Measuring Genealogical Stability of Typological Data

Michael Cysouw
Max Planck Institute for Evolutionary Anthropology, Leipzig

or:

A Plea for Genealogically *Biased* Sampling

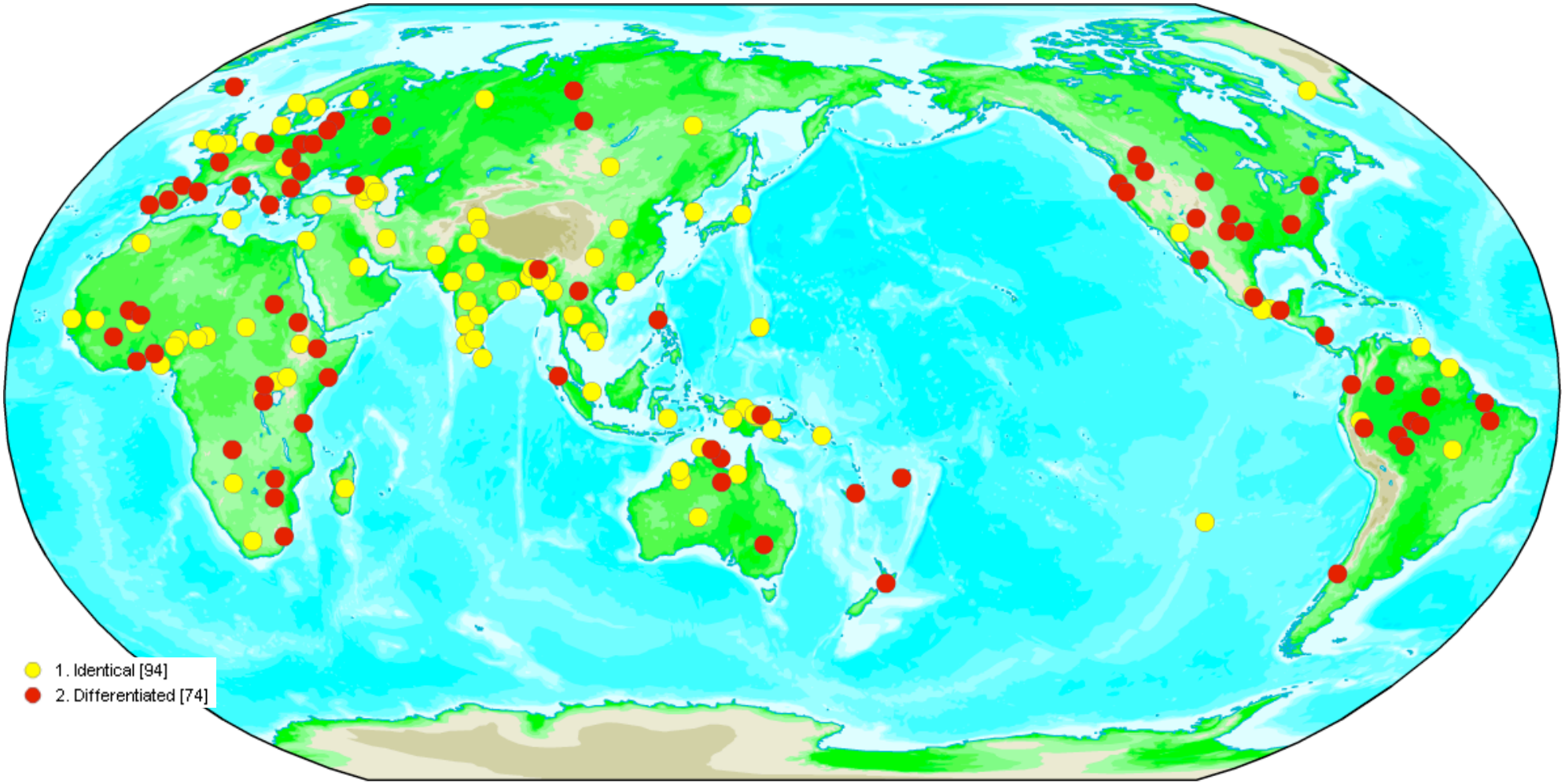
Michael Cysouw

Max Planck Institute for Evolutionary Anthropology, Leipzig

Questions for Typology

- How to explain world-wide typological distributions ?
  - a-historical, intra-linguistic, *universal* influences
  - historical, extra-linguistic, *contingent* influences
- How to deal with *large-areal consistency* in genealogically balanced samples ?

Intensifiers & Reflexive Pronouns



König, Ekkehard & Peter Siemund (with Stephan Töpper). 2005. Intensifiers and Reflexive Pronouns. In: Haspelmath *et al.* (eds.) *The World Atlas of Linguistic Structures.* Oxford University Press.

Reactions to *Large Areal Consistencies*

- Matthew Dryer (starting from 1989):

  Problem for universals !

- Johanna Nichols (starting from 1992):

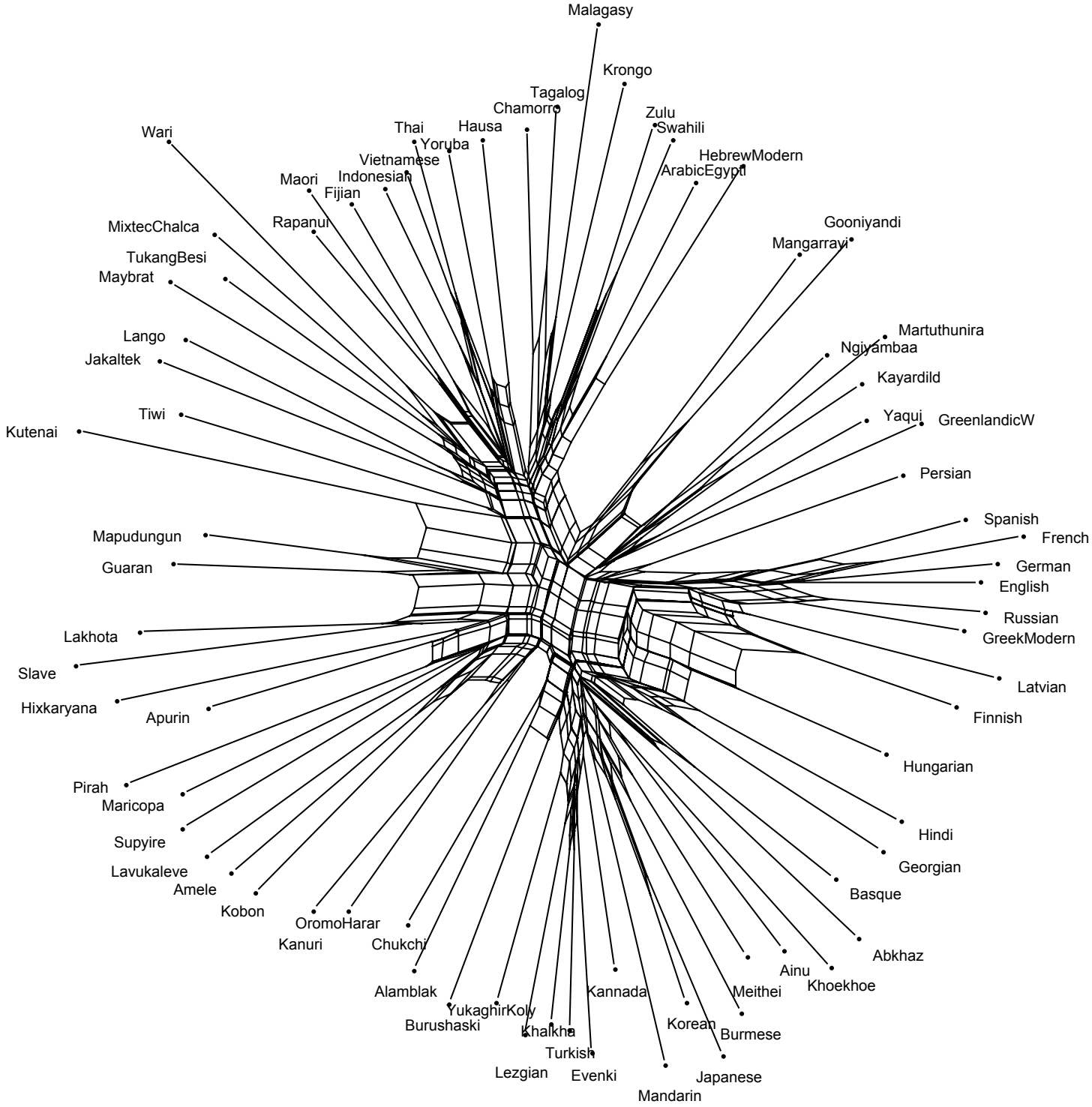
  Great for investigation of history !

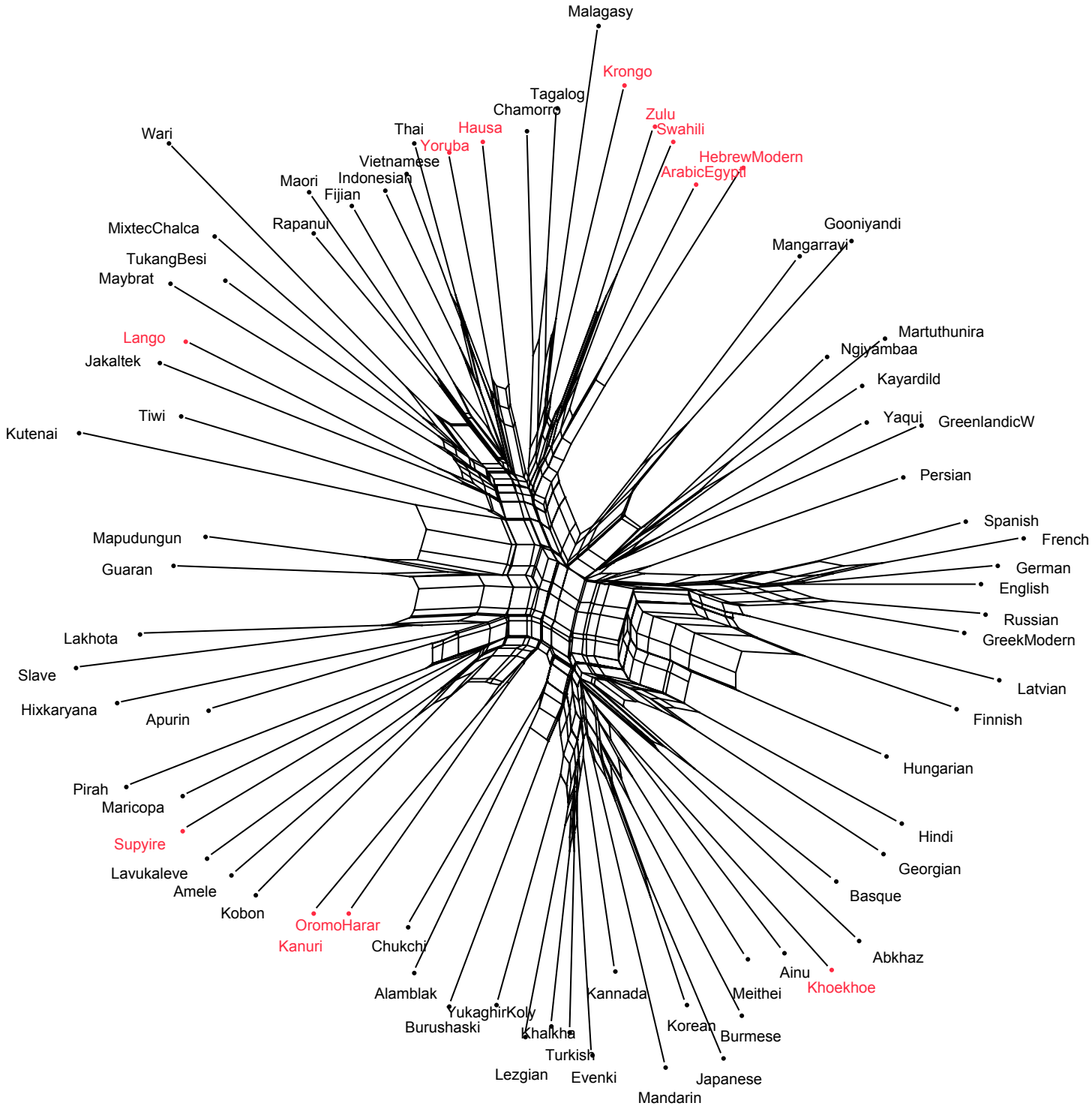
- Michael Cysouw:

