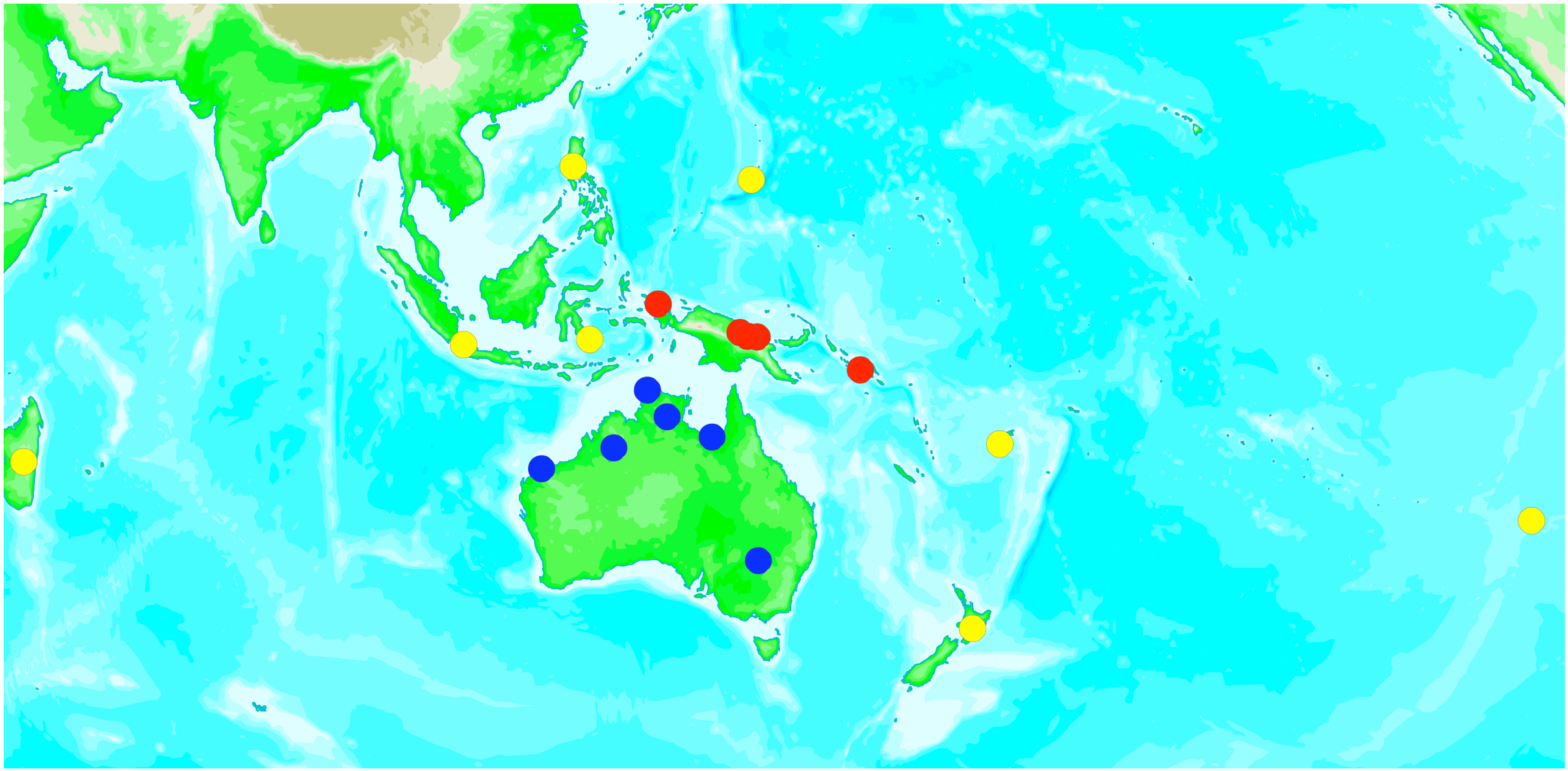
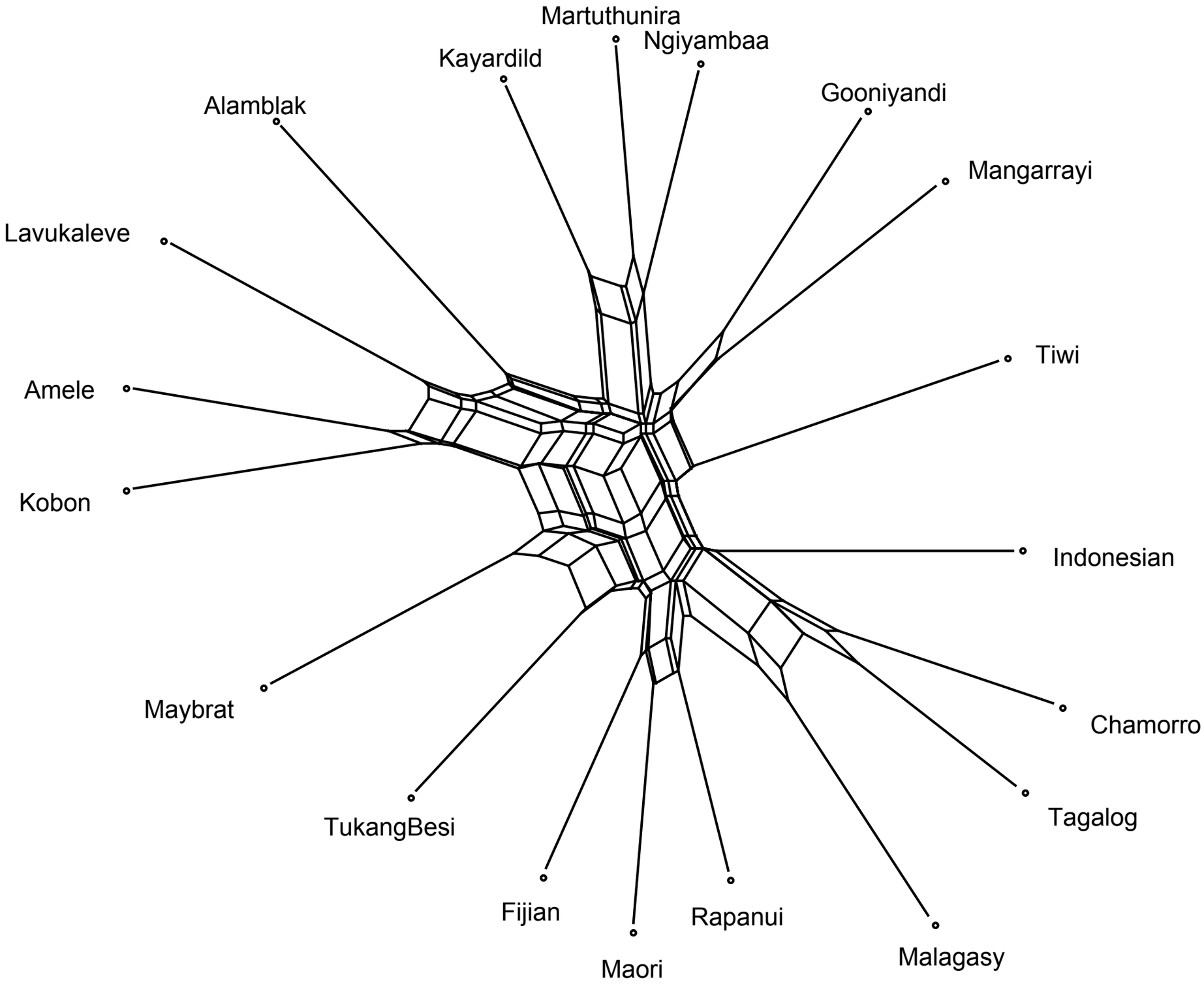
Geographic distance as a predictor of linguistic similarity

