	0.23	0.32	0.32	0.28	0.33	0.41	0.3	0.42	0.4	0.38	0.28	0.36	0.41	0.28	0.36	0.4	0.31	0.31	0.37	0.36	0.36	0.3
YAB	0.33	0.32	0.29	0.19	0.27	0.31	0.29	0.17	0.27	0.32	0.25	0.25	0.25	0.28	0.26	0.19	0.33	0.2	0.22	0.29	0.32	0.33	0.39	0.25	0.26	0.24	0.25	0.33	0.21	0.24	0.3	0.21	0.24	0.26	0.22	0.32	0.2
ZAP	0.5	0.47	0.52	0.26	0.36	0.51	0.48	0.22	0.36	0.35	0.38	0.37	0.37	0.34	0.39	0.28	0.36	0.31	0.3	0.41	0.51	0.37	0.58	0.33	0.36	0.28	0.35	0.45	0.34	0.3	0.44	0.29	0.35	0.38	0.33	0.47	0.3


ACH
MIS
SNG
KHA
ENG
NAM
CAK
FIN
LIV
SWE
LTV
GEB
SWA
FIS
ZAP
SPA
GUA
TAJ
MRD
SGO
GRC
VIE
TRK
ITA
GRE
LIT
MAR
FRE
RHA
BSQ
RUS
MAP
MAO
HAI
TAG
SAM
CHI
HUN
dimension 2
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
dimension 1

5. Improving Typology

- Finer-graind Coding
- ‘Deconstructing’ Typology
- Select Suitable Characteristics

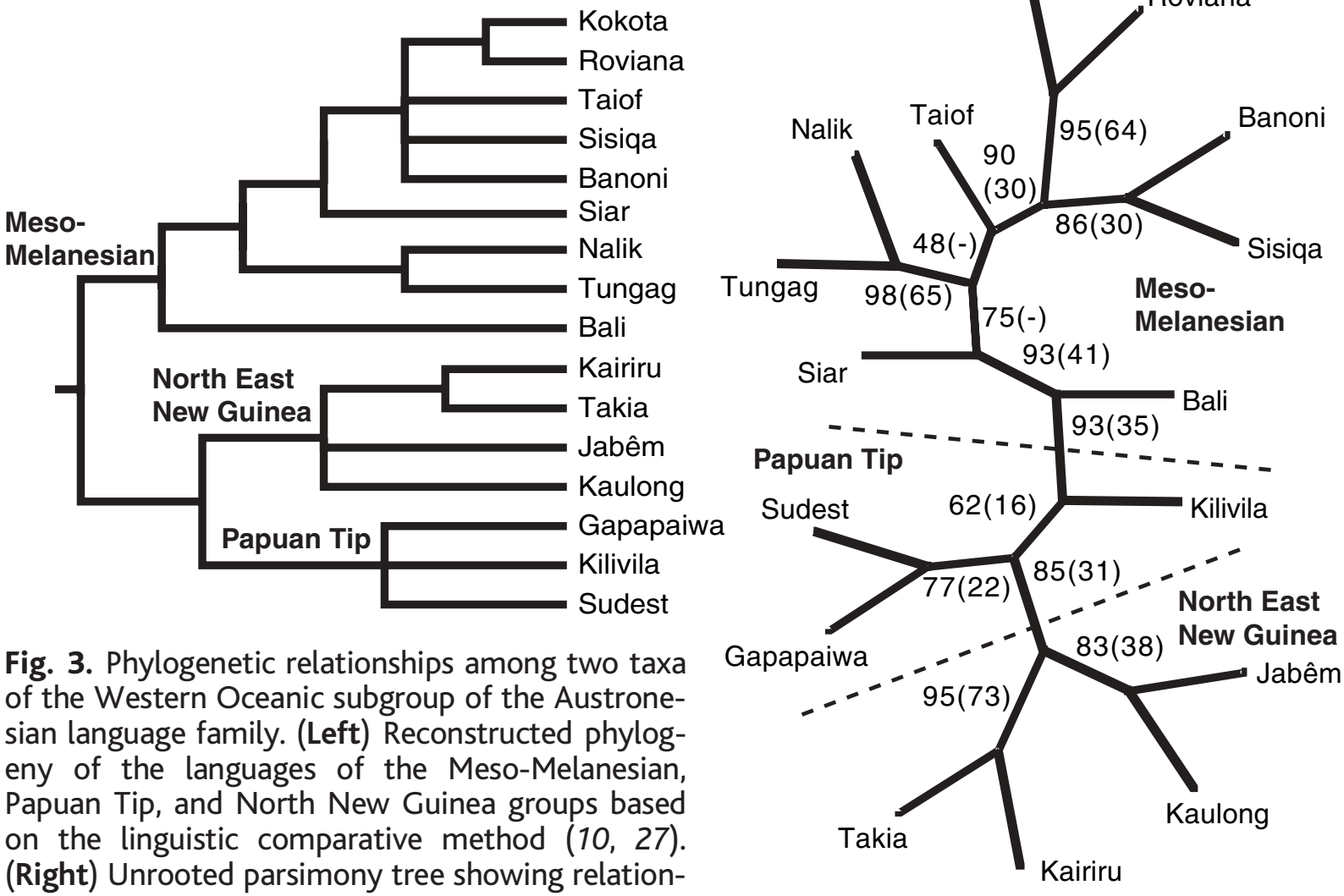
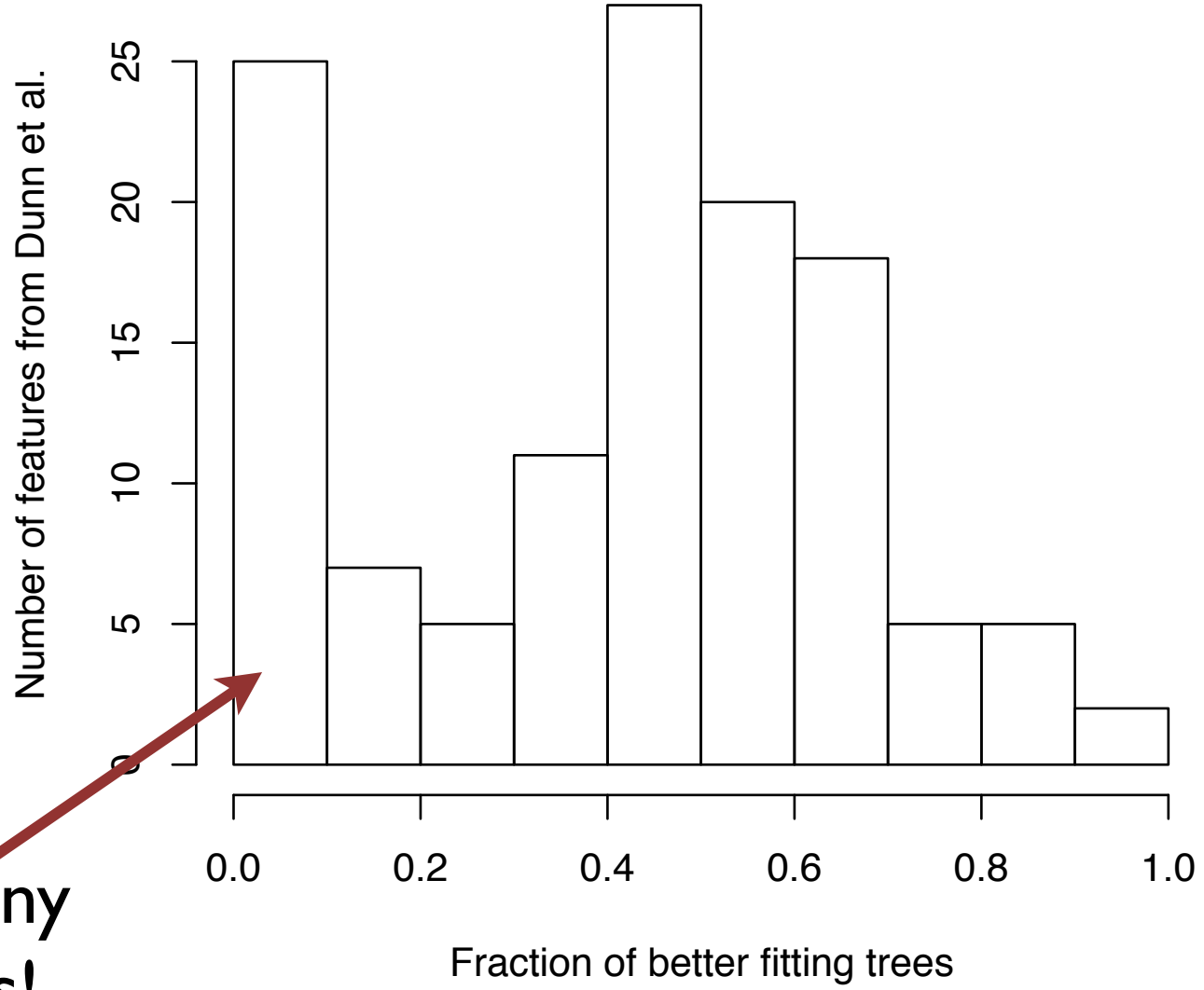