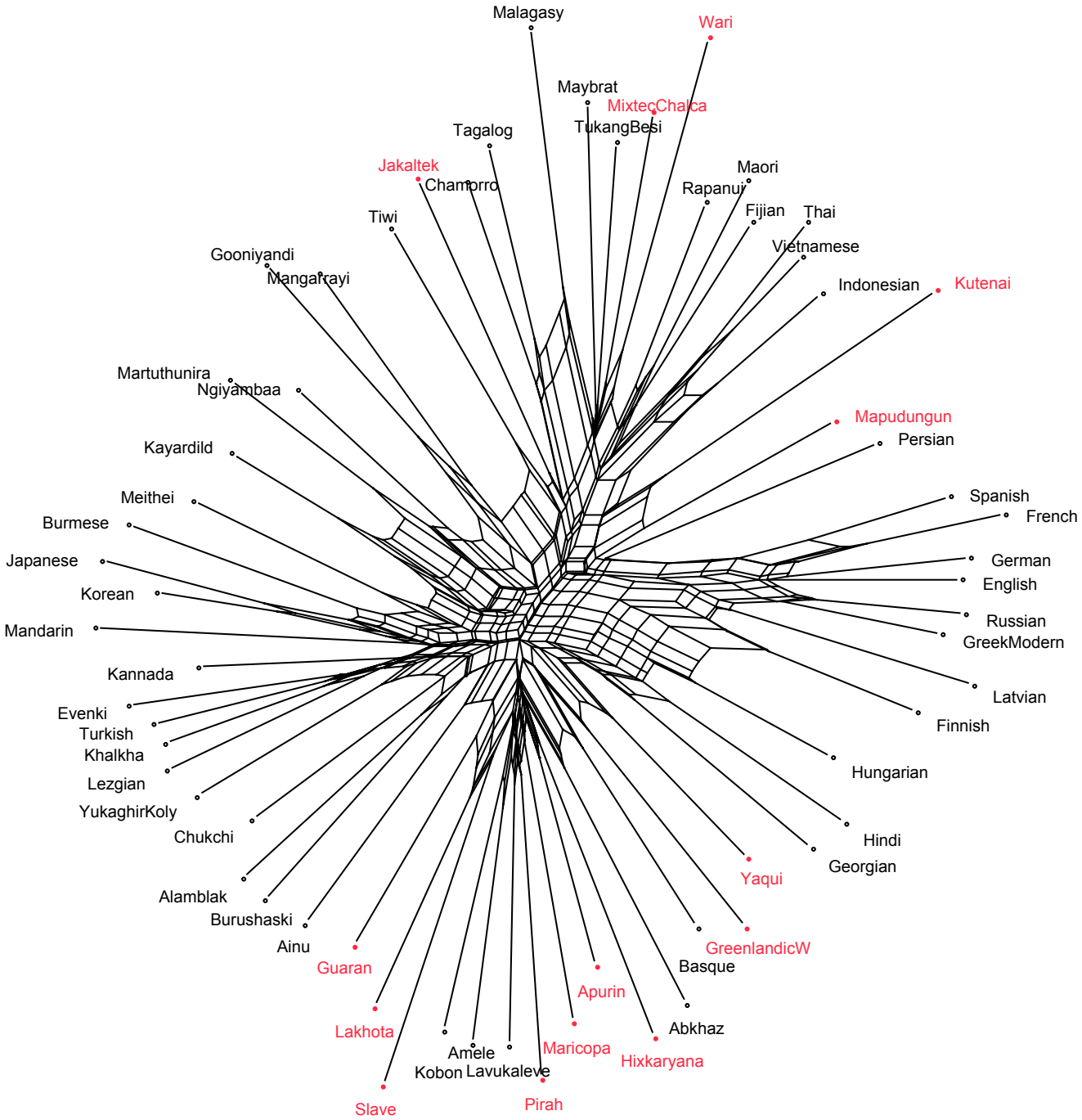
  there is indeed history to be uncovered,
  but it is far from clear with which methods

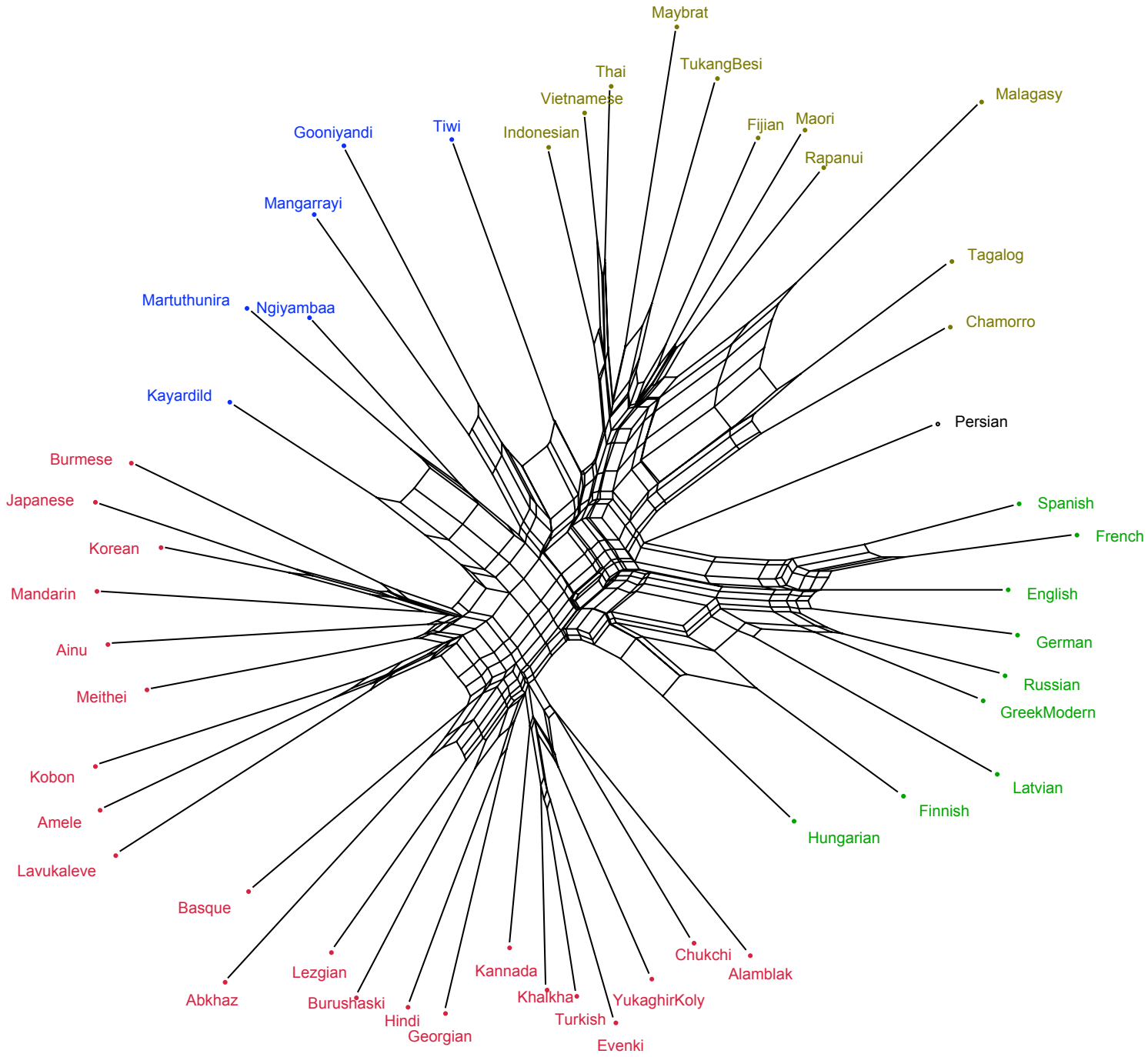
How tree-like is typological data?

- Using the *World Atlas of Language Structures*
- Investigating overall typologically similarity of the world's languages
- Calculate overall similarity between all pairs of languages
- Are there any clusters of similar languages ?