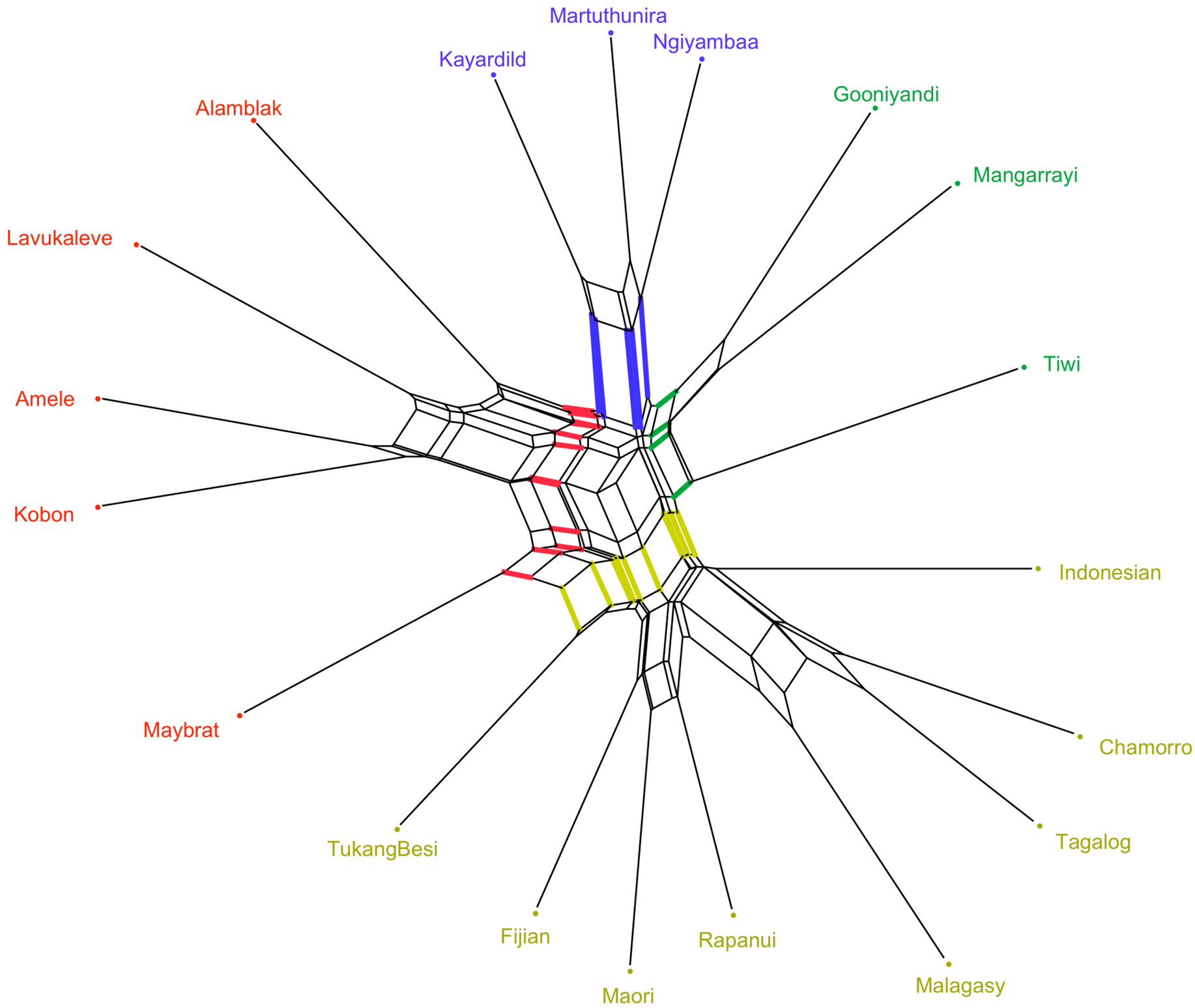
Michael Cysouw
MPI for Evolutionary Anthropology, Leipzig