Fig. 3. Phylogenetic relationships among two taxa of the Western Oceanic subgroup of the Austronesian language family. (**Left**) Reconstructed phylogeny of the languages of the Meso-Melanesian, Papuan Tip, and North New Guinea groups based on the linguistic comparative method (*10*, *27*). (**Right**) Unrooted parsimony tree showing relationships among the Meso-Melanesian and Papuan Tip groups based on grammatical traits only (that is, discarding abundant lexical evidence) (the figure shows reweighted and raw bootstrap values). The two trees show a high degree of concordance, with monophyly in both major taxa and the similar geographical structuring of within-taxon diversity.

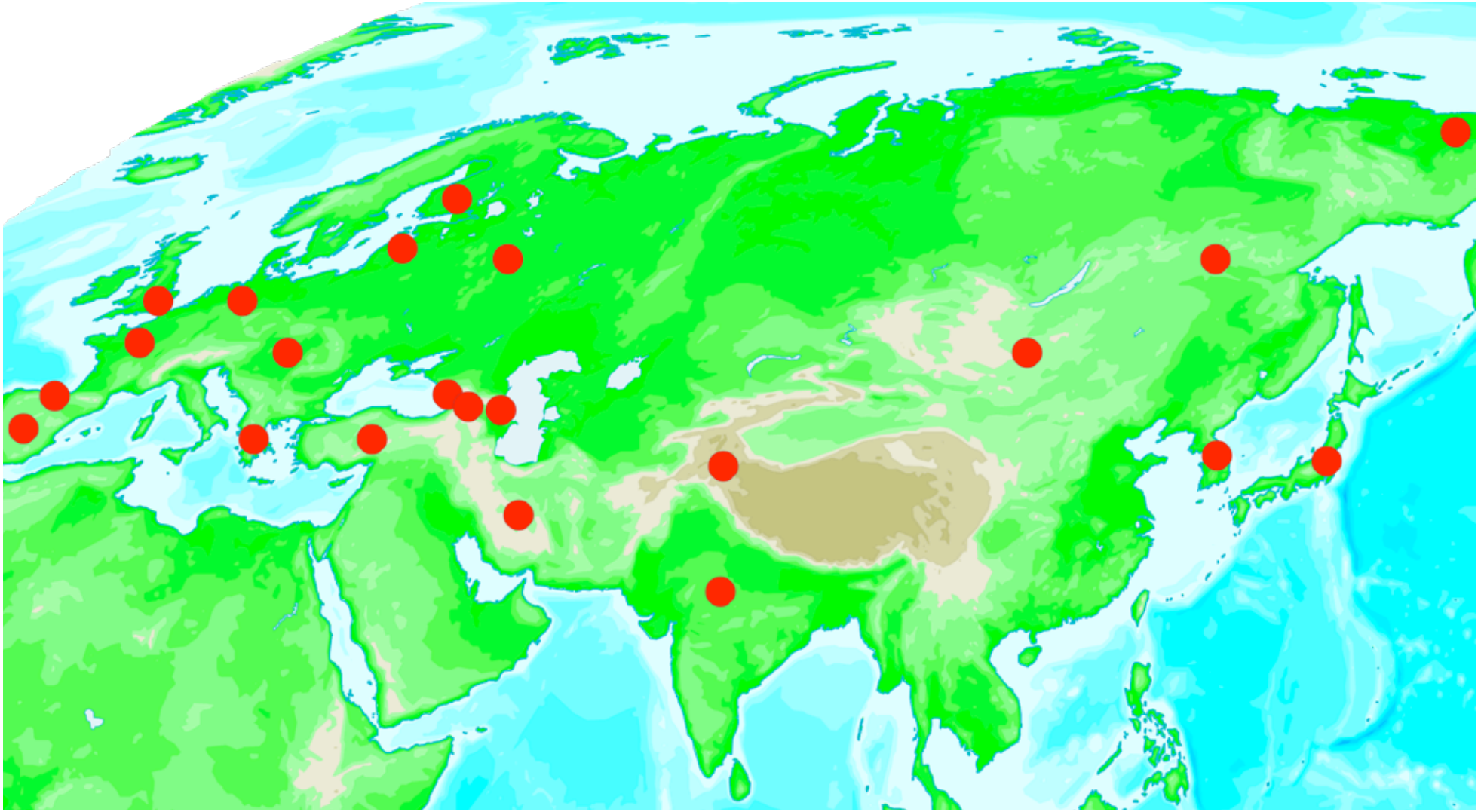
Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural Phylogenetics and the Reconstruction of Ancient Language History. Science 309: 2072-2075.