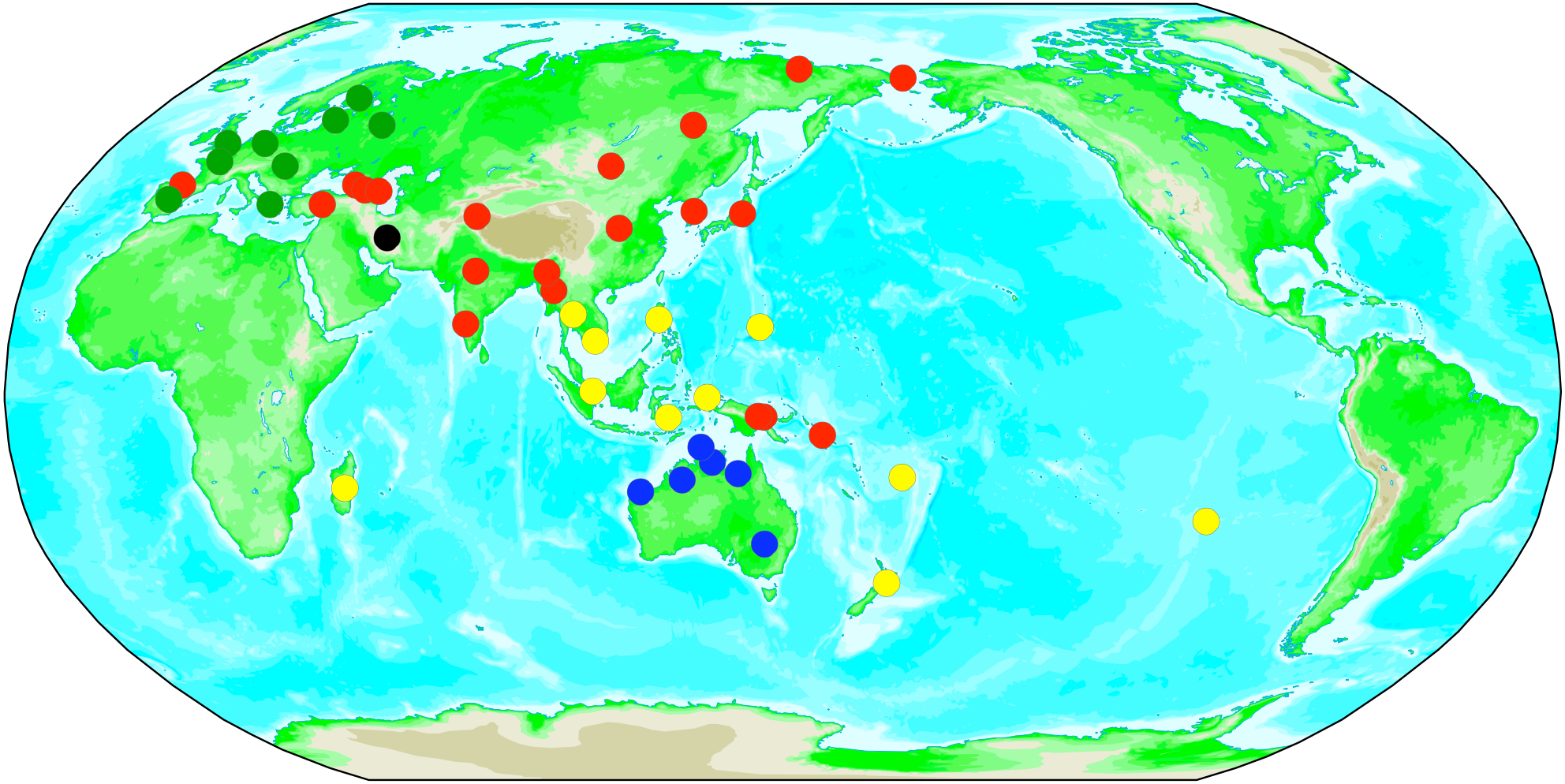

Malagasy
Krongo
Tagalog
Chamorro
Zulu
Swahili
Thai
Hausa
Yoruba
HebrewModern
ArabicEgypt
Vietnamese
Indonesian
Maori
Fijian
Wari
Rapanui
MixtecChalca
TukangBesi
Maybrat
Gooniyandi
Mangarrayi
Martuthunira
Ngiyambaa
Kayardild
Yaqui
GreenlandicW
Persian
Lango
Jakaltek
Tiwi
Kutenai
Spanish
French
German
English
Russian
GreekModern
Latvian
Finnish
Hungarian
Hindi
Georgian
Basque
Abkhaz
Ainu
Khoekhoe
Meithei
Korean
Burmese
Japanese
Mandarin
Evenki
Lezgian
Turkish
Khalkha
Kannada
Alamblak
YukaghirKoly
Burushaski
Chukchi
OromoHarar
Kanuri
Kobon
Amele
Lavukaleve
Supyire
Maricopa
Pirah
Apurin
Hixkaryana
Slave
Lakhota
Guaran
Mapudungun


Malagasy
Krongo
Tagalog
Chamorro
Zulu
Swahili
Thai
Hausa
Yoruba
HebrewModern
ArabicEgypt
Vietnamese
Indonesian
Maori
Fijian
Wari
Rapanui
MixtecChalca
TukangBesi
Maybrat
Gooniyandi
Mangarrayi
Martuthunira
Ngiyambaa
Kayardild
Yaqui
GreenlandicW
Persian
Lango
Jakaltek
Tiwi
Kutenai
Spanish
French
German
English
Russian
GreekModern
Latvian
Finnish
Hungarian
Hindi
Georgian
Basque
Abkhaz
Ainu
Khoekhoe
Meithei
Burmese
Japanese
Korean
Kannada
Alamblak
YukaghirKoly
Burushaski
Khalkha
Turkish
Lezgian
Evenki
Mandarin
Chukchi
OromoHarar
Kanuri
Kobon
Amele
Lavukaleve
Supyire
Maricopa
Pirah
Apurin
Hixkaryana
Slave
Lakhota
Guaran
Mapudungun

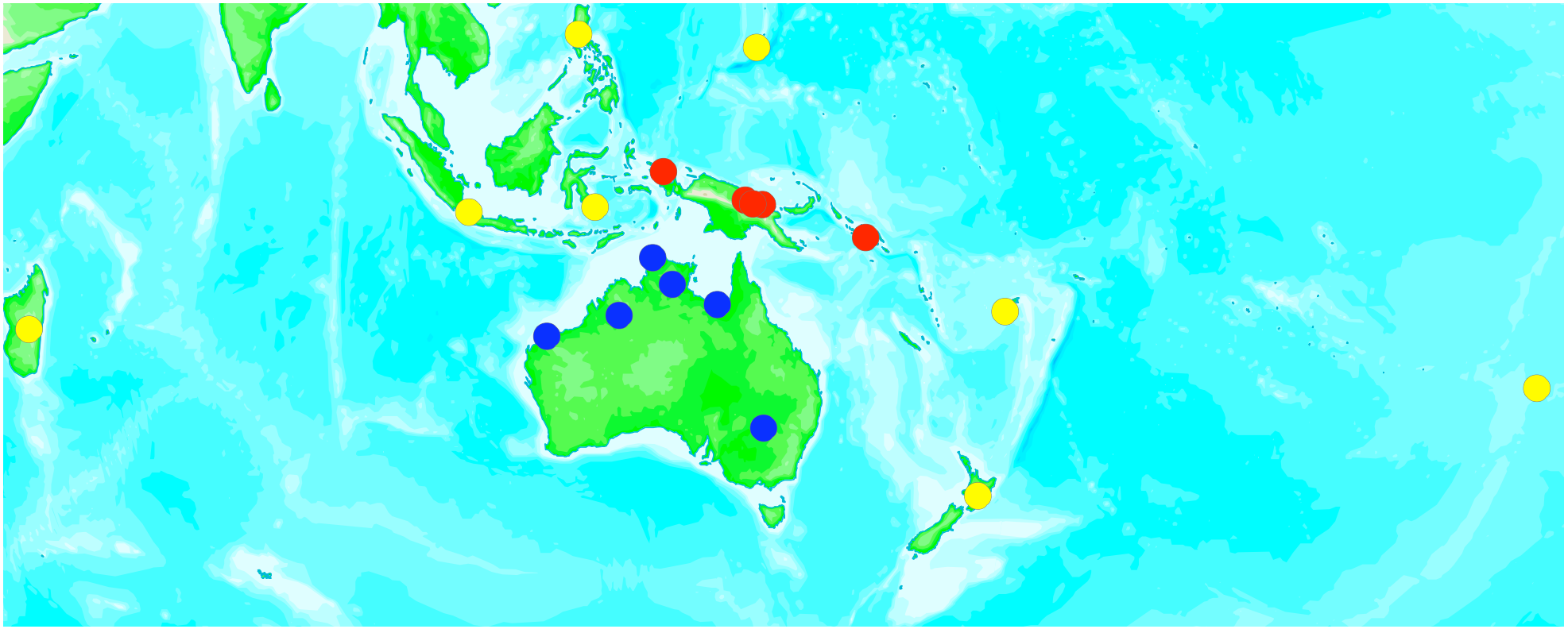

Malagasy
Wari
Maybrat
MixtecChalca
TukangBesi
Tagalog
Jakaltek
Chamorro
Maori
Rapanui
Fijian
Thai
Tiwi
Vietnamese
Gooniyandi
Mangarrayi
Indonesian
Kutenai
Martuthunira
Ngiyambaa
Mapudungun
Persian
Kayardild
Spanish
French
Meithei
Burmese
German
Japanese
English
Korean
Russian
Mandarin
GreekModern
Kannada
Latvian
Evenki
Finnish
Turkish
Khalkha
Hungarian
Lezgian
YukaghirKoly
Hindi
Chukchi
Georgian
Yaqui
Alamblak
Burushaski
GreenlandicW
Ainu
Basque
Guaran
Apurin
Lakhota
Abkhaz
Amele
Maricopa
Kobon
Lavukaleve
Hixkaryana
Slave
Pirah


Maybrat
TukangBesi
Thai
Vietnamese
Malagasy
Gooniyandi
Tiwi
Indonesian
Fijian
Maori
Rapanui
Mangarrayi
Tagalog
Martuthunira
Ngiyambaa
Chamorro
Kayardild
Persian
Burmese
Japanese
Spanish
French
Korean
Mandarin
English
Ainu
German
Russian
Meithei
GreekModern
Kobon
Latvian
Amele
Finnish
Hungarian
Lavukaleve
Basque
Chukchi
Lezgian
Kannada
Alamblak
Abkhaz
Khalkha
YukaghirKoly
Burushaski
Turkish
Hindi
Georgian
Evenki

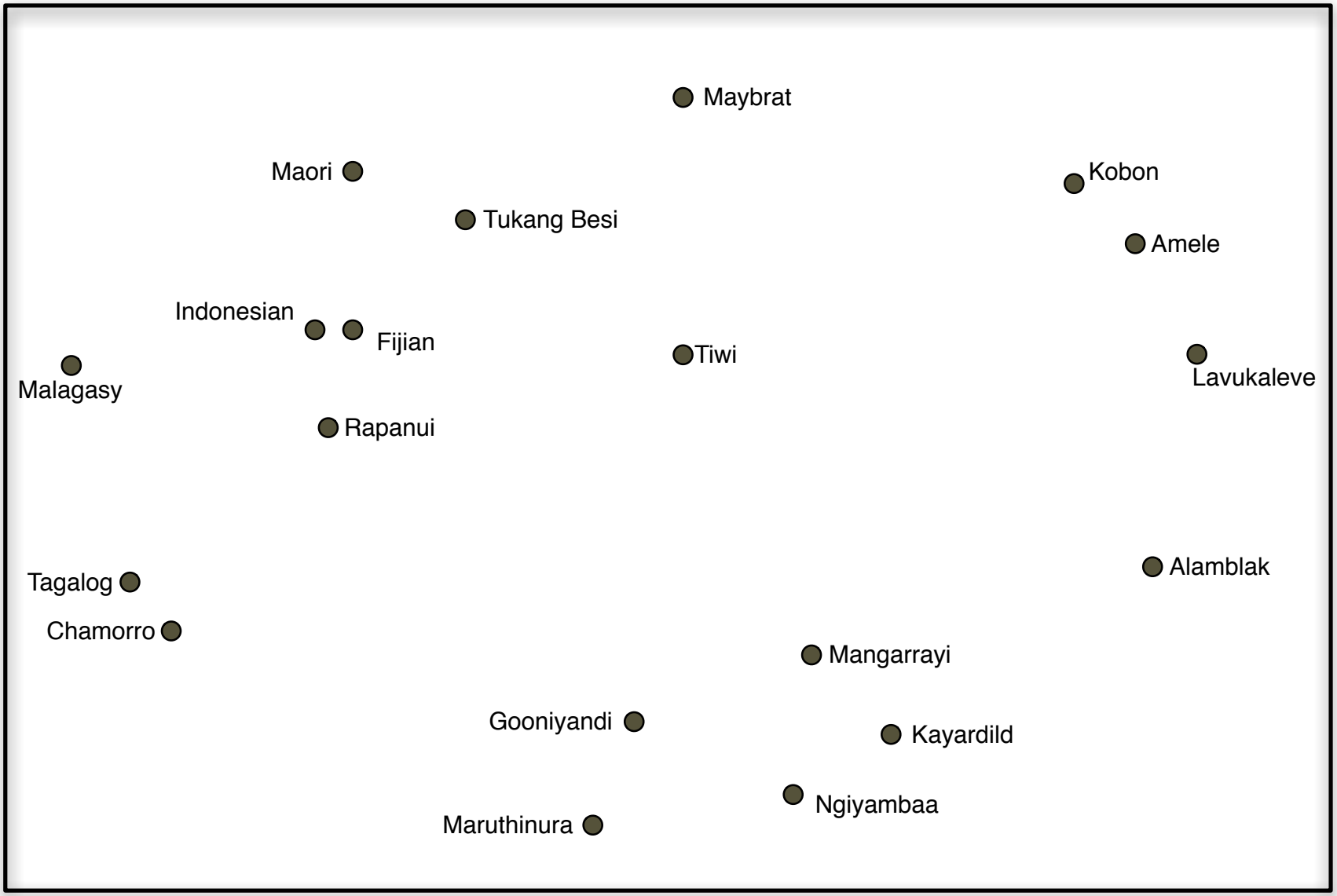
Overall Similarities show aspects of Geography *and* Genealogy



