On the geographical distribution of phoneme inventories: a reply to Atkinson

Michael Cysouw
(with Dan Dediu, MPI Nijmegen)

Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa

Quentin D. Atkinson[1,2]*

Human genetic and phenotypic diversity declines with distance from Africa, as predicted by a serial founder effect in which successive population bottlenecks during range expansion progressively reduce diversity, underpinning support for an African origin of modern humans. Recent work suggests that a similar founder effect may operate on human culture and language. Here I show that the number of phonemes used in a global sample of 504 languages is also clinal and fits a serial founder–effect model of expansion from an inferred origin in Africa. This result, which is not explained by more recent demographic history, local language diversity, or statistical non-independence within language families, points to parallel mechanisms shaping genetic and linguistic diversity and supports an African origin of modern human languages.

The number of phonemes—perceptually distinct units of sound that differentiate words—in a language is positively correlated with the size of its speaker population (*1*) in such a way that small populations have fewer phonemes. Languages continually gain and lose phonemes because of stochastic processes (*2*, *3*). If phoneme distinctions are more likely to be lost in small founder populations, then a succession of founder events during range expansion should progressively reduce phonemic diversity with increasing distance from the point of origin, paralleling the serial founder effect observed in population genetics (*4*–*9*). A founder effect has already been used to explain patterns of variation in other cultural replicators, including human material culture (*10*–*13*) and birdsong (*14*). A range of possible mechanisms (*15*) predicts similar dynamics governing the evolution of phonemes (*11*, *16*) and language generally (*17*–*20*). This raises the possibility that the serial founder–effect model used to trace our genetic origins to a recent expansion from Africa (*4*–*9*) could also be applied to global phonemic diversity to investigate the origin and expansion of modern human languages. Here I examine geographic variation in phoneme inventory size using data on vowel, consonant, and tone inventories taken from 504 languages in the World Atlas of Language Structures (WALS) (*21*), together with information on language location, taxonomic affiliation, and speaker demography (Fig. 1 and table S1) (*15*).

Consistent with previous work (*1*), speaker population size is a significant predictor of phonemic diversity (Pearson's correlation $r = 0.385$, df = 503, $P < 0.001$), with smaller population size predicting smaller overall phoneme inventories (fig. S1A). The same relationship holds for vowel ($r = 0.378$, df = 503, $P < 0.001$) and tone ($r = 0.230$, df = 503, $P < 0.001$) inventories separately, with a weaker, though still significant, effect of population size on consonant diversity ($r = 0.131$, df = 503, $P = 0.003$). To account for any non-independence within language families, the analysis was repeated, first using mean values at the language family level (table S2) and then using a hierarchical linear regression framework to model nested dependencies in variation at the family, subfamily, and genus levels (*15*). These analyses confirm that, consistent with a founder effect model, smaller population size predicts reduced phoneme inventory size both between families (family-level analysis $r = 0.468$, df = 49, $P < 0.001$; fig. S1B) and within families, controlling for taxonomic affiliation {hierarchical linear model: fixed-effect coefficient (β) = 0.0338 to 0.0985 [95% highest posterior density (HPD)], $P = 0.009$}.