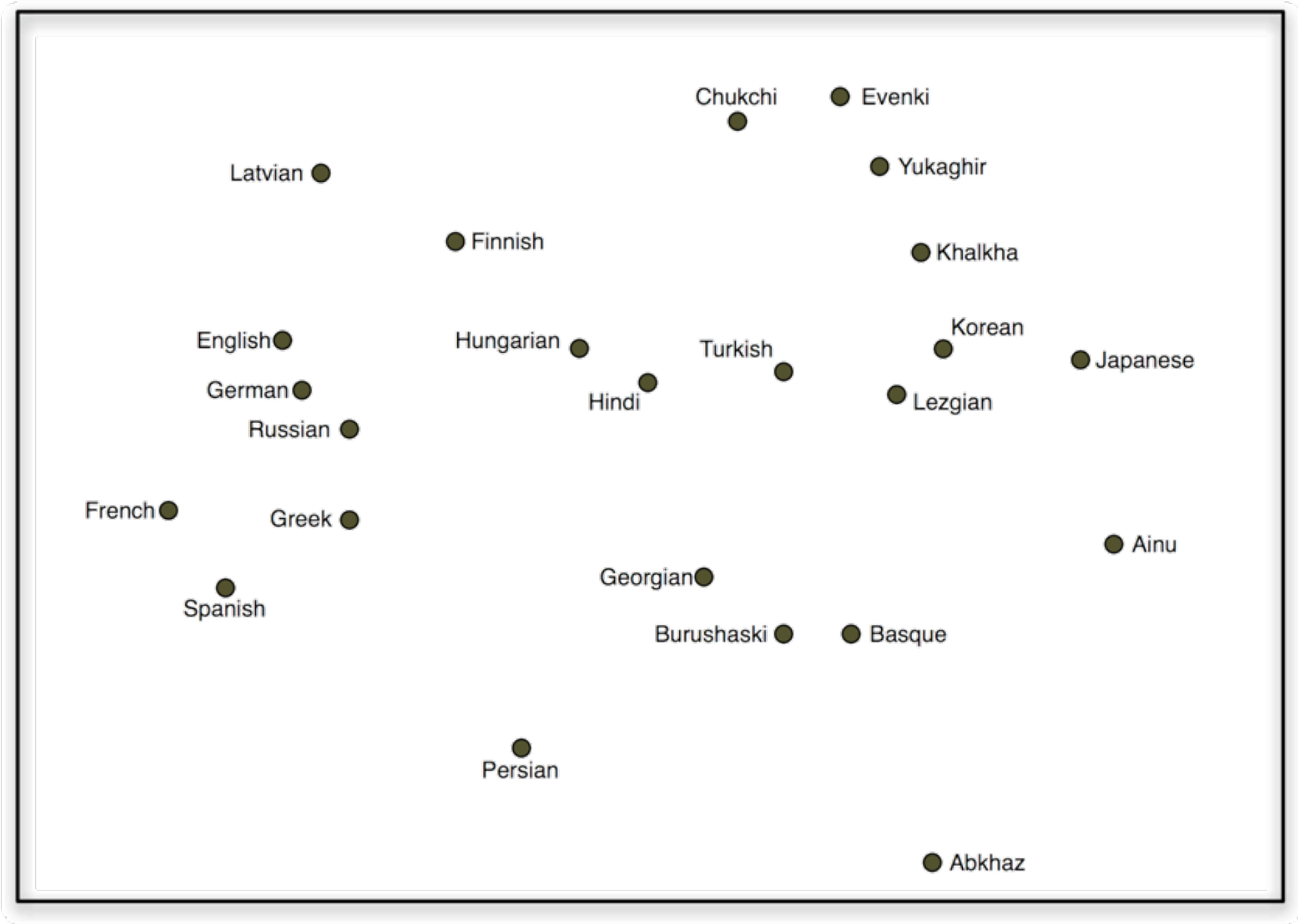
Fit for all 125 features

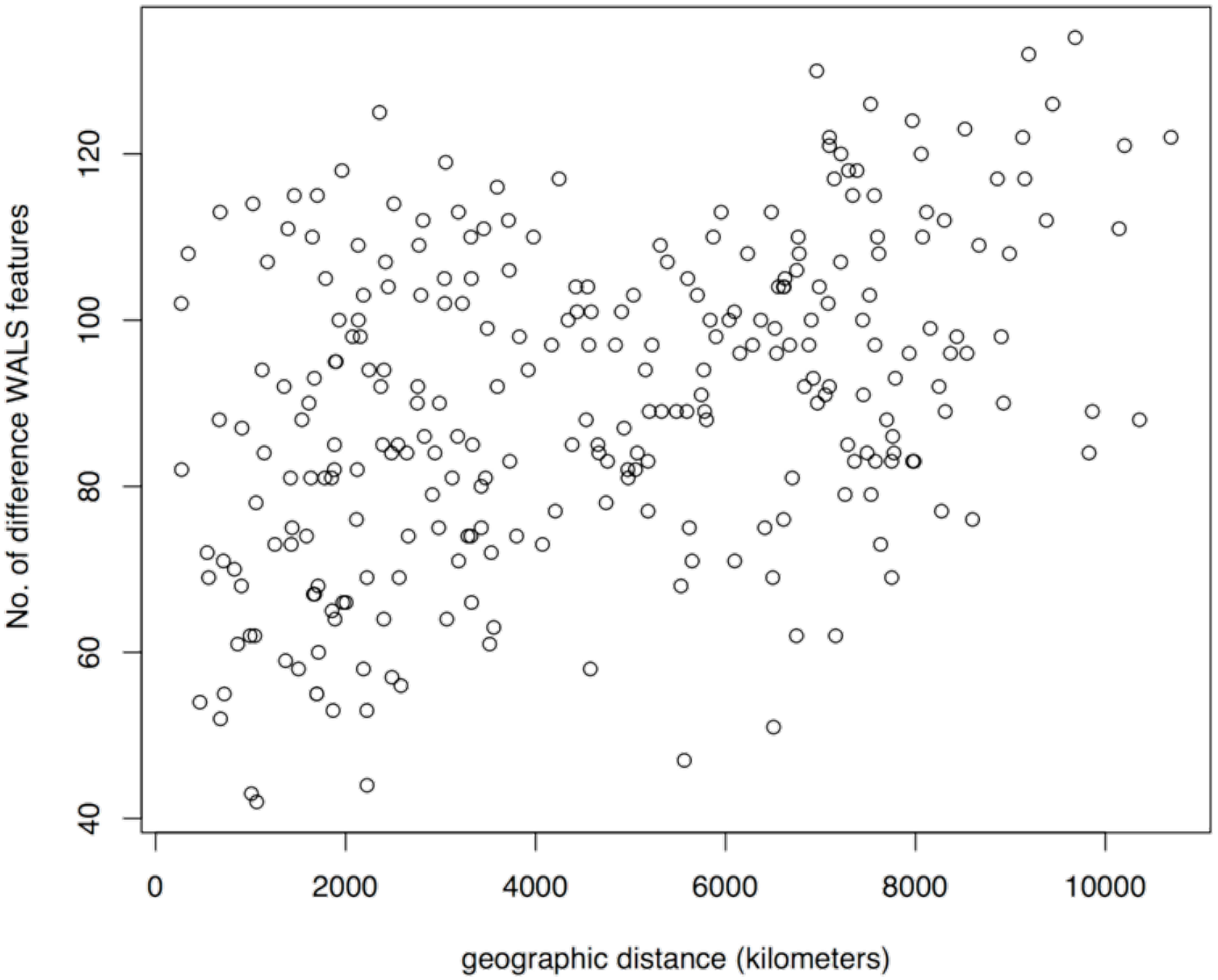