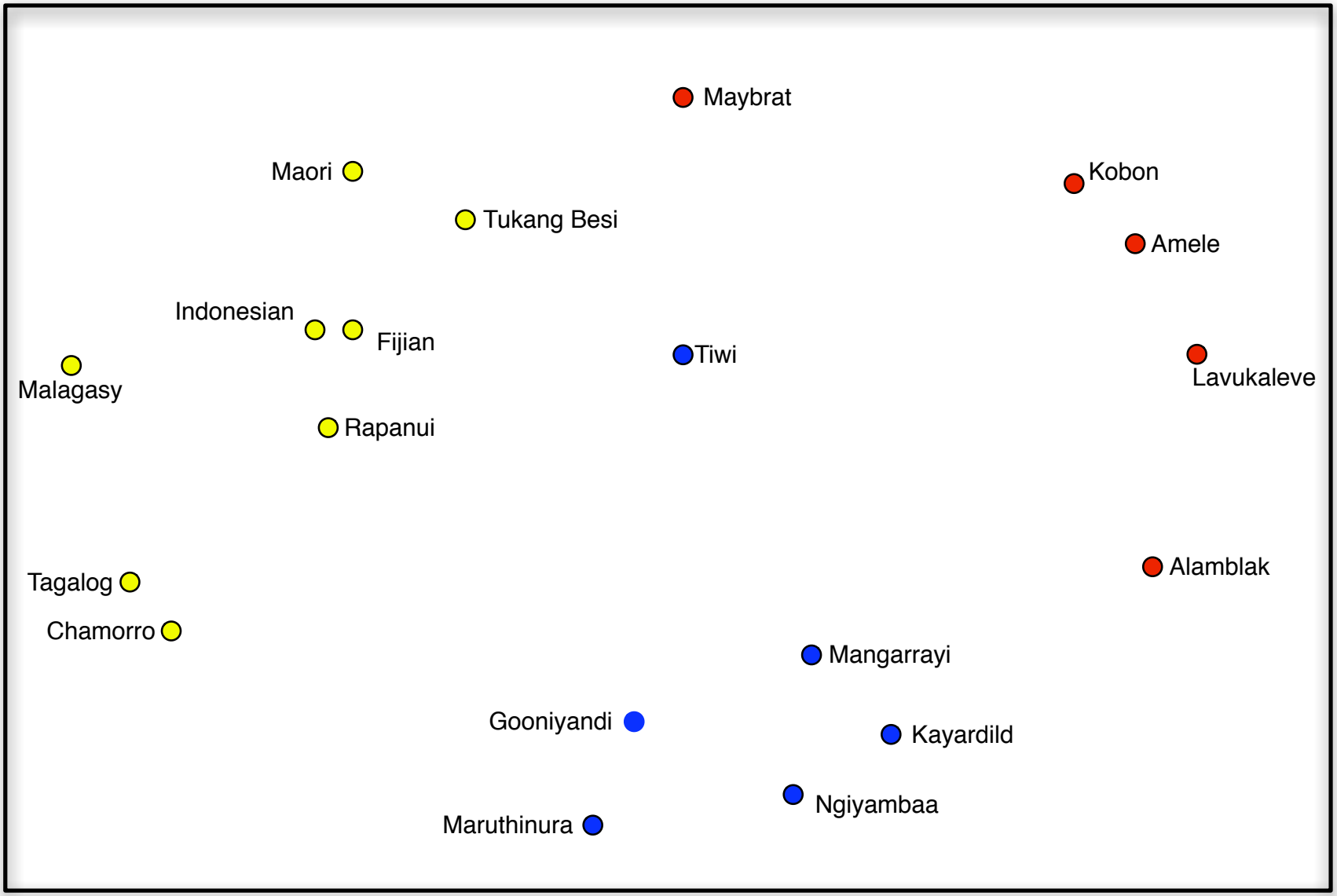
A closer look at geography: the case of Oceania