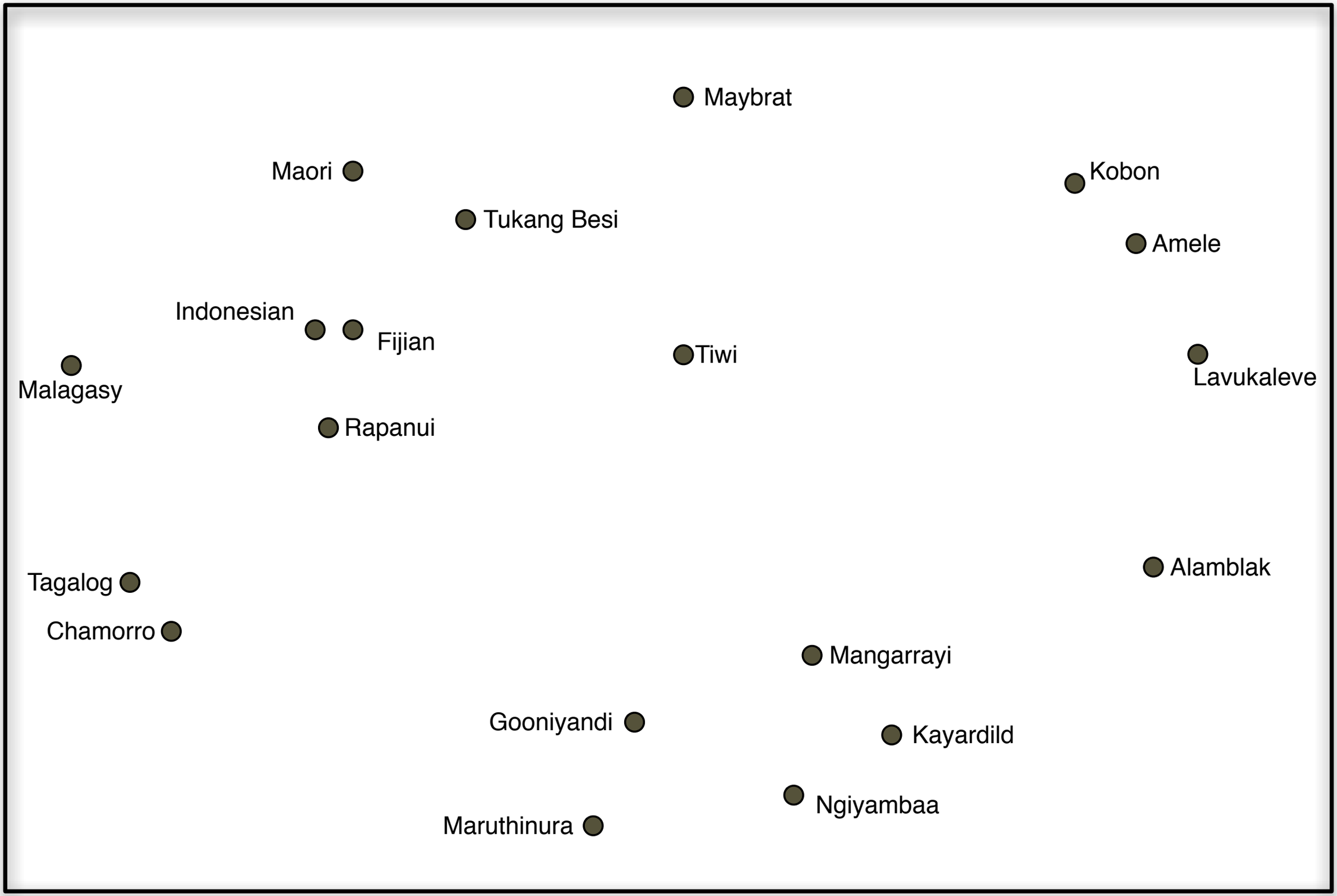
Geogr ance as a tor of li tic si arity

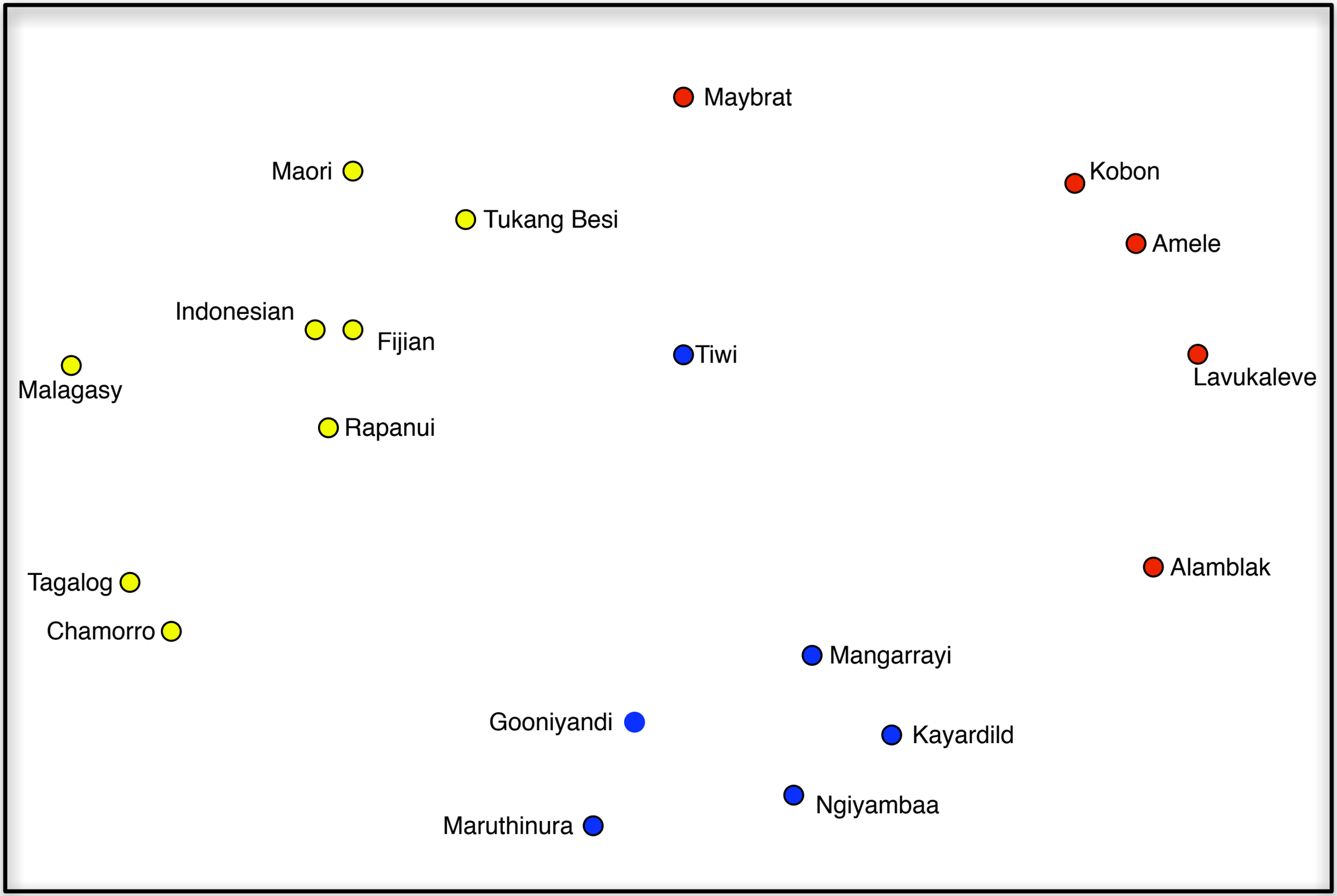
Michael Cysouw
MPI for Evolutionary Anthropology, Leipzig




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


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


Maybrat
Maori
Kobon
Tukang Besi
Amele
Indonesian
Fijian
Tiwi
Lavukaleve
Malagasy
Rapanui
Alamblak
Tagalog
Chamorro
Mangarrayi
Gooniyandi
Kayardild
Ngiyambaa
Maruthinura


Maybrat
Maori
Kobon
Tukang Besi
Amele
Indonesian
Fijian
Tiwi
Lavukaleve
Malagasy
Rapanui
Alamblak
Tagalog
Chamorro
Mangarrayi
Gooniyandi
Kayardild
Ngiyambaa
Maruthinura