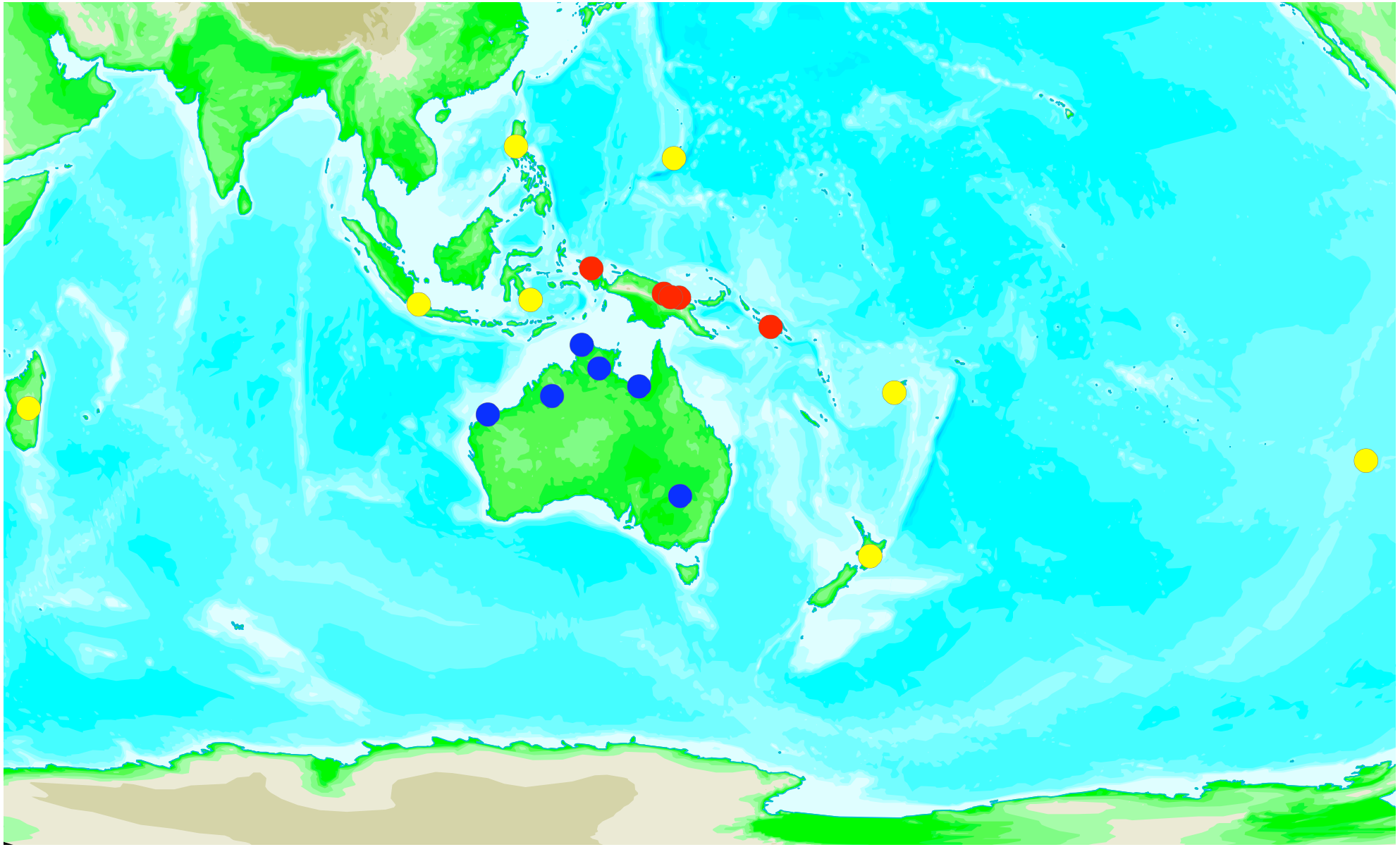
(Too) many good fits!

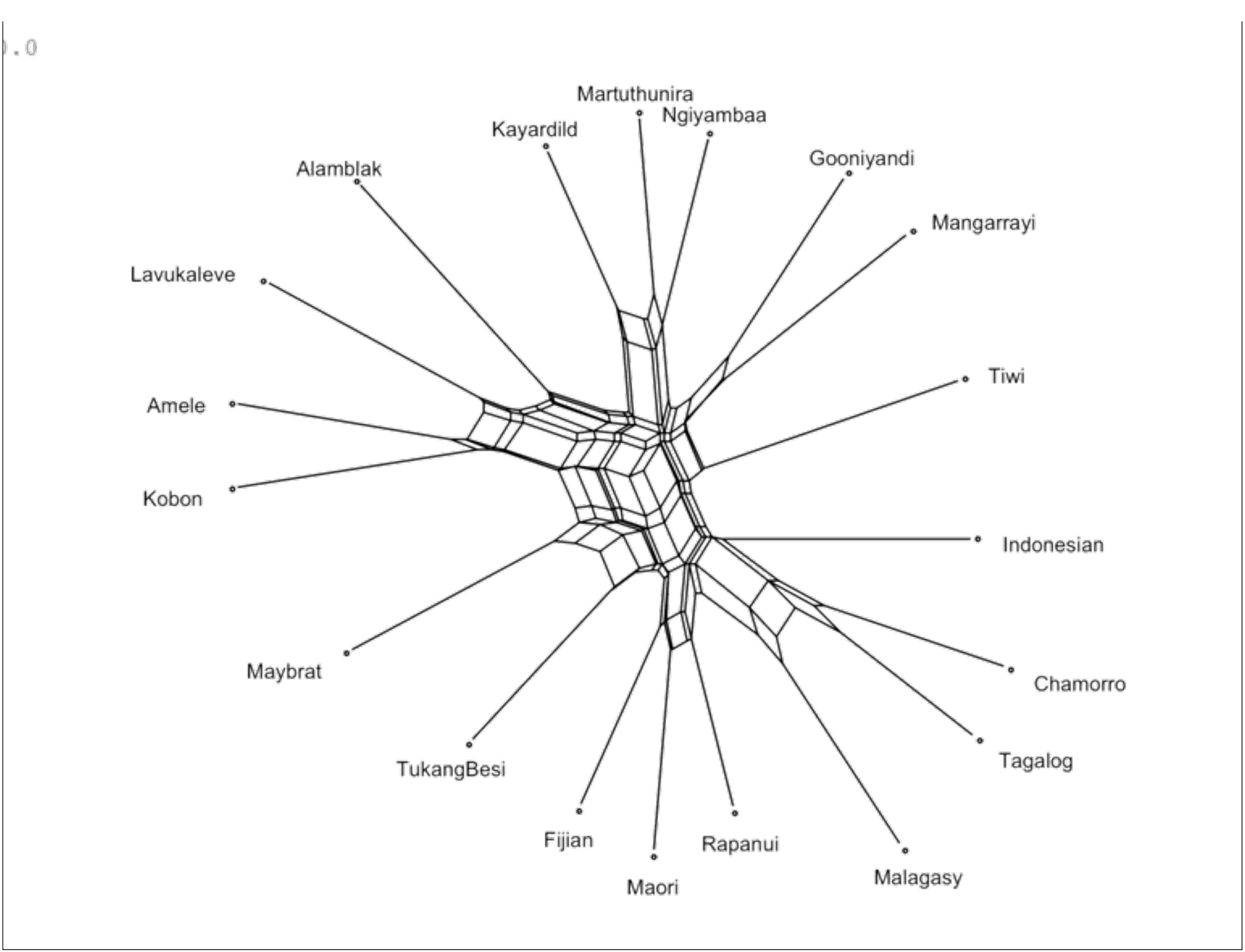
6. The Problem of Geographical Similarities