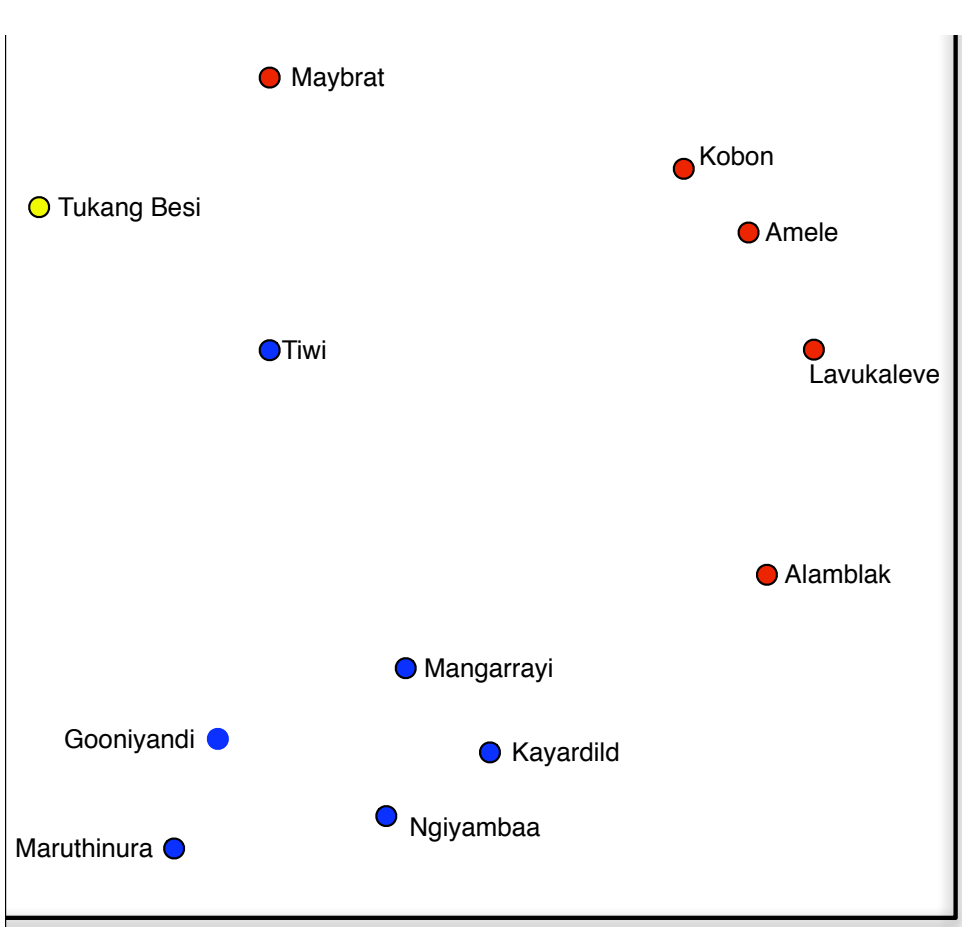


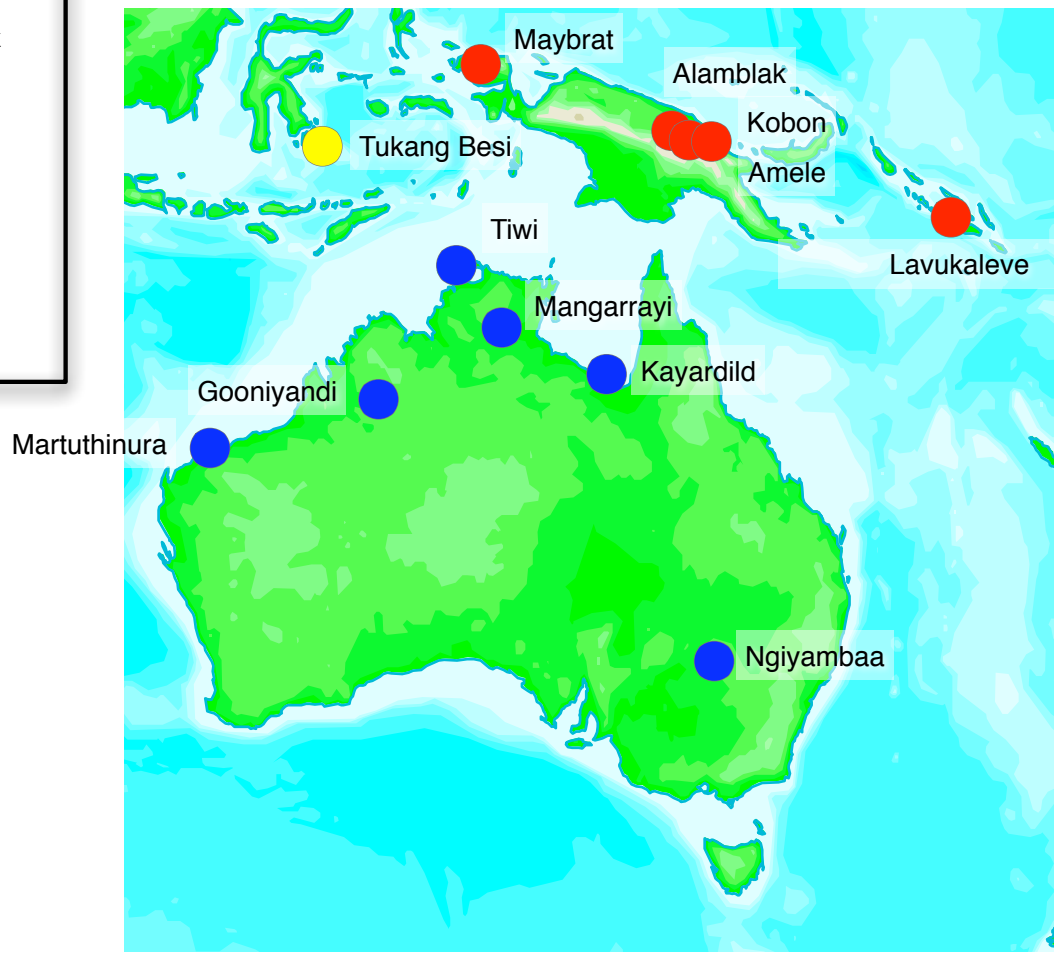
MDS of typological distances



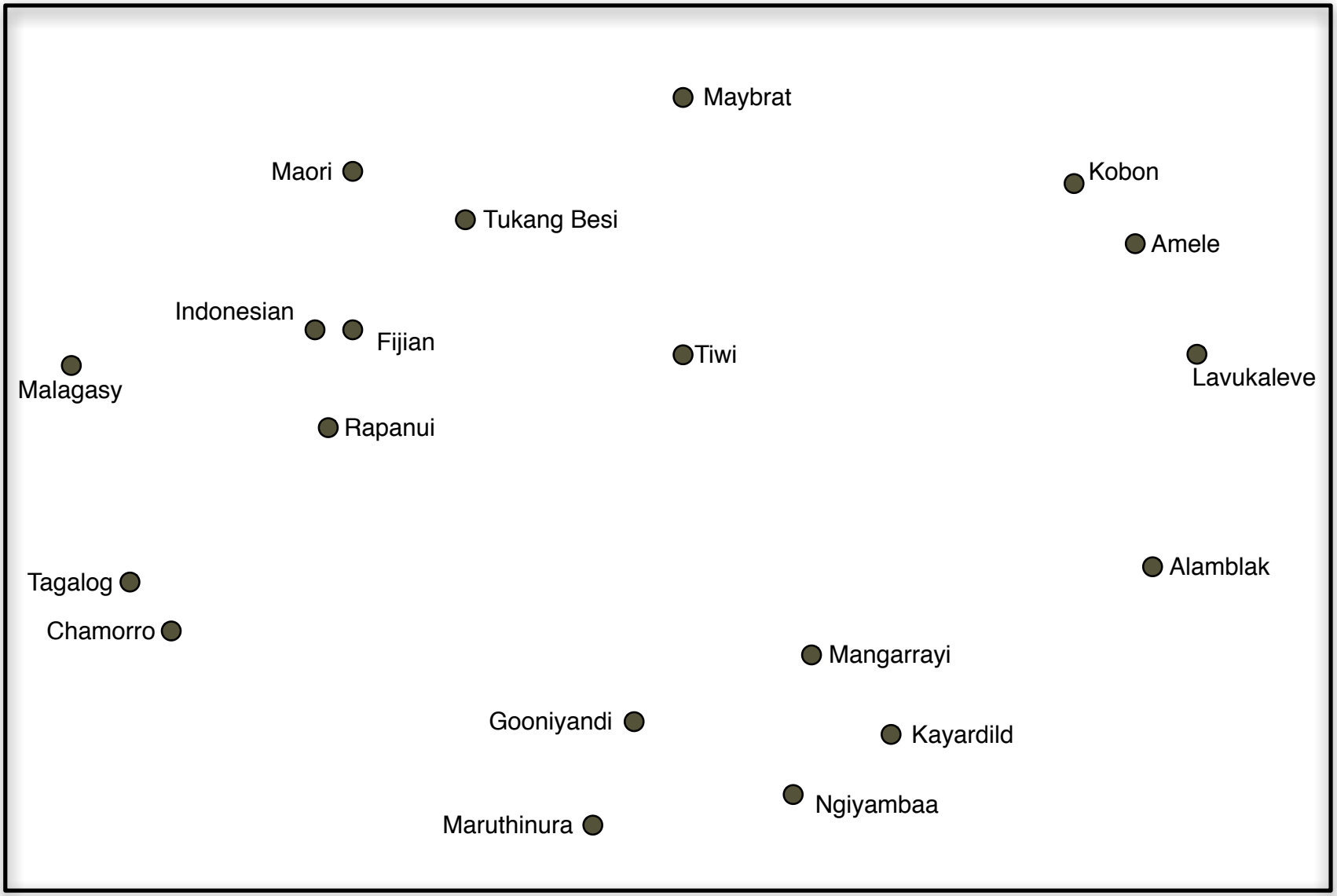
MDS of typological distances



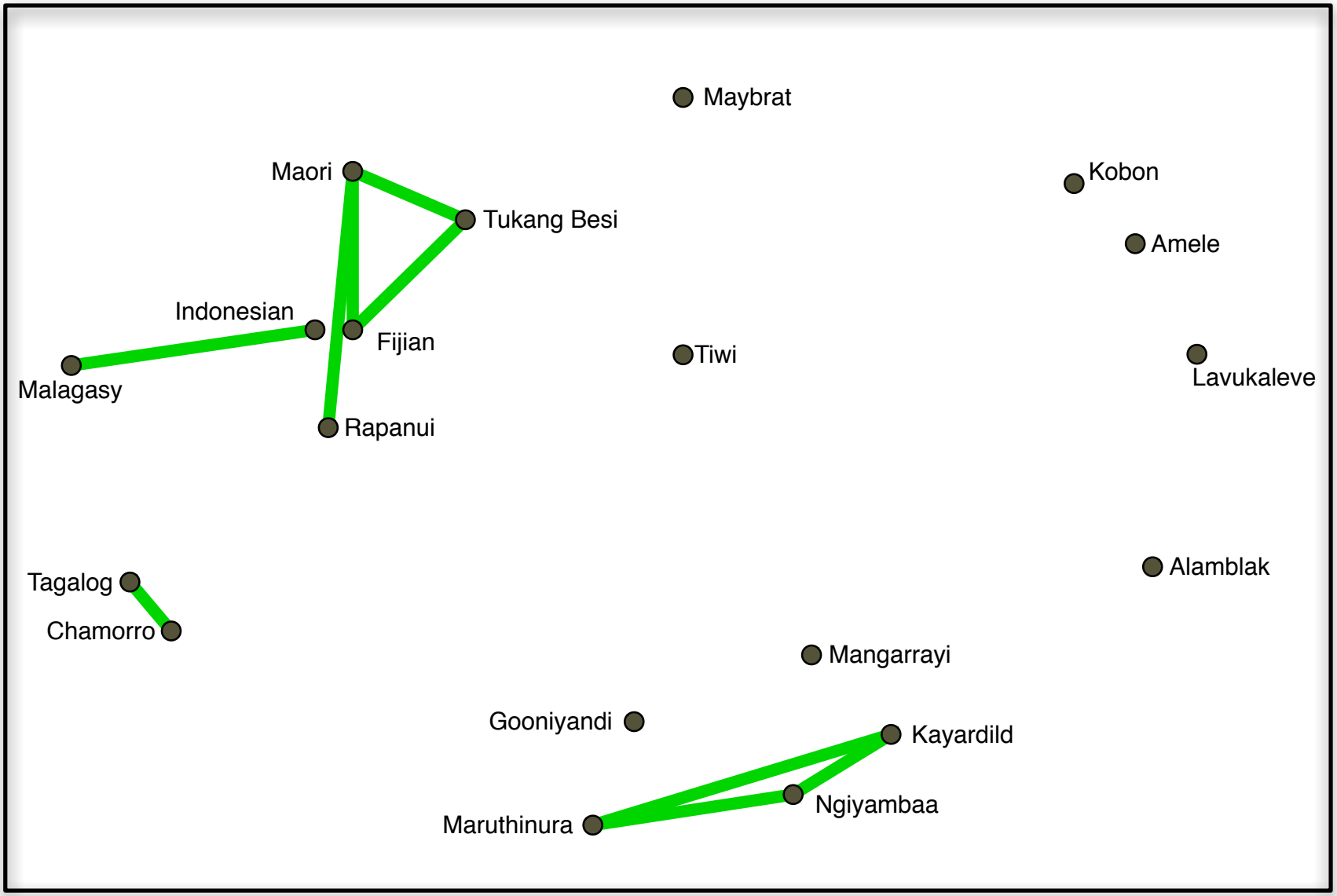

Maybrat
Kobon
Tukang Besi
Amele
Tiwi
Lavukaleve
Alamblak
Mangarrayi
Gooniyandi
Kayardild
Ngiyambaa
Maruthinura


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

MDS of typological distances



MDS of typological distances



Pairs of linguistically (too) similar languages

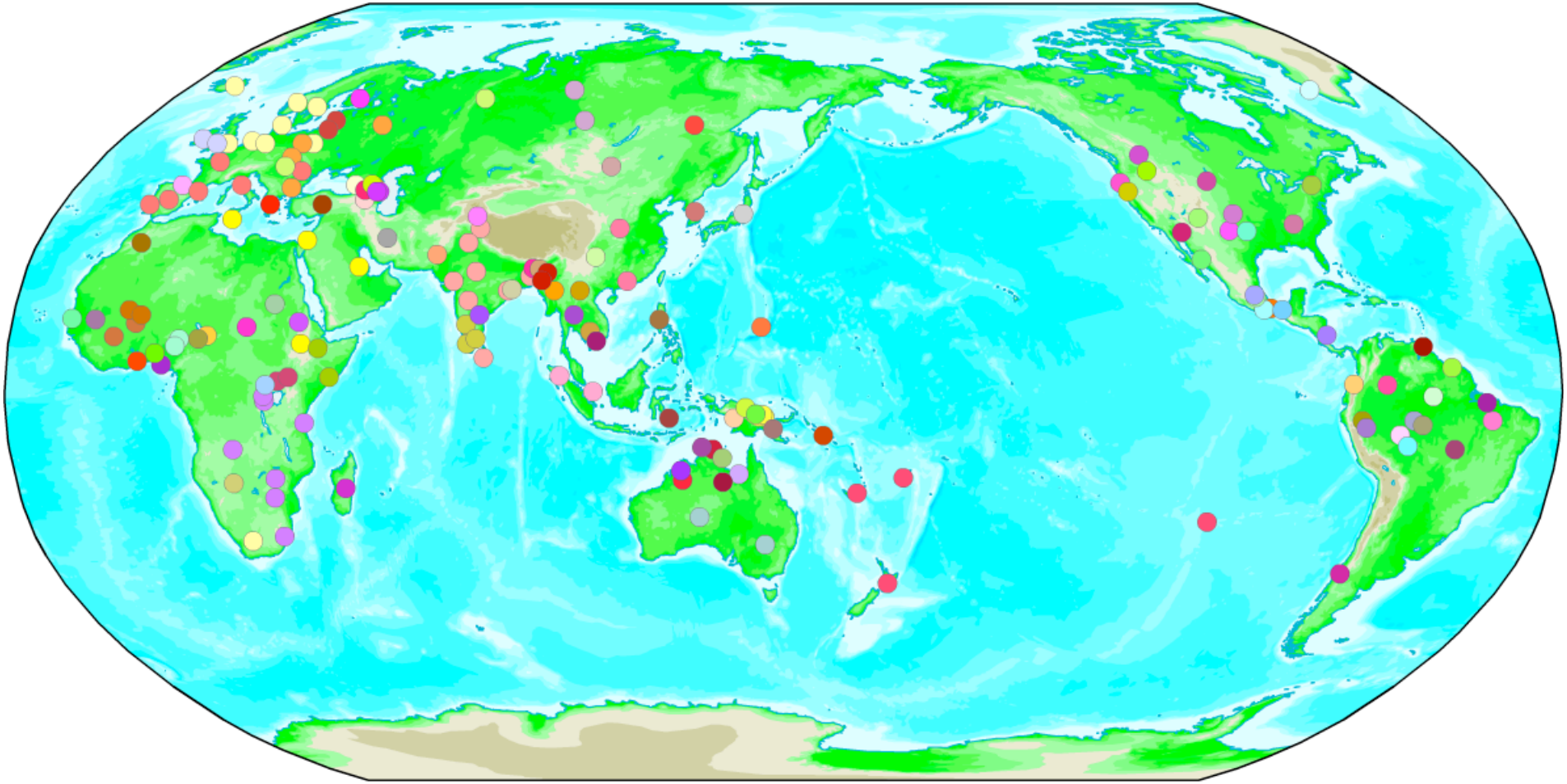
What does this tell us ?

- There appear to be historical “contingencies” hidden in typological data
- Combining typological similarity and geographic distribution gives some clues
- But: can we derive genealogical stability from typological distributions ?

Measuring typological stability (Version I)

- Given a world-wide typological distribution, try to estimate stability
- Attempt A) correlation between individual features and overall similarity (Andreas Dress)
- Attempt B) probability value for diversity on genus level (Søren Wichmann)
- Problem: it does not (yet ?) seem to work

Intensifiers & Reflexive Pronouns



König, Ekkehard & Peter Siemund (with Stephan Töpper). 2005. Intensifiers and Reflexive Pronouns. In: Haspelmath *et al.* (eds.) *The World Atlas of Linguistic Structures.* Oxford University Press.