Maybrat

Kobon

Tukang Besi

Amele

Tiwi

Lavukaleve

Alamblak

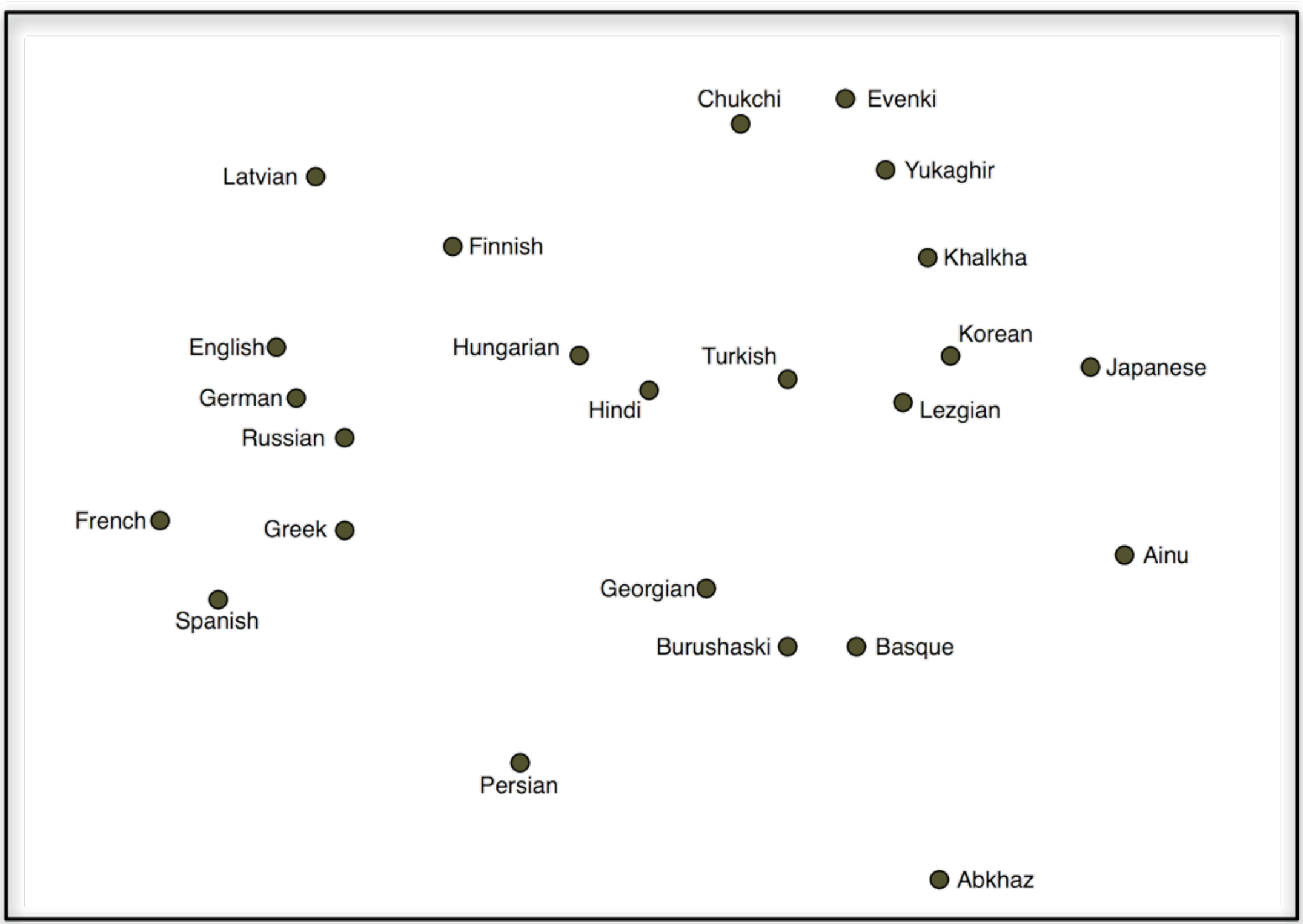
Mangarrayi

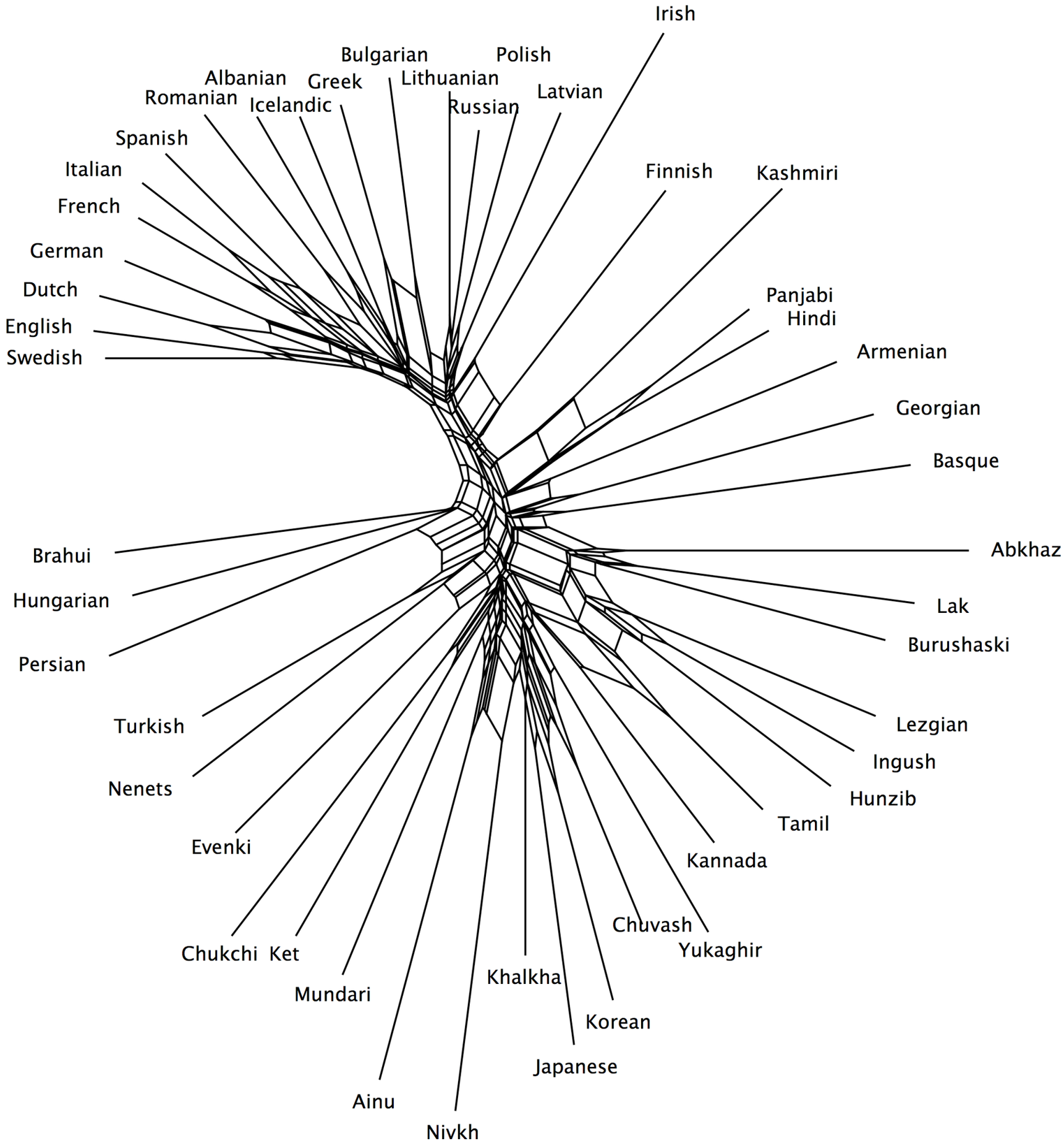
Gooniyandi

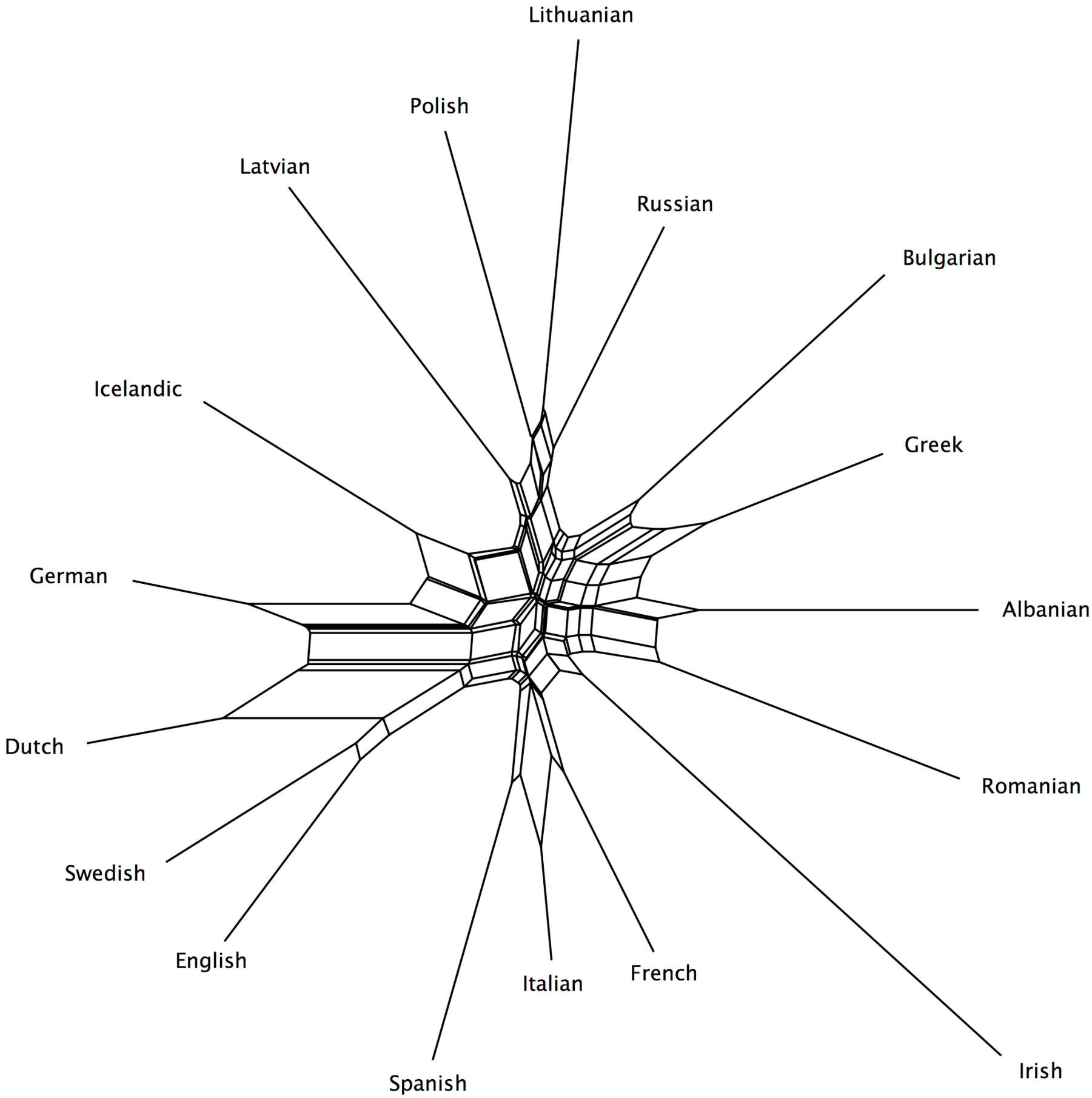
Kayardild

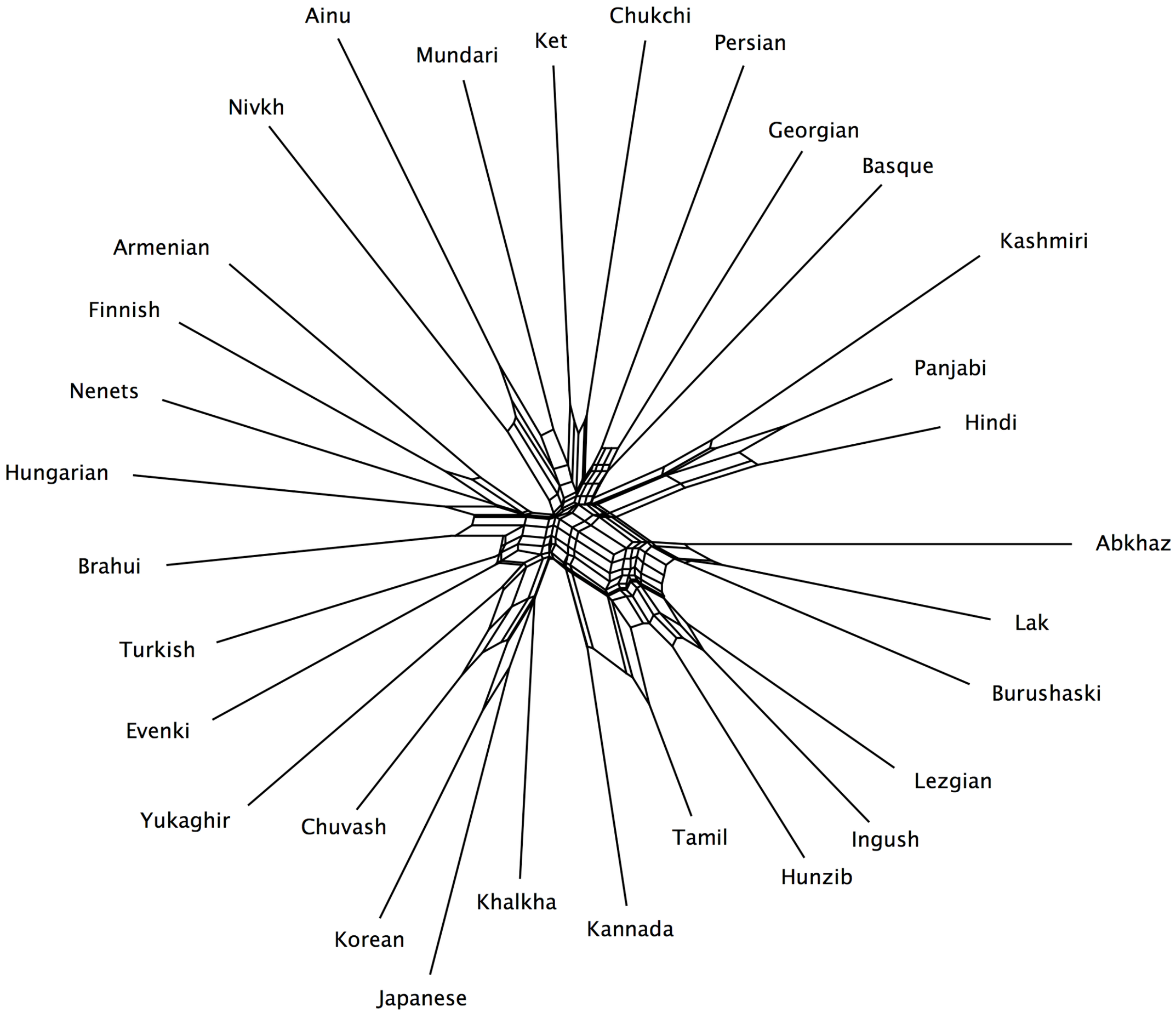
Ngiyambaa

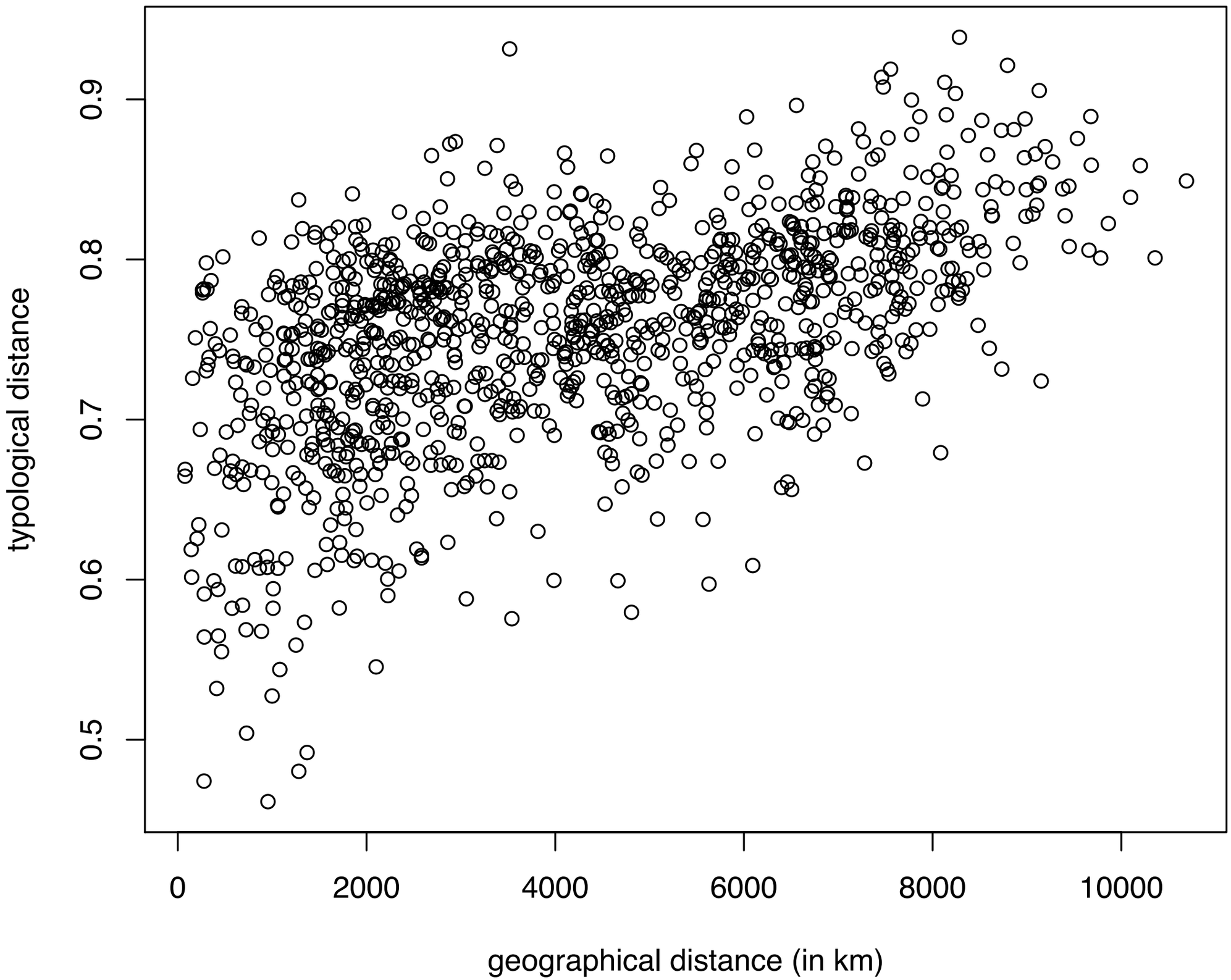
Maruthinura