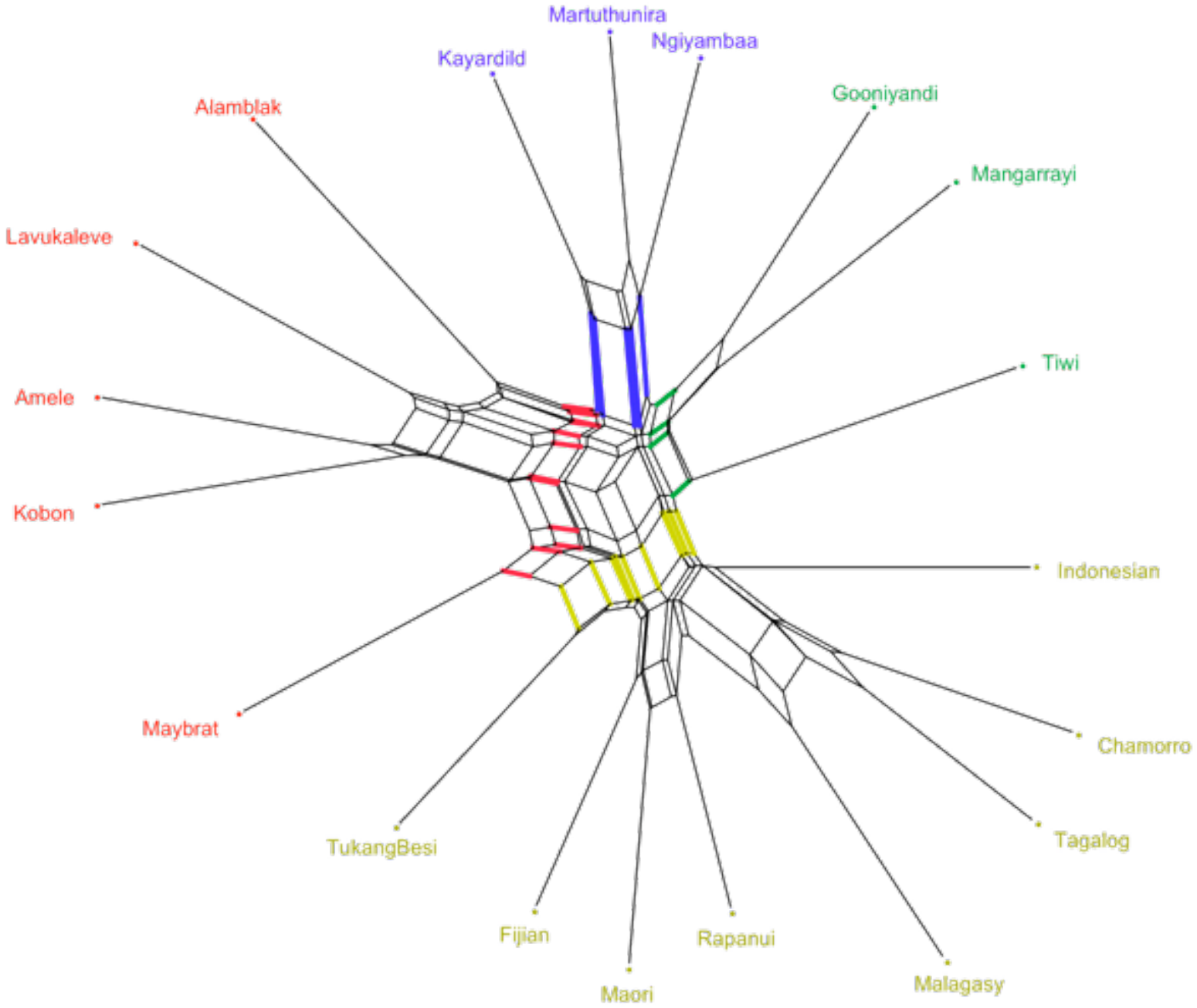



Chukchi
Evenki
Latvian
Yukaghir
Finnish
Khalkha
English
Hungarian
Turkish
Korean
Japanese
German
Hindi
Lezgian
Russian
French
Greek
Ainu
Georgian
Spanish
Burushaski
Basque
Persian
Abkhaz

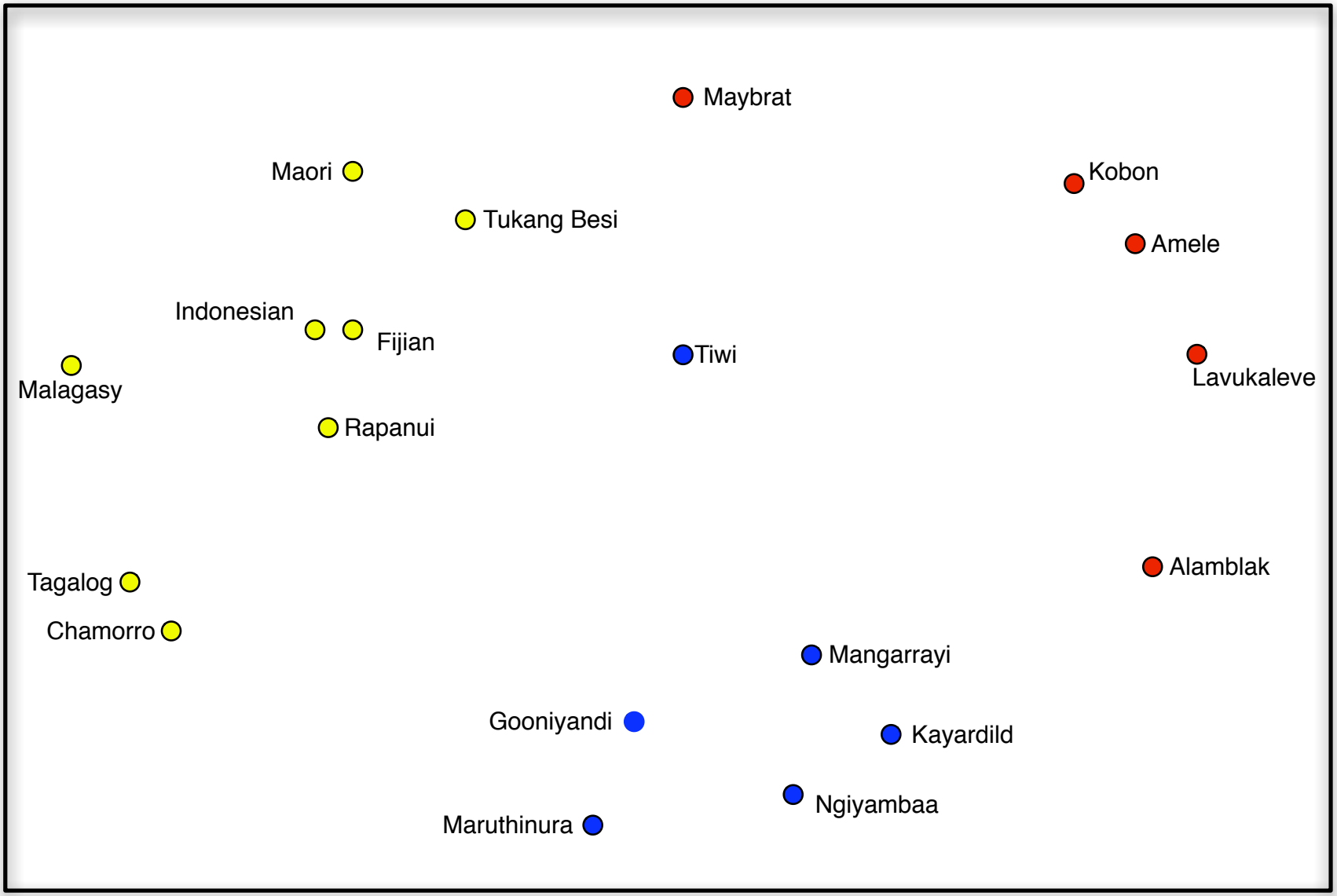

No. of difference WALS features
40
60
80
100
120
0
2000
4000
6000
8000
10000
geographic distance (kilometers)