Measuring typological stability: Variables to consider

- 168 languages
- 24 genera with more than one language *(bias!)*
- in total 79 languages in these 24 groups
- minimally *only 8 changes* needed:
  17 genera are consistent,
  6 genera need one change,
  1 genus needs two changes

Measuring typological stability: Next steps ?

- Given a world-wide typological distribution, try to estimate stability
- Biased samples are needed for this !
- No success yet
- There are many possible methods to try

Using Typological Data for Genealogical Investigations

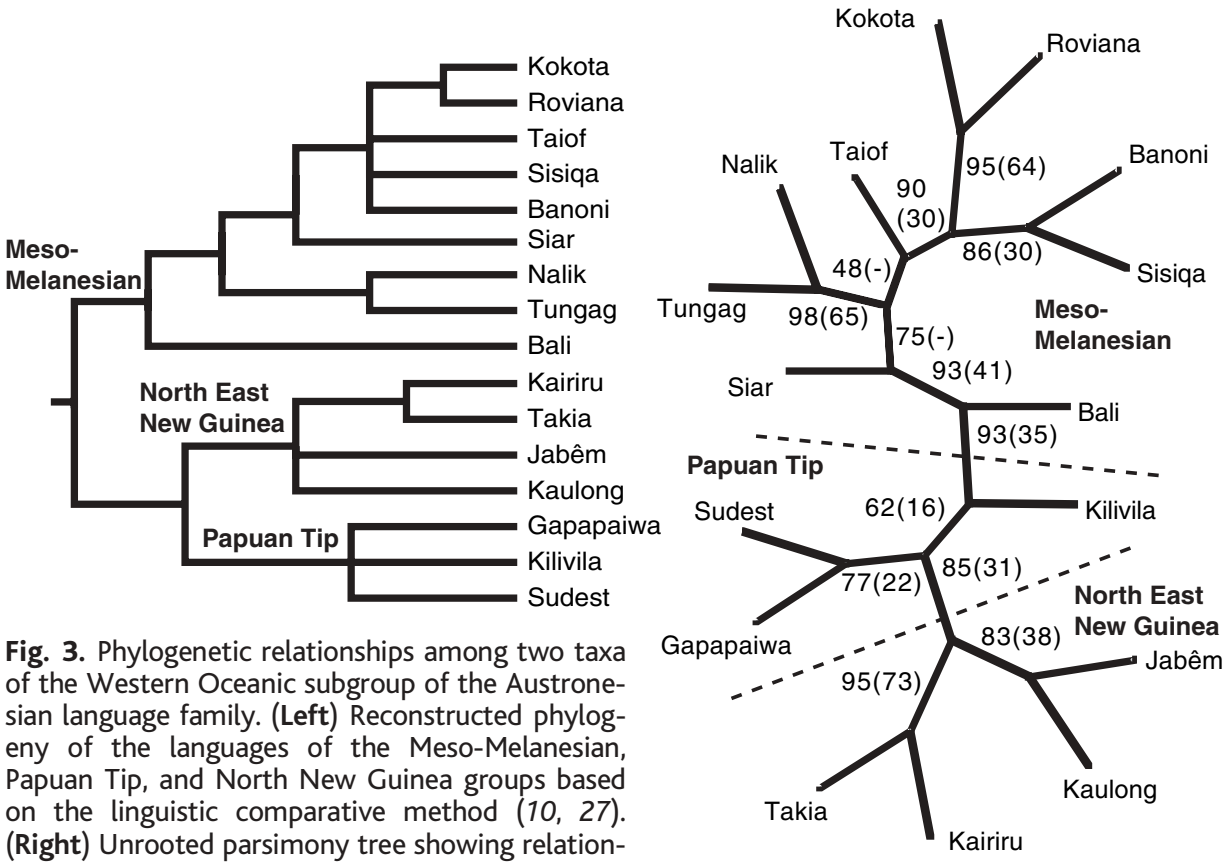
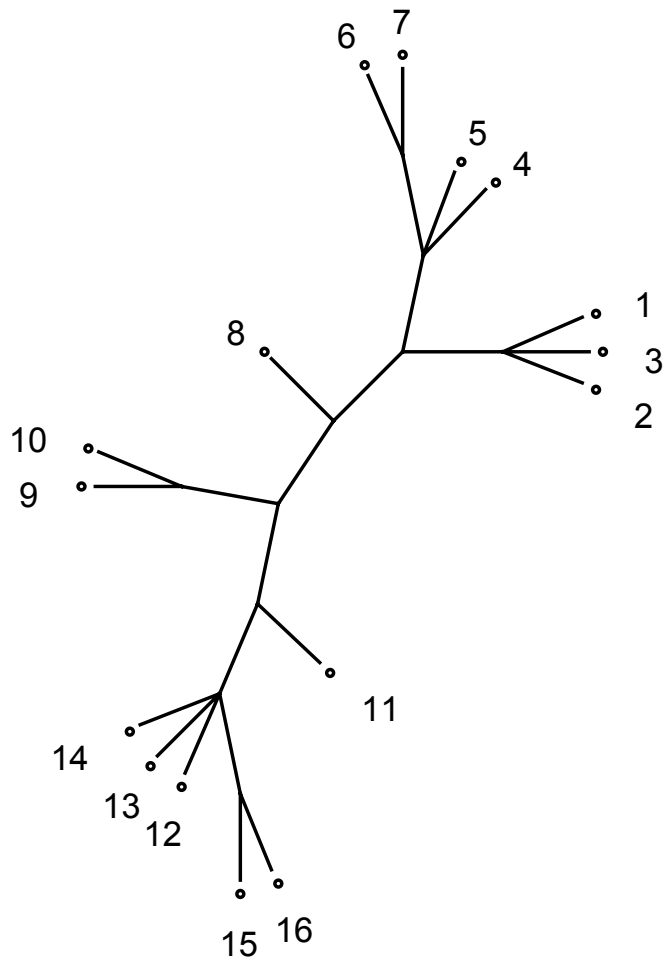


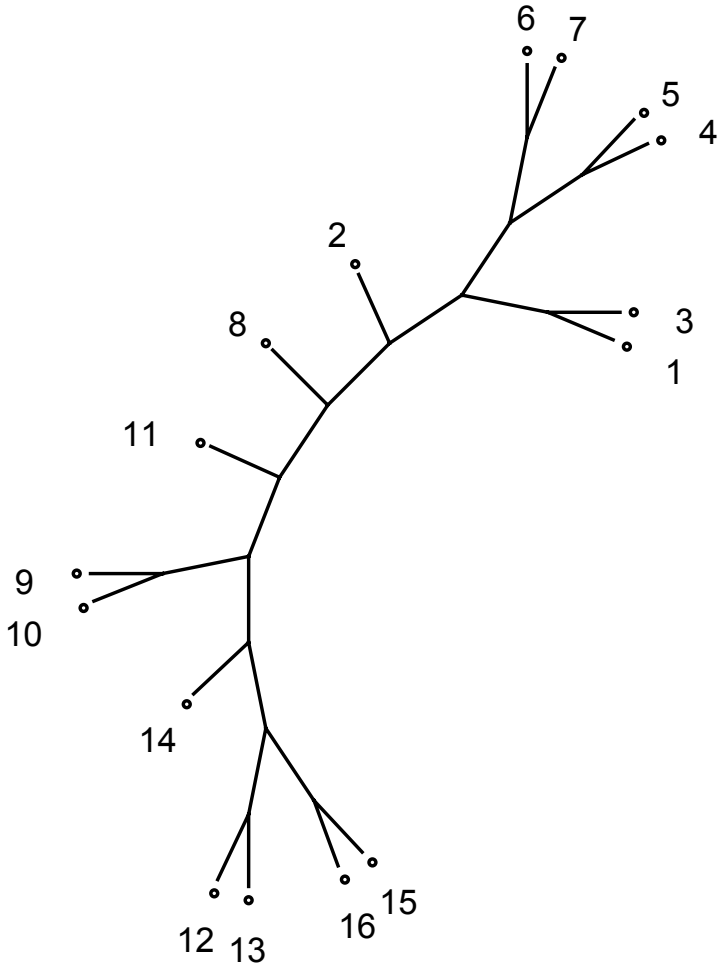
Fig. 3. Phylogenetic relationships among two taxa of the Western Oceanic subgroup of the Austronesian language family. (**Left**) Reconstructed phylogeny of the languages of the Meso-Melanesian, Papuan Tip, and North New Guinea groups based on the linguistic comparative method (*10*, *27*). (**Right**) 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.

Traditional Tree



Dunn *et al.* tree based on typological data

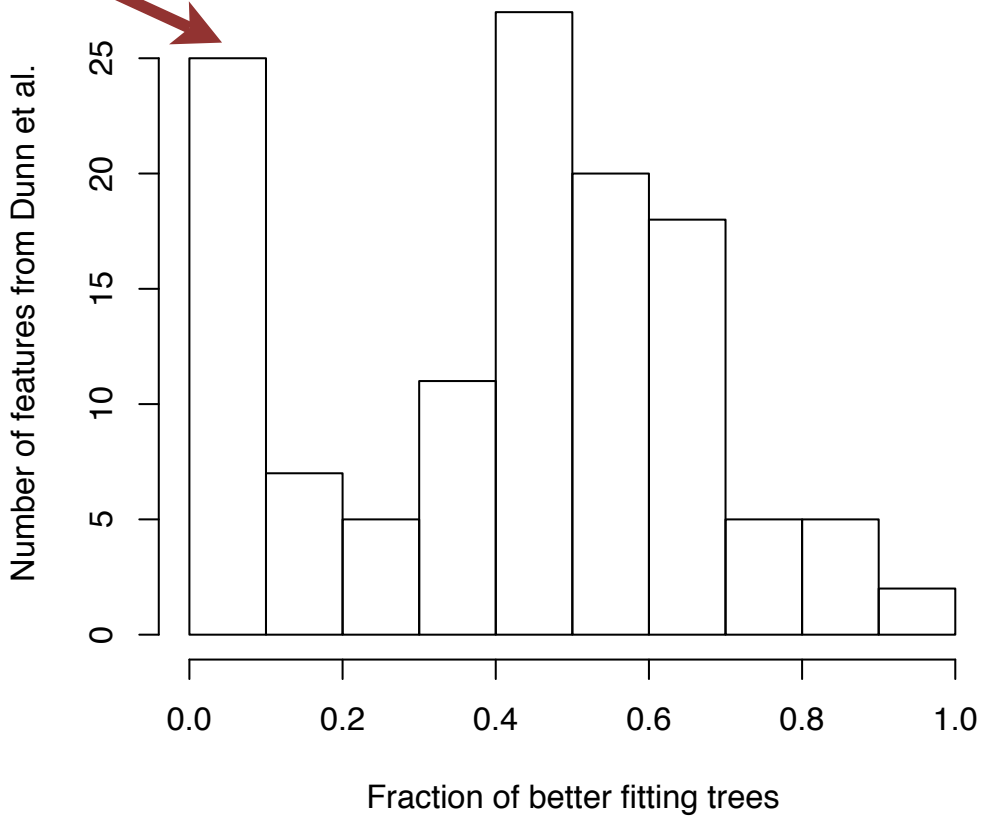


Measuring typological stability, (Version II)

- Given an accepted tree, with many languages sampled from this tree
- how good does a typological feature predict this tree
- Energy-based measurement of fit between a dataset and a tree (work by Mihai Albu)
- Take a large set of random trees, and determine how good the “real” tree fits

Distribution of fits of all 125 features

(Too) many good fits!