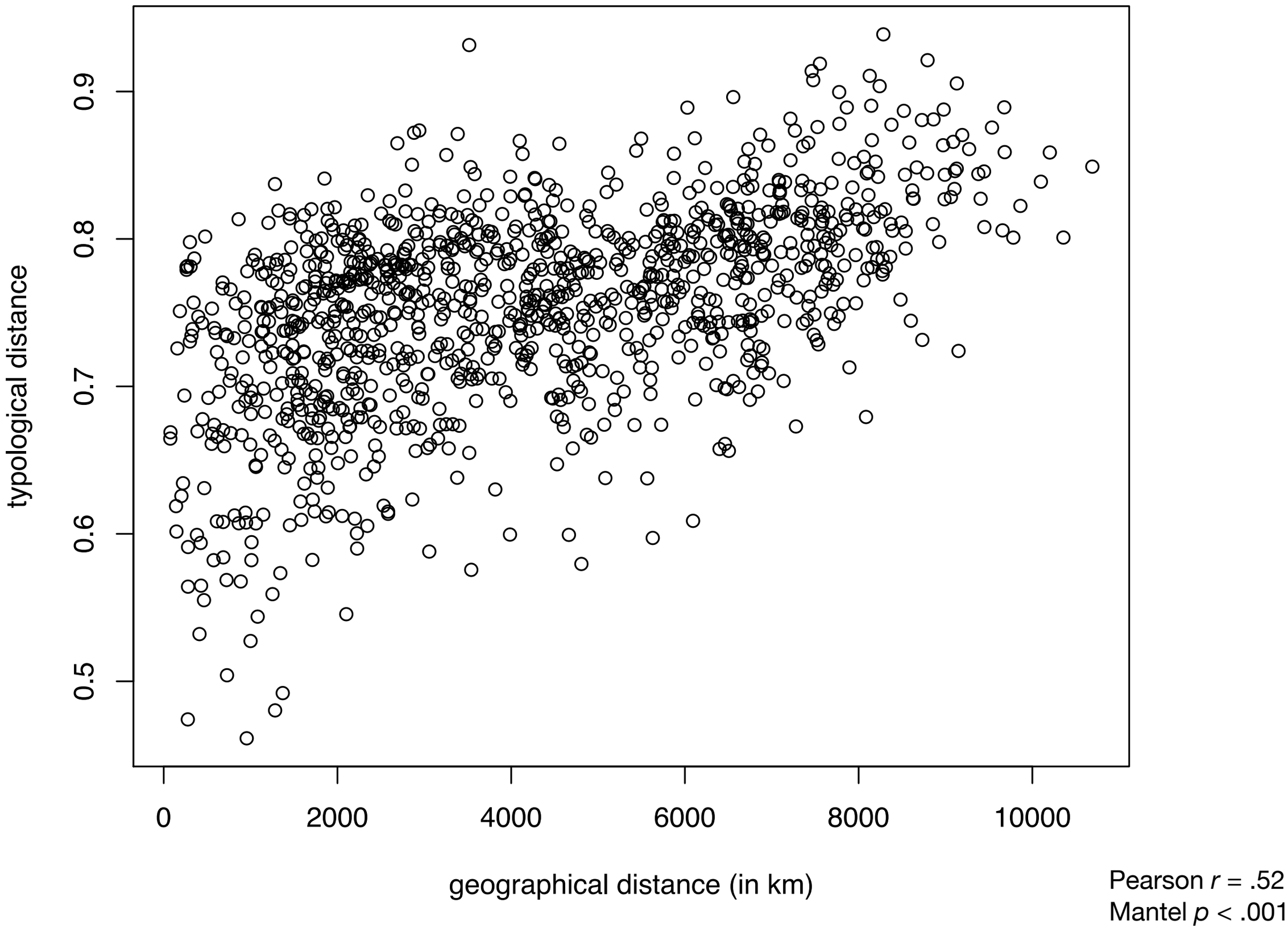

Chukchi
Evenki
Latvian
Yukaghir
Finnish
Khalkha
English
Hungarian
Turkish
Korean
Japanese
German
Hindi
Lezgian
Russian
French
Greek
Ainu
Georgian
Spanish
Burushaski
Basque
Persian
Abkhaz

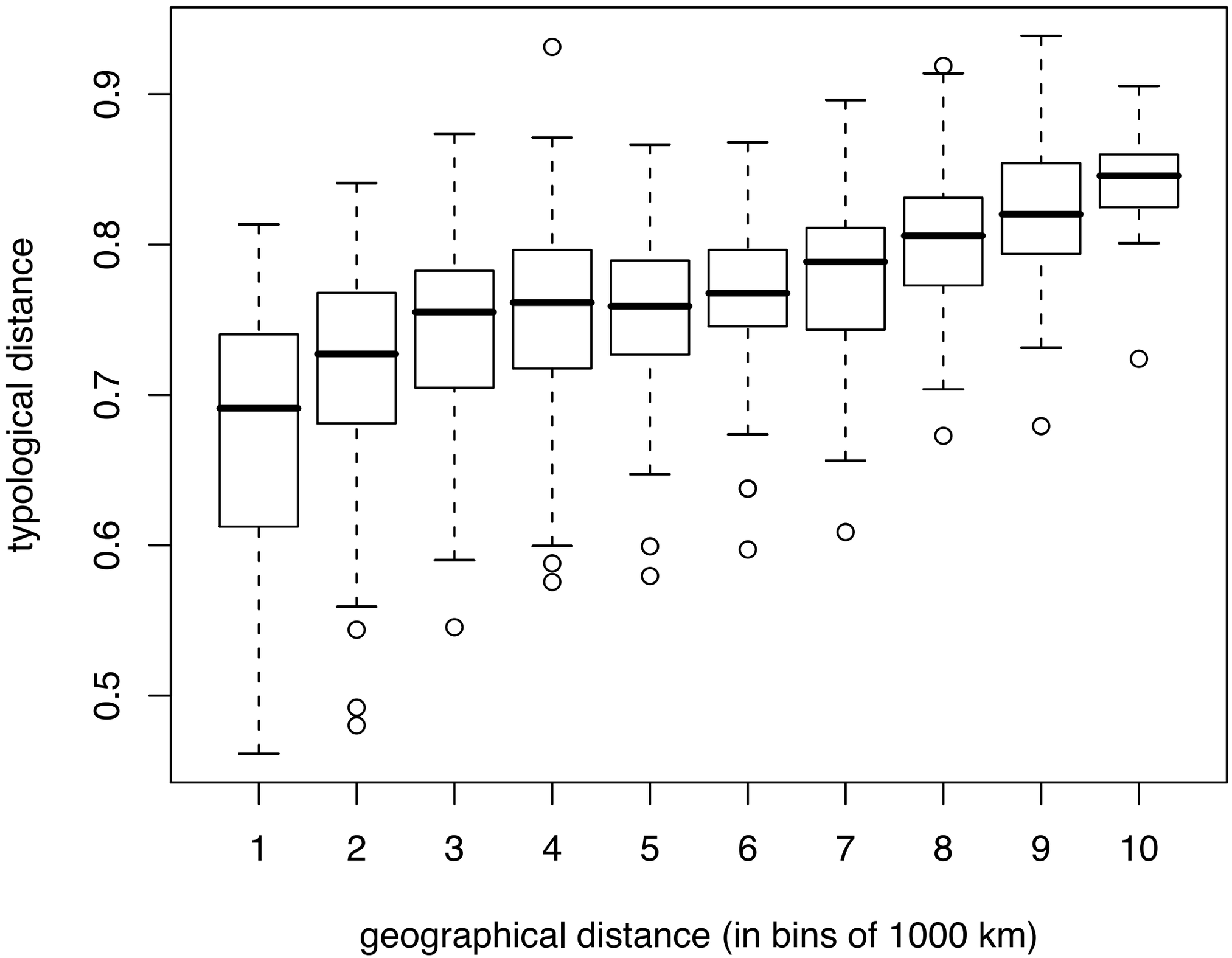

Irish
Bulgarian
Polish
Albanian
Greek
Lithuanian
Latvian
Romanian
Icelandic
Russian
Spanish
Italian
Finnish
Kashmiri
French
German
Dutch
Panjabi
Hindi
English
Armenian
Swedish
Georgian
Basque
Brahui
Abkhaz
Hungarian
Lak
Burushaski
Persian
Lezgian
Turkish
Ingush
Hunzib
Nenets
Tamil
Evenki
Kannada
Chuvash
Chukchi
Ket
Yukaghir
Mundari
Khalkha
Korean
Japanese
Ainu
Nivkh