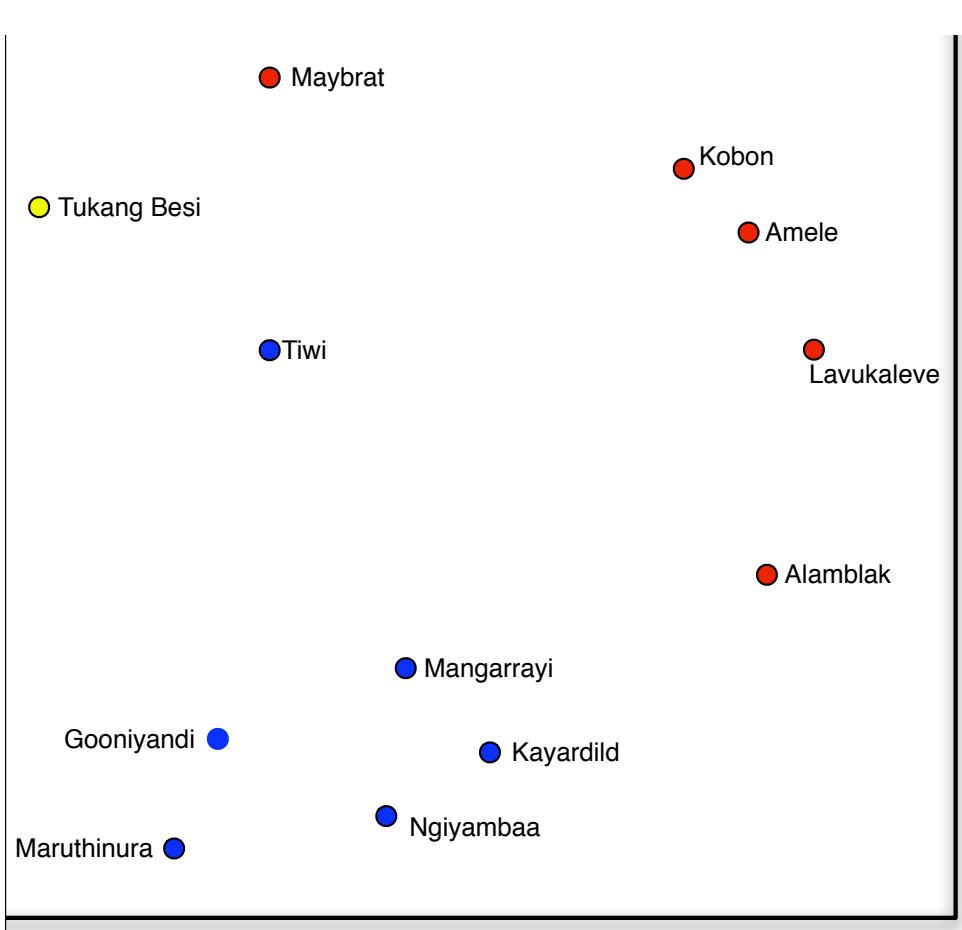



Martuthunira
Ngiyambaa
Kayardild
Alamblak
Gooniyandi
Mangarrayi
Lavukaleve
Tiwi
Amele
Kobon
Indonesian
Maybrat
Chamorro
Tagalog
TukangBesi
Fijian
Rapanui
Malagasy
Maori


Martuthunira
Ngiyambaa
Kayardild
Gooniyandi
Alamblak
Mangarrayi
Lavukaleve
Tiwi
Amele
Kobon
Indonesian
Maybrat
Chamorro
Tagalog
TukangBesi
Fijian
Rapanui
Malagasy
Maori

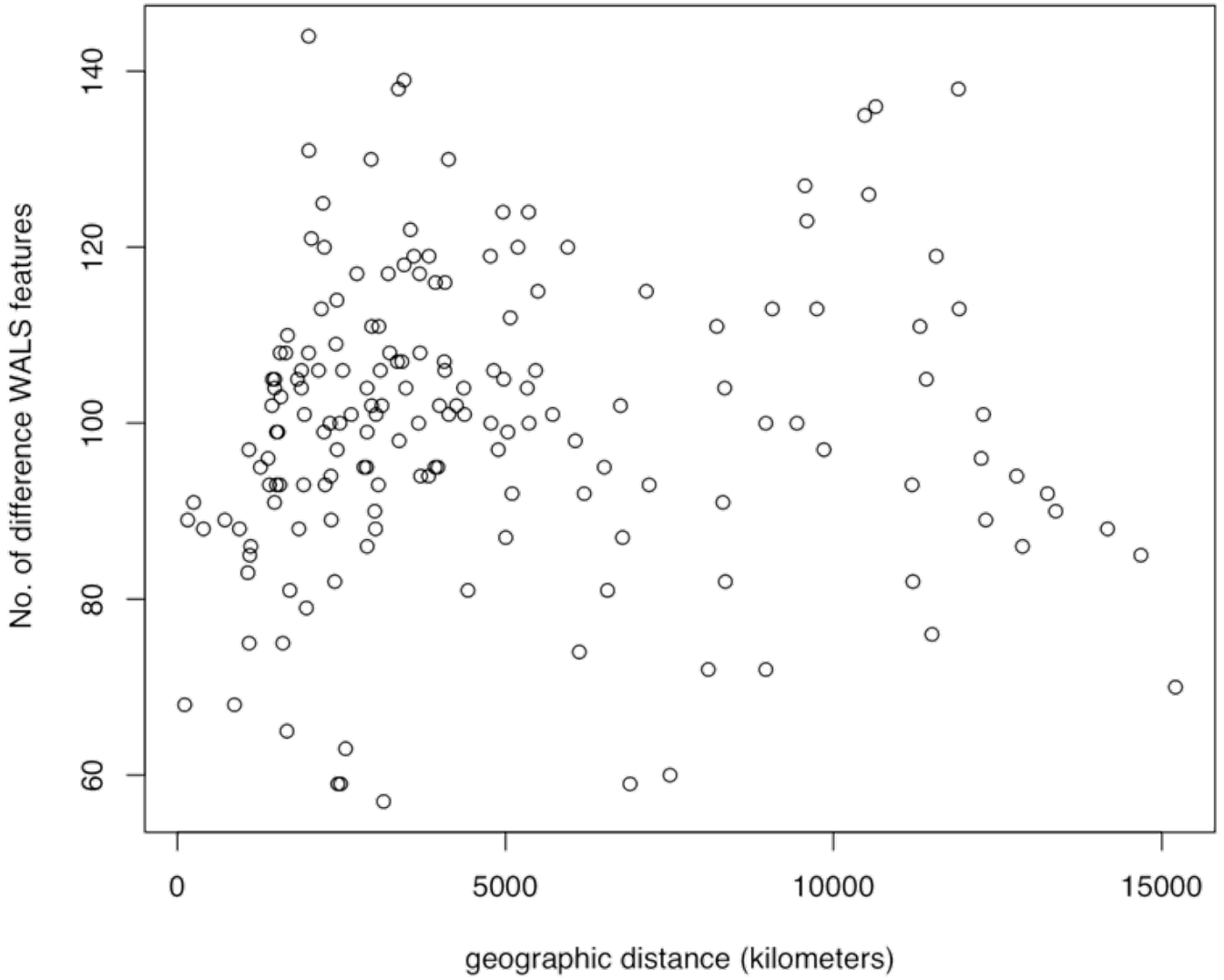
MDS of typological distances




Maybrat
Kobon
Tukang Besi
Amele
Tiwi
Lavukaleve
Alamblak
Mangarrayi
Gooniyandi
Kayardild
Ngiyambaa
Maruthinura


Maybrat
Alamblak
Kobon
Tukang Besi
Amele
Tiwi
Lavukaleve
Mangarrayi
Kayardild
Gooniyandi
Martuthinura
Ngiyambaa