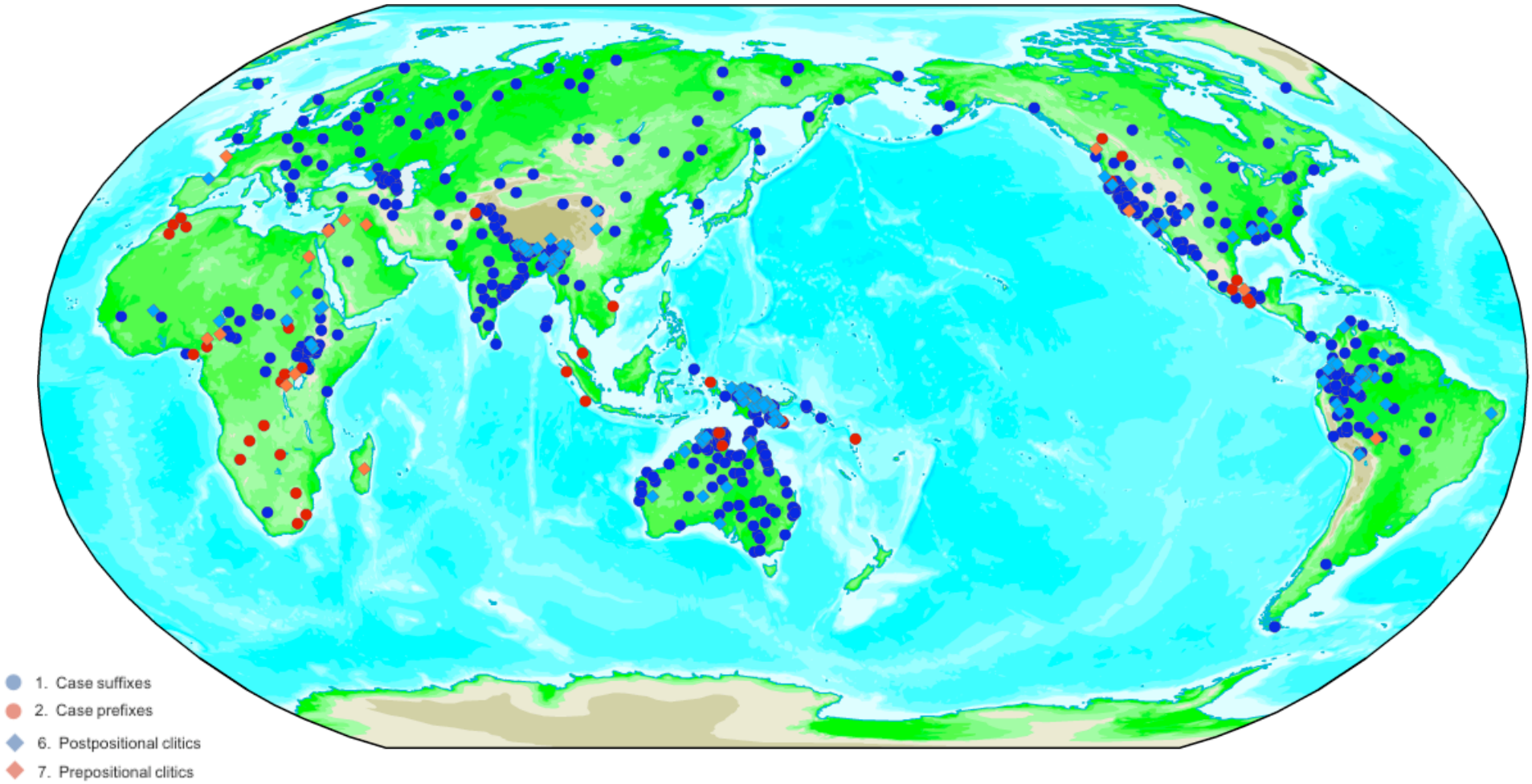
Best features for Austronesian ($p < .05$)

97: VS intransitive clauses
59: S-prefix
61: A-prefix
44: Decimal numerals
83: Reflexive morphology (including affixes and clitics)
14: Article-noun order
35: Possessive classifiers
13: Indefinite or non-specific articles
12: Definite or specific articles
52: Postpositions
85: Verb classifiers
41: Marked possessor
74: Recipient object
66: Verb variation clause type
67: Verb variation person
98: V-initial transitive clauses
112: Clause chaining

Measuring Genealogical Stability (Version III)

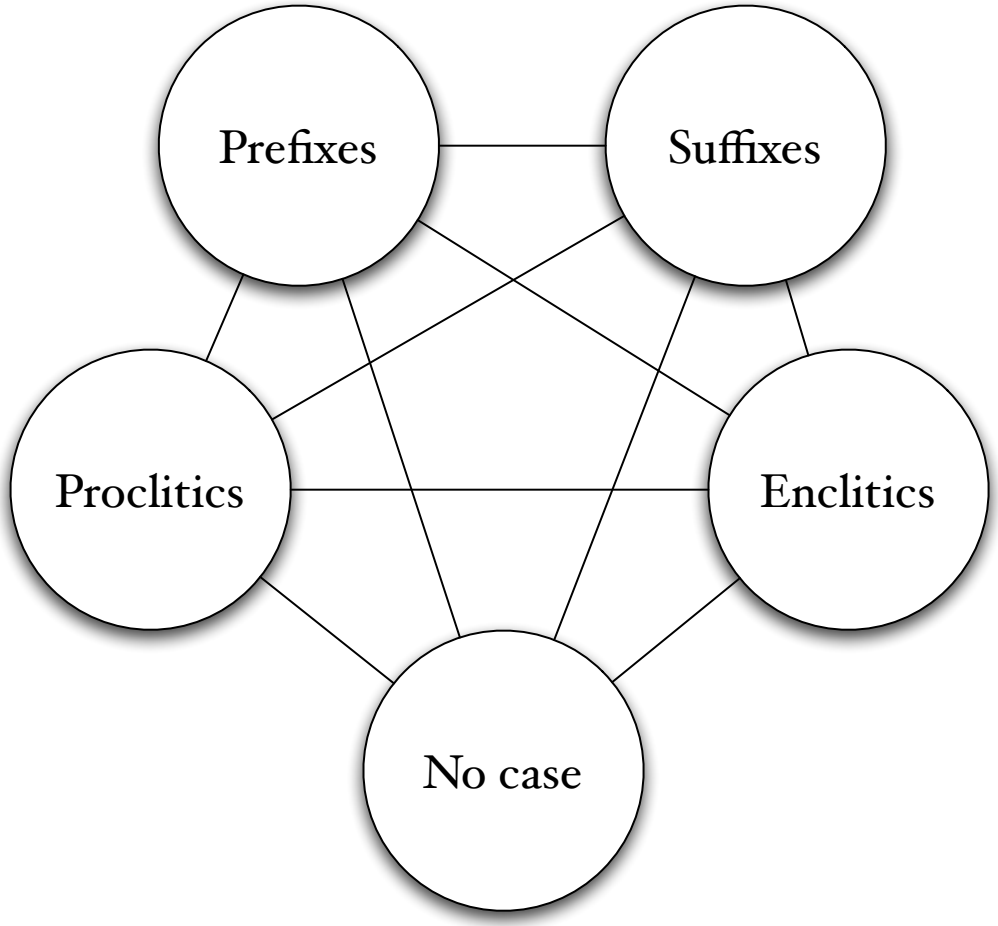
- Are there regularities in typological change ?
- To investigate this, a *sample from accepted family-trees* is needed
- The result would be a matrix of *Transition Probabilities* (Maslova 2000)

Example: *WALS*, map 51: Position of Case affixes (selection)



Dryer, Matthew. 2005. Position of Case Affixes. In: Haspelmath *et al.* (eds.)
The World Atlas of Linguistic Structures. Oxford University Press.