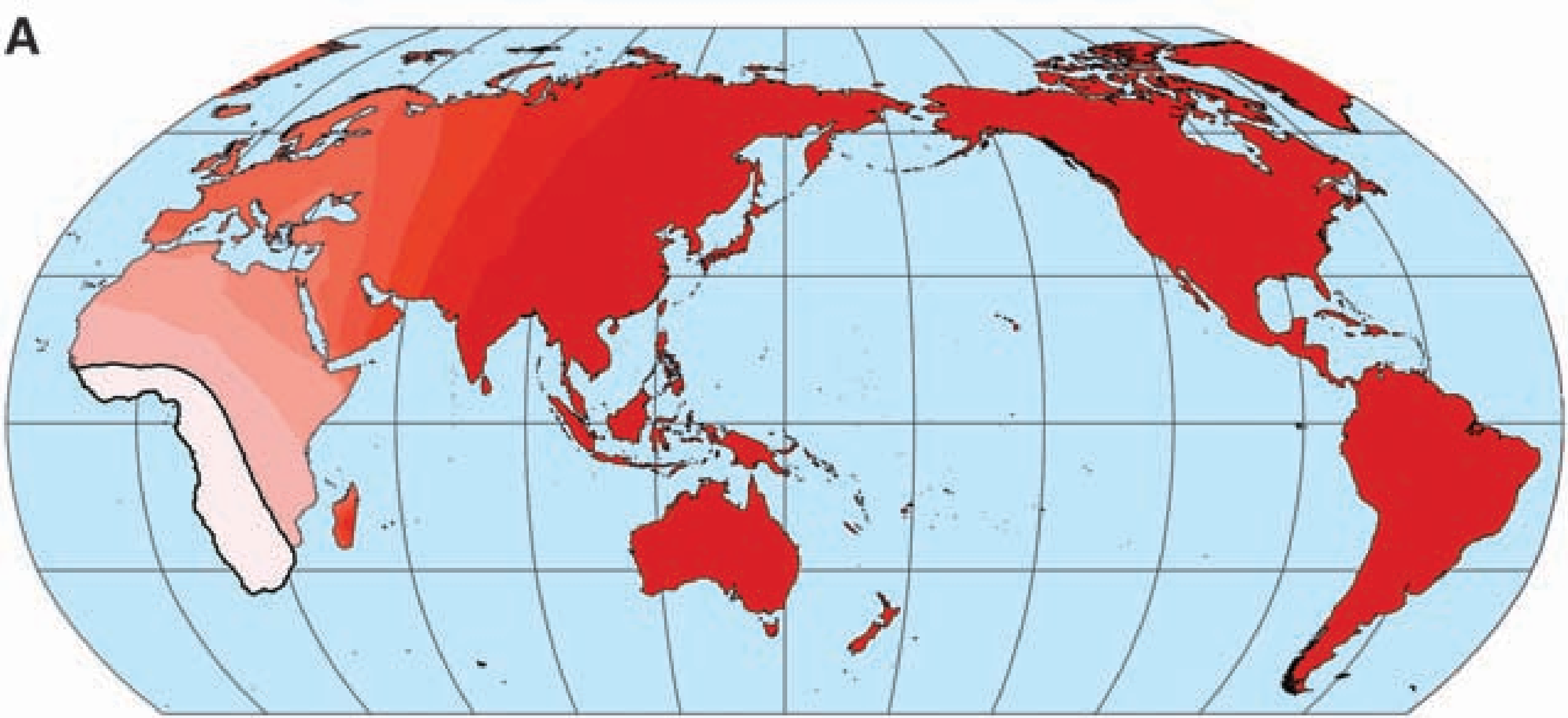
Figure 1B shows clear regional differences in phonemic diversity, with the largest phoneme inventories in Africa and the smallest in South America and Oceania. A series of linear regressions was used to predict phoneme inventory size from the log of speaker population size and distance from 2560 potential origin locations around the world (*15*). Incorporating modern speaker population size into the model controls for geographic patterning in population size and means that the analysis is conservative about the amount of variation attributed to ancient demography. Model fit was evaluated with the Bayesian information criterion (BIC) (*22*). Following previous work (*5*, *6*), the set of origin locations within four BIC units of the best-fit location was taken to be the most likely area of origin under a serial founder–effect model.

The origin locations producing the strongest decline in phonemic diversity and best-fit model lie across central and southern Africa (Fig. 2A). This region could represent either a single origin for modern languages or the main origin under a polygenesis scenario. The best-fit model incorporating population size and distance from the origin explains 31% of the variance in phoneme inventory size [correlation coefficient (R) = 0.558, $F_{2,501} = 113.463$, $P < 0.001$] (Fig. 3). Both population size ($r_{population} = 0.146$, $P = 0.002$)

[1]Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand. [2]Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 64 Banbury Road, Oxford OX2 6PN, UK.