Typology/geography correlation

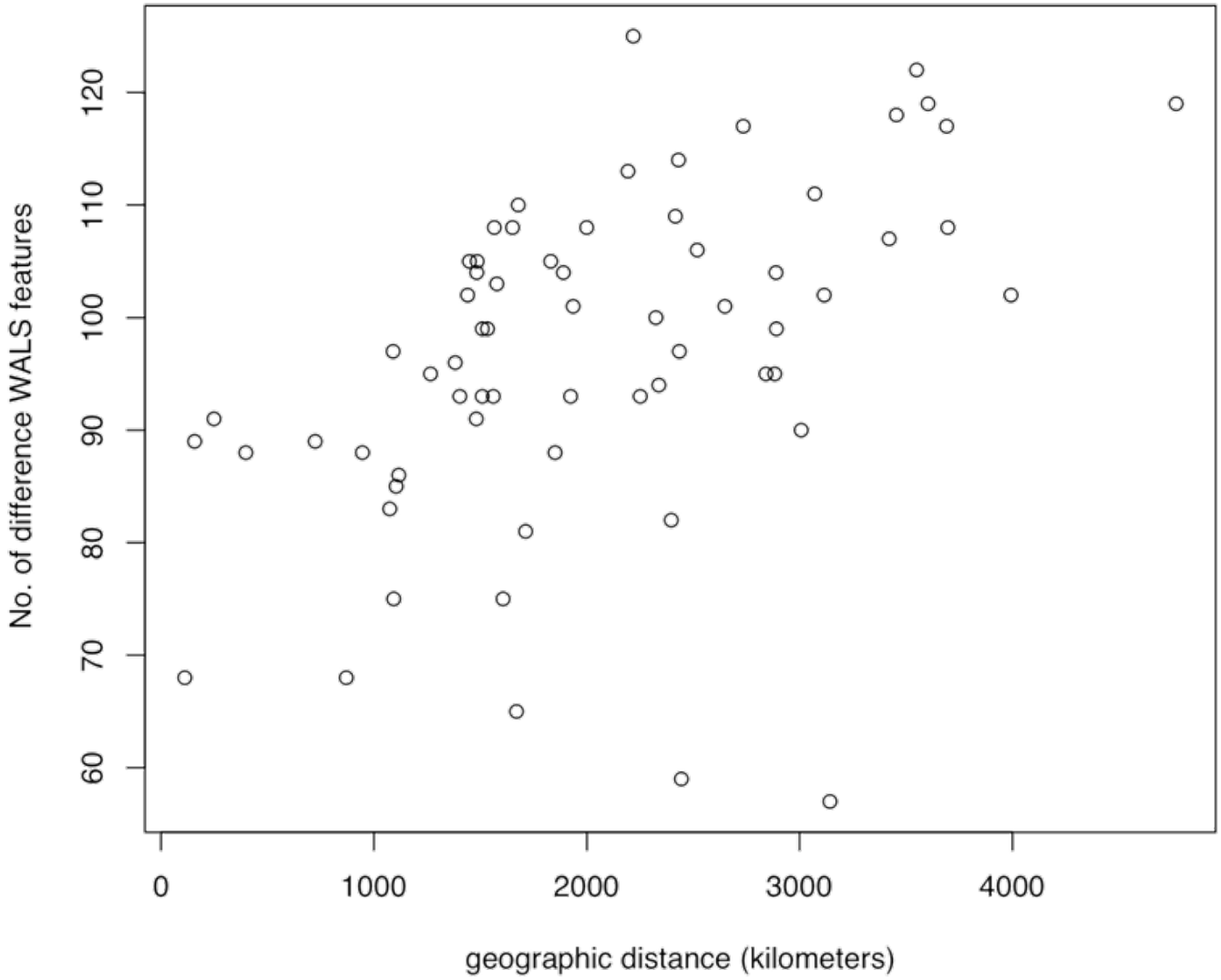


Mantel test
p = .349

When does correlation improve?

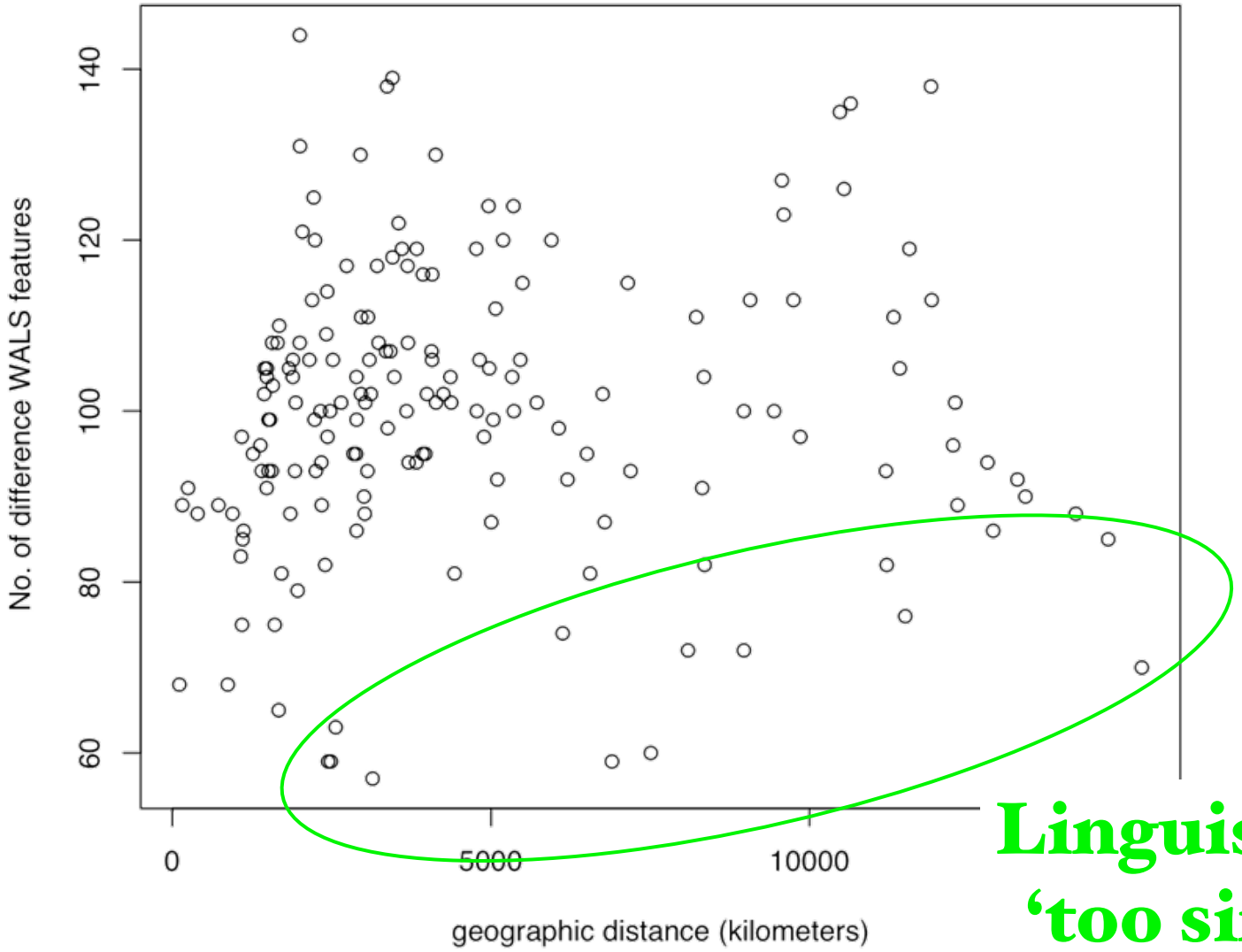
	Pearson's *r*
Nothing removed	.035
Rapanui	.186
Chamorro	.086
Indonesian	.076
Fijian	.073
Tagalog	.071
Maori	.062
Tukang Besi	.048