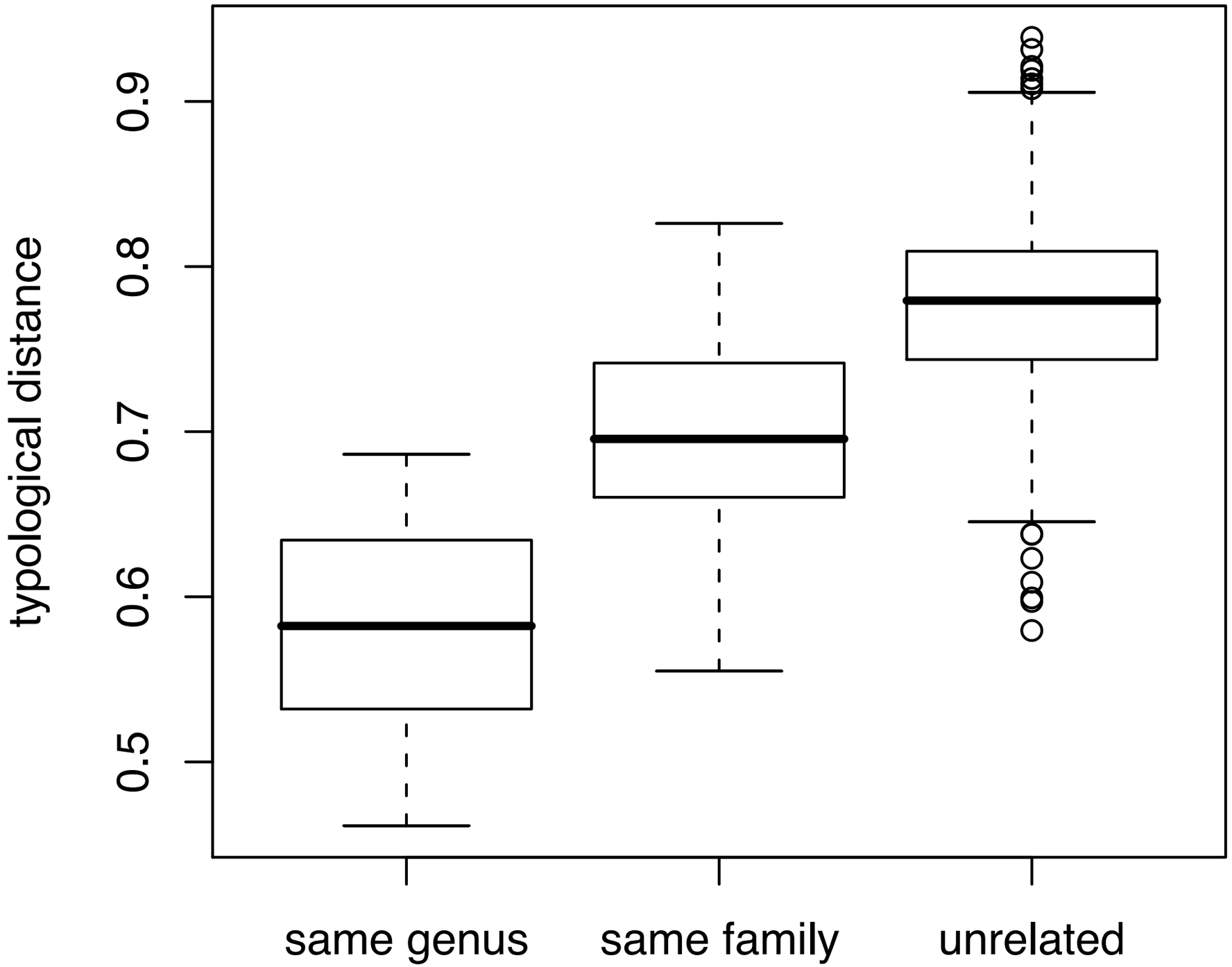

Lithuanian
Polish
Latvian
Russian
Bulgarian
Icelandic
Greek
German
Albanian
Dutch
Romanian
Swedish
English
Italian
French
Spanish
Irish

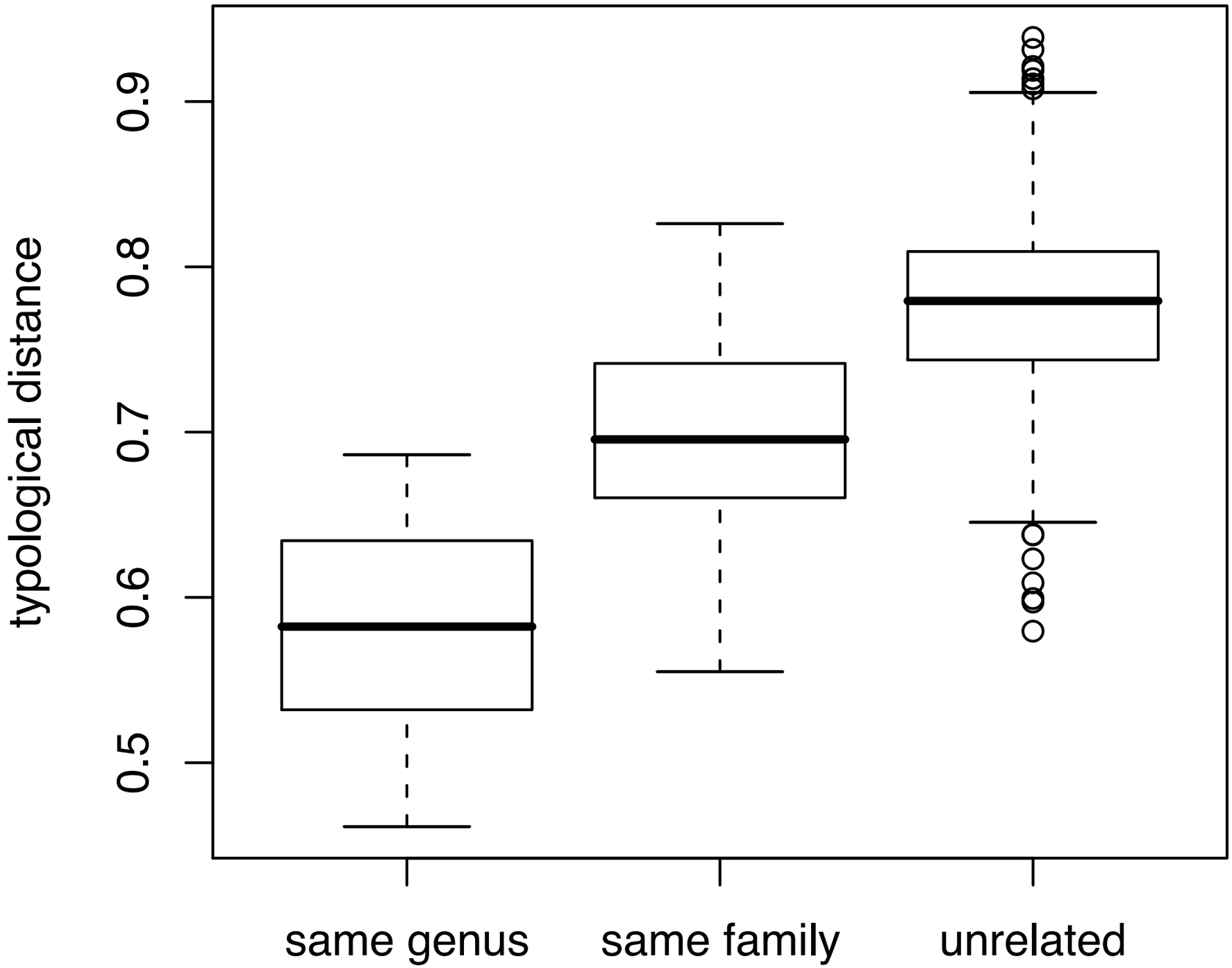

Ainu
Chukchi
Ket
Mundari
Persian
Nivkh
Georgian
Basque
Armenian
Kashmiri
Finnish
Panjabi
Nenets
Hindi
Hungarian
Abkhaz
Brahui
Lak
Turkish
Burushaski
Evenki
Lezgian
Yukaghir
Chuvash
Tamil
Ingush
Hunzib
Khalkha
Korean
Kannada
Japanese