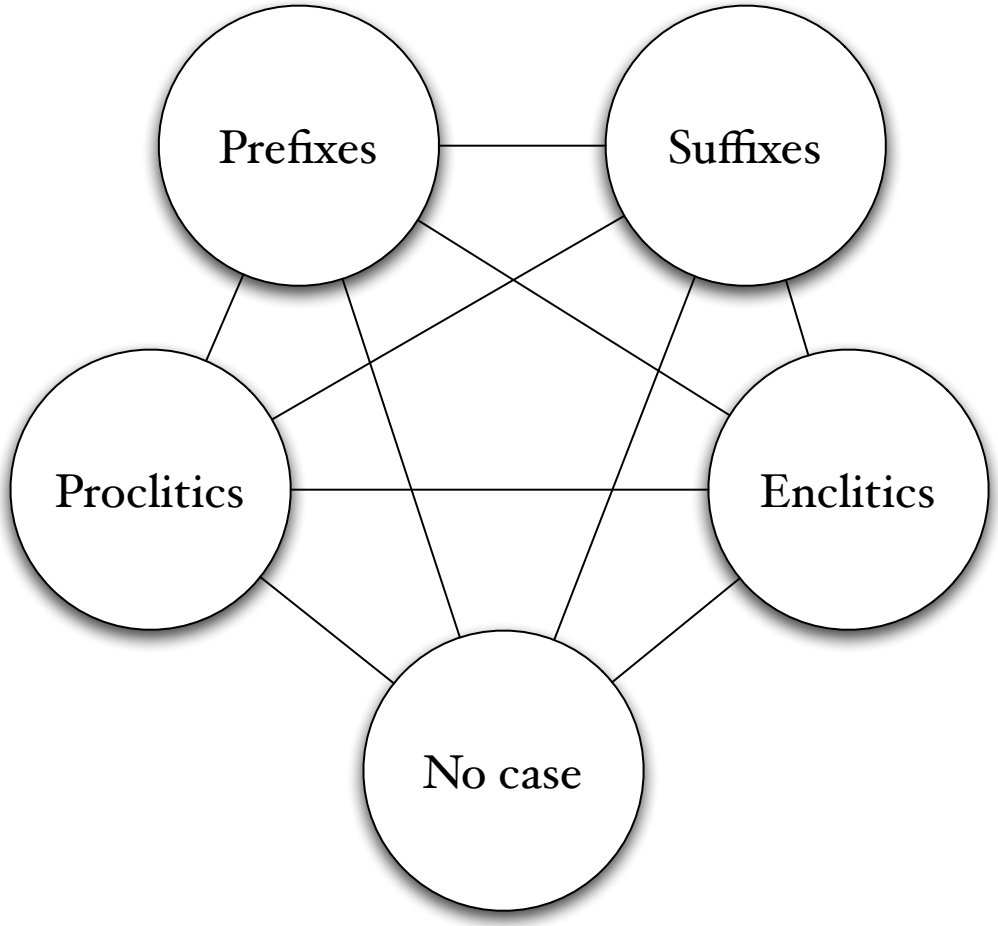
Undifferentiated typology



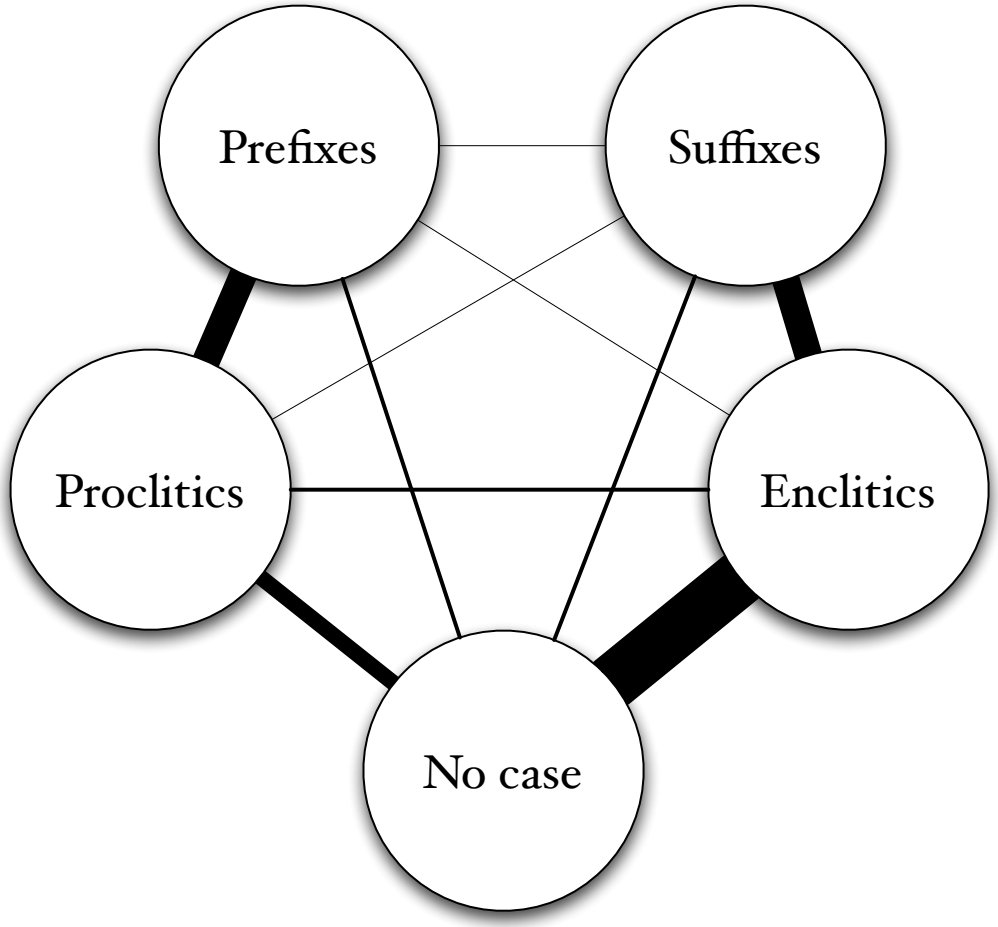
Undifferentiated typology

	no case	proclitics	enclitics	prefixes	suffixes
no case	0	1	1	1	1
proclitics	1	0	1	1	1
enclitics	1	1	0	1	1
prefixes	1	1	1	0	1
suffixes	1	1	1	1	0

Undifferentiated typology



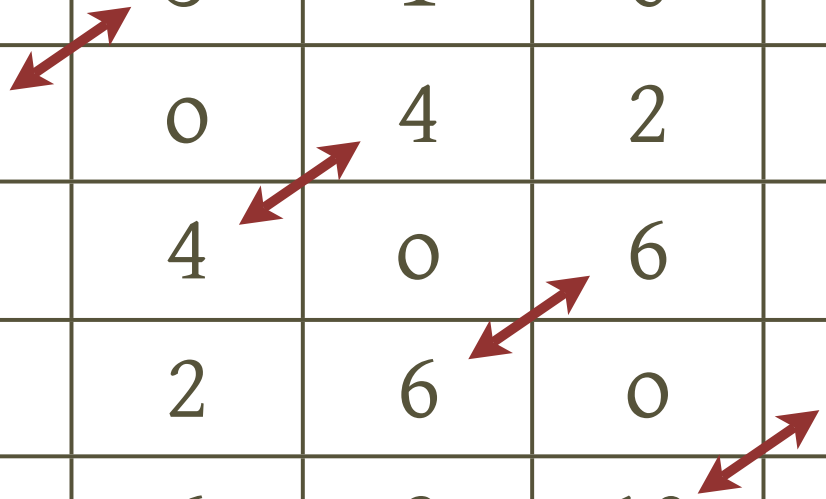
Similarities



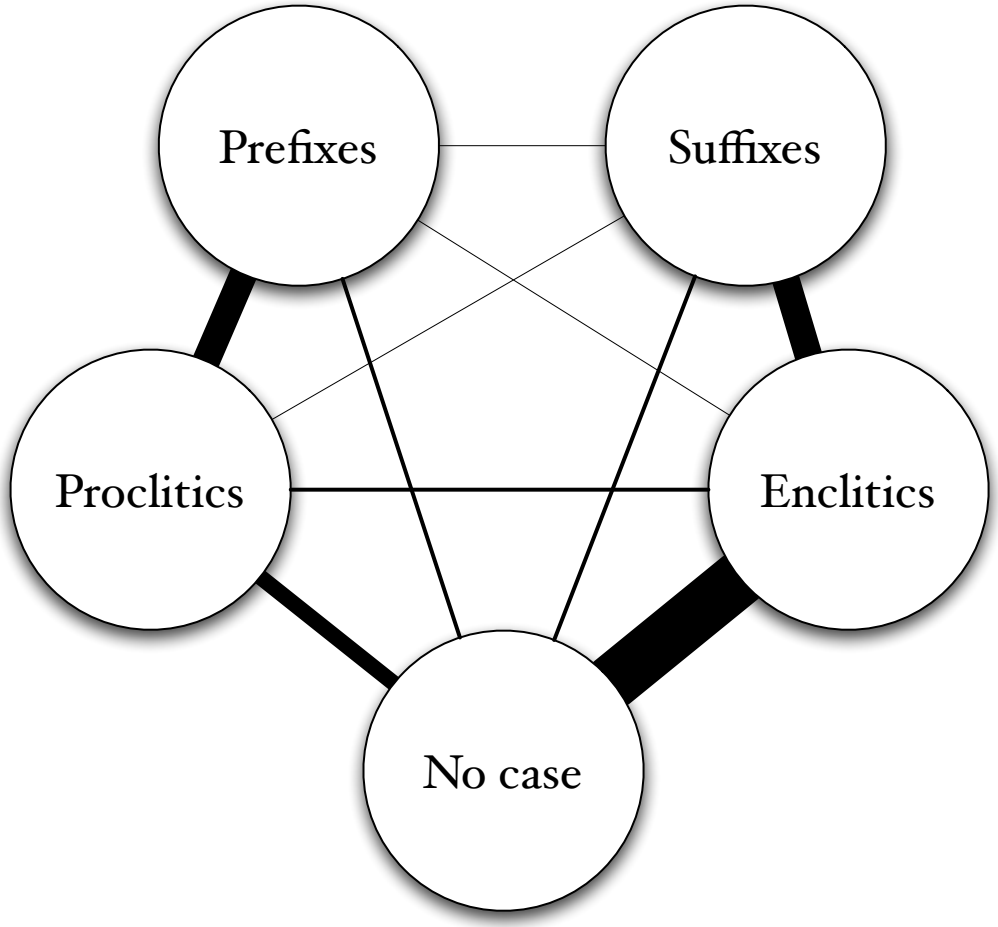
Similarities

(relative values; higher values are less likely)

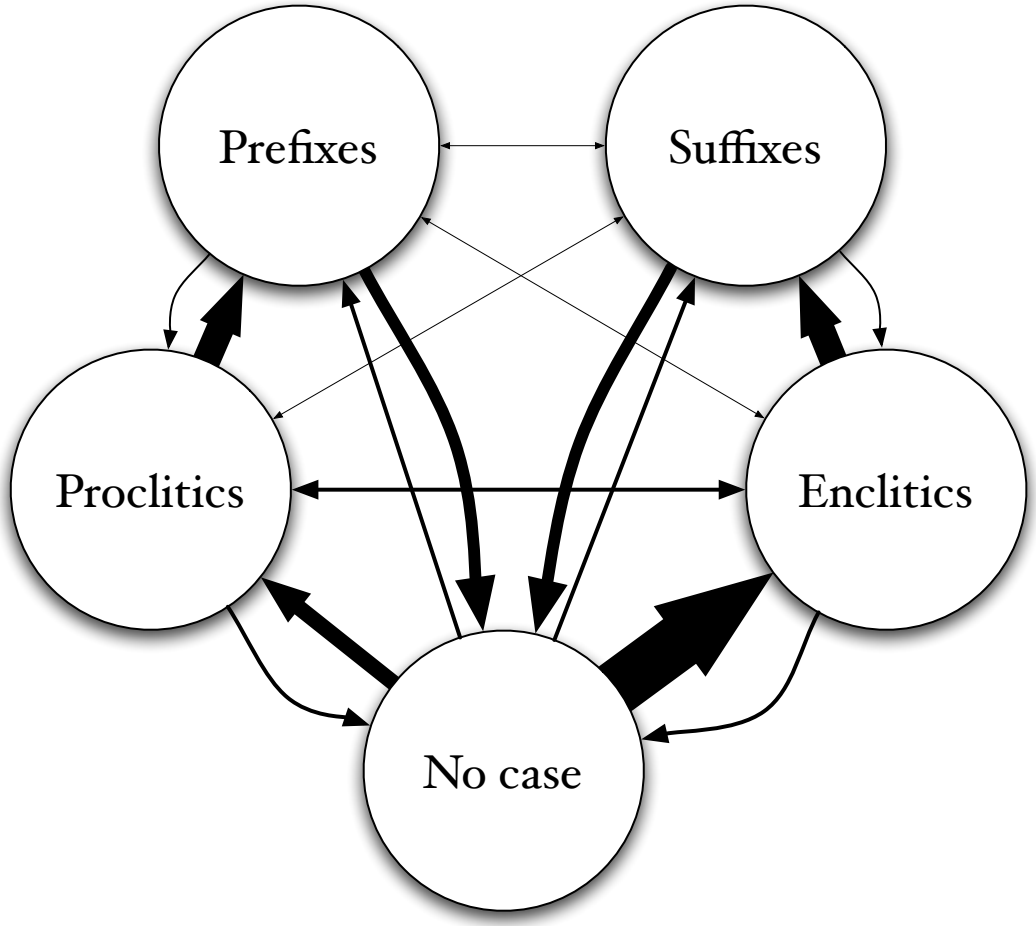
	no case	proclitics	enclitics	prefixes	suffixes
no case	0	3	1	6	6
proclitics	3	0	4	2	6
enclitics	1	4	0	6	2
prefixes	6	2	6	0	10
suffixes	6	6	2	10	0



Similarities



Transition probabilities



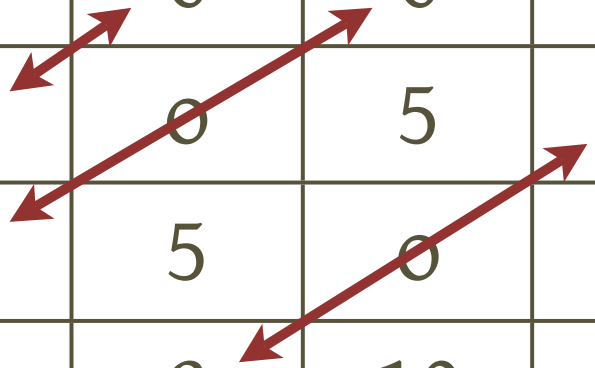
Transition probabilities

(relative values; higher values are less likely)

From:

To:

	no case	proclitics	enclitics	prefixes	suffixes
no case	0	6	6	4	4
proclitics	3	0	5	8	10
enclitics	1	5	0	10	8
prefixes	6	2	10	0	10
suffixes	6	10	2	10	0



Work to do ...

- Yes, historical "coincidences" are important
- But: There are different historical scenarios
  - Unknown language genealogy
  - Spread of features individually (substrate, superstrate, long-term borrowing)
- The real problem of typology is to distinguish between these historical scenarios
- For universals we need something different:

Structural Coevolution

The End