Correlation for selection



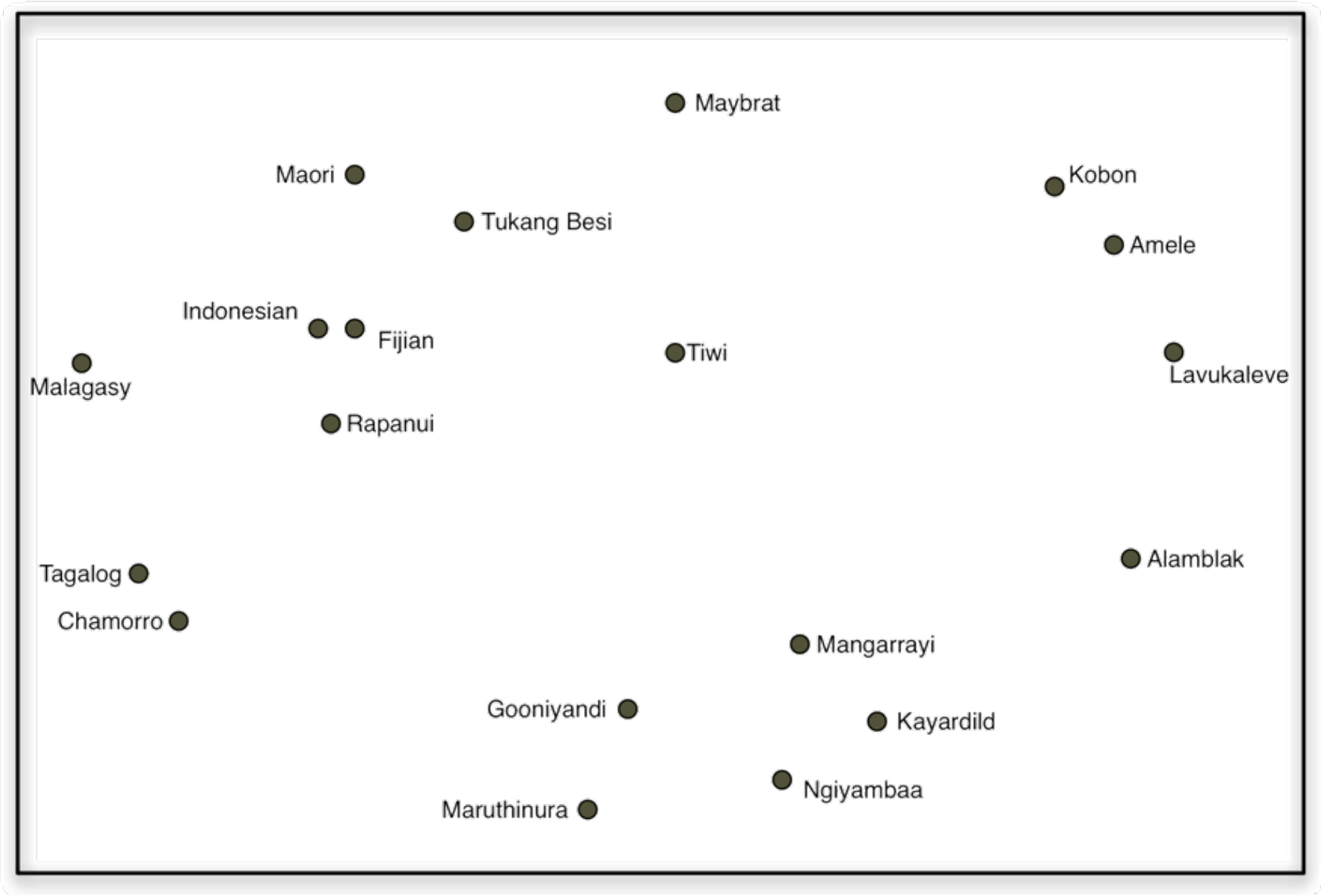
Mantel test p = .001

Investigation typology/geography relation

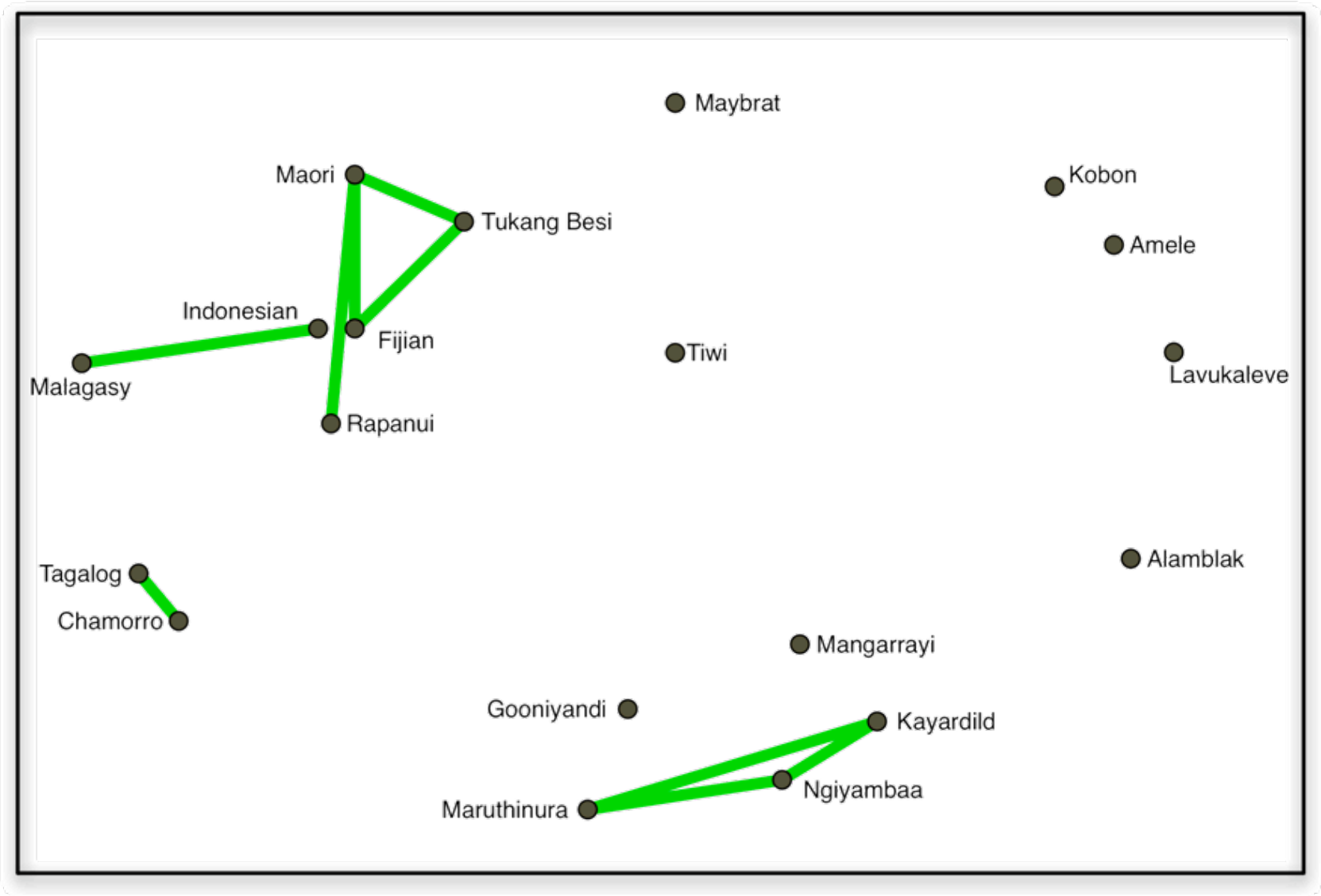


Linguistically 'too similar'

Linguistically ‘too similar’



Linguistically ‘too similar’



MAX-PLANCK-GESELLSCHAFT

Max Planck Institute
for Evolutionary Anthropology