typological distance
0.5
0.6
0.7
0.8
0.9
0
2000
4000
6000
8000
10000
geographical distance (in km)


typological distance
0.5
0.6
0.7
0.8
0.9
0
2000
4000
6000
8000
10000
geographical distance (in km)
Pearson r = .52
Mantel p < .001


typological distance
0.5
0.6
0.7
0.8
0.9
1
2
3
4
5
6
7
8
9
10
geographical distance (in bins of 1000 km)


typological distance
0.5
0.6
0.7
0.8
0.9
same genus
same family
unrelated



Pearson r = .61
Mantel p < .001

Mantel Test

Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography	0.52	< .001

Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography	0.52	< .001
Typology ~ Genealogy	0.61	< .001

Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography	0.52	< .001
Typology ~ Genealogy	0.61	< .001
Geography ~ Genealogy	0.33	< .001

Partial Mantel Test

Partial Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography + Genealogy	0.42	< .001

Partial Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography + Genealogy	0.42	< .001
Typology ~ Genealogy + Geography	0.54	< .001

Multivariate Matrix Regression

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

Multivariate Matrix Regression

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

Multivariate Matrix Regression

	Sums of Sqs	Mean Sqs	F Model	R^2
family : genus	13.65	0.02	-3.29	0.77
latitude : longitude	0.28	0.28	-39.01	0.02
Residuals	3.82	-0.01		0.22

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

How much geography is there left after factoring out genealogy ?

How much geography is there left after factoring out genealogy ?

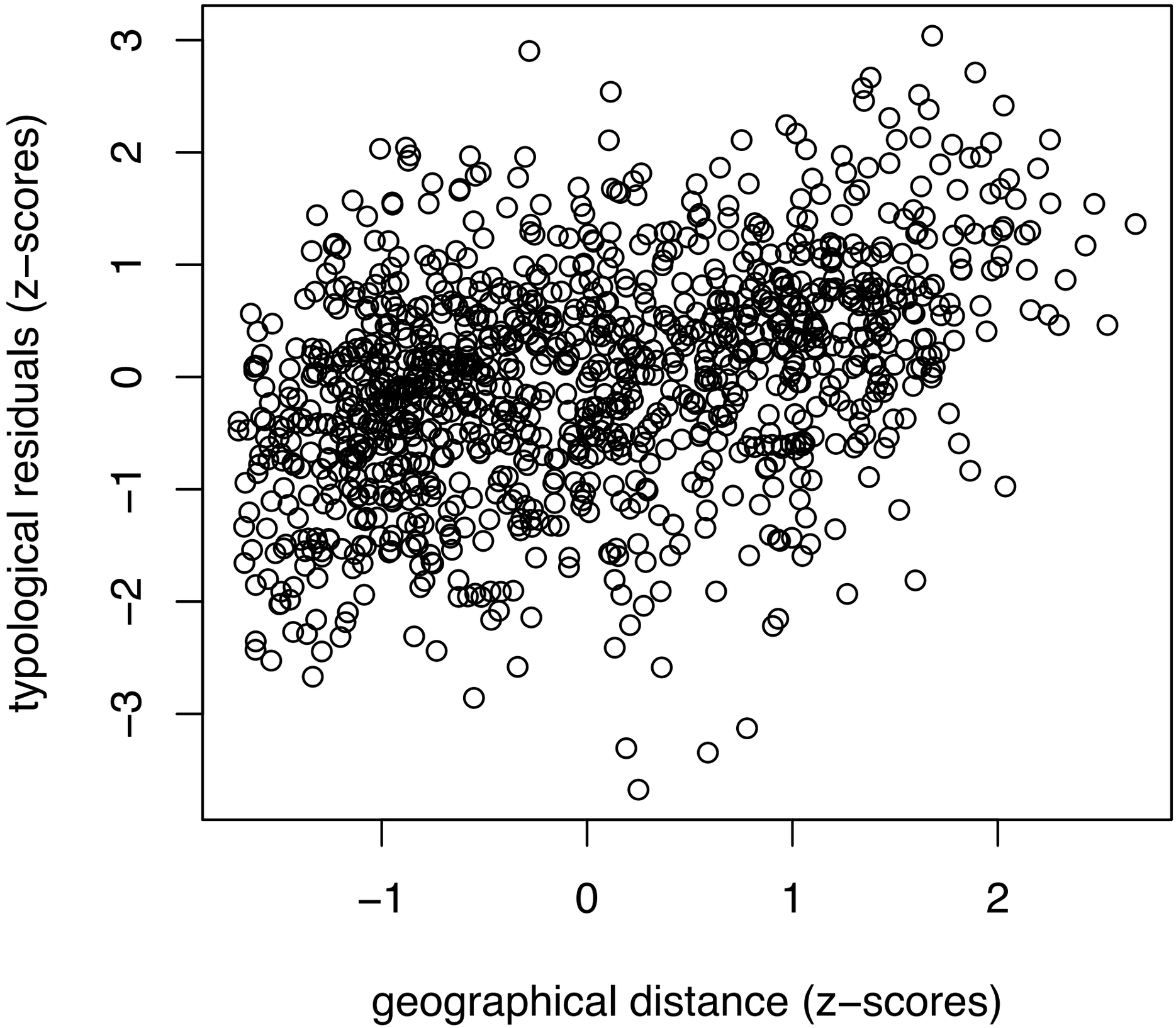
- Regression Typology ~ Genealogy

How much geography is there left after factoring out genealogy ?

- Regression Typology ~ Genealogy
- Negative residuals after regression show ‘more similarity than expected by genealogy’

How much geography is there left after factoring out genealogy ?

- Regression Typology ~ Genealogy
- Negative residuals after regression show ‘more similarity than expected by genealogy’
- This surpluss similarity is probably contact


typological residuals (z−scores)
3
2
1
0
−1
−2
−3
−1
0
1
2
geographical distance (z−scores)

Greek	Bulgarian
German	Dutch
Italian	French
Greek	Albanian
Korean	Japanese
German	French
Russian	Lithuanian
Latvian	Finnish
Swedish	English
French	Dutch
Russian	Finnish
Lezgian	Ingush
Romanian	Albanian
Tamil	Ingush
Tamil	Burushaski

German	Dutch
German	French
French	Dutch

Italian	French

Swedish	English

Greek	Bulgarian
Greek	Albanian
Romanian	Albanian

Russian	Lithuanian
Latvian	Finnish
Russian	Finnish

Korean	Japanese

Lezgian	Ingush

Tamil	Ingush

Tamil	Burushaski

Summary

Summary

- Overall typological similarity is both correlated with
  - genealogical closeness
  - geographical closeness

Summary

- Overall typological similarity is both correlated with
  - genealogical closeness
  - geographical closeness
- Still, genealogy seems to be most important

Summary

- Overall typological similarity is both correlated with
  - genealogical closeness
  - geographical closeness
- Still, genealogy seems to be most important
- Leftover variation points towards contact