*E-mail: q.atkinson@auckland.ac.nz

Out-of-Africa Gradient



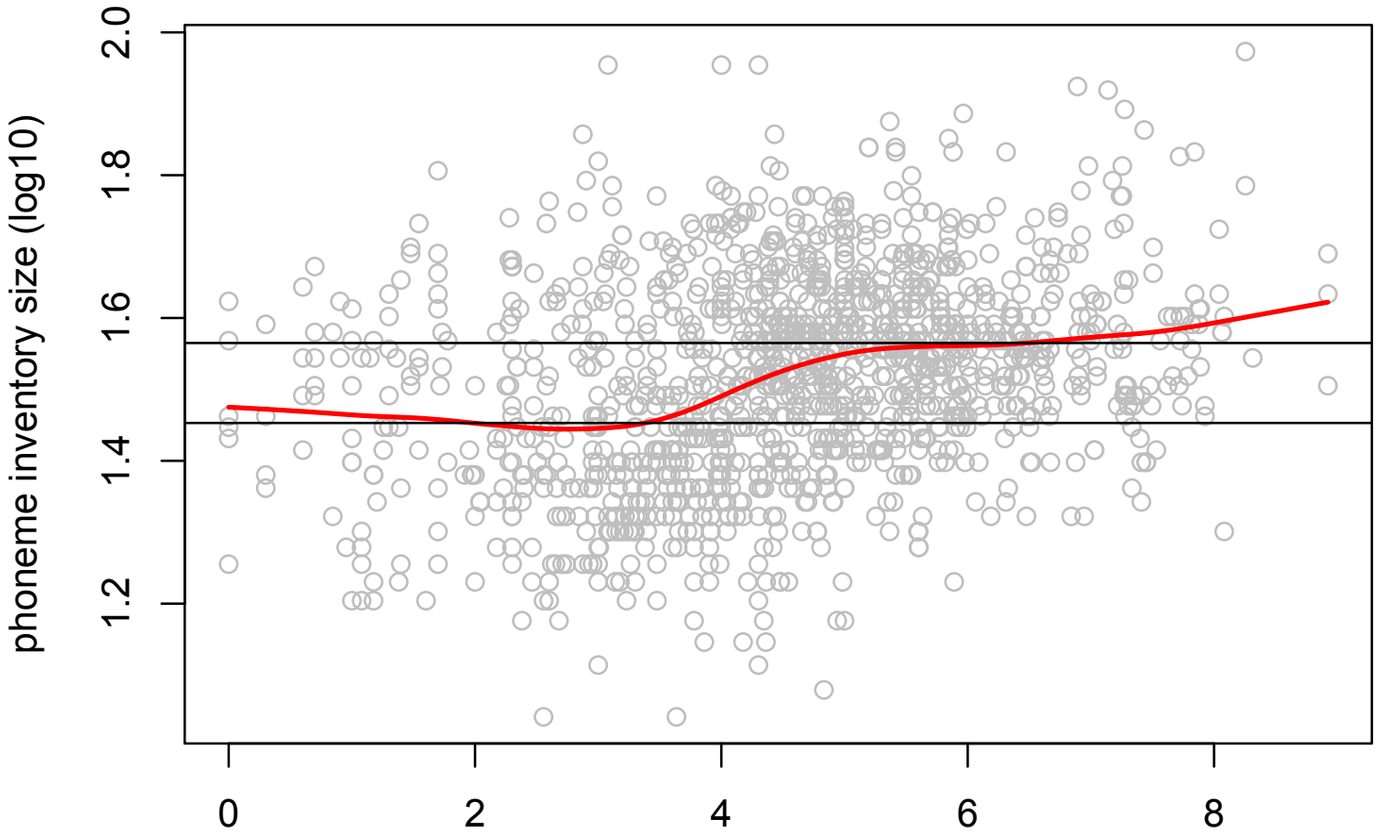
Atkinson, Quentin D. 2011. Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. Science 332: 346-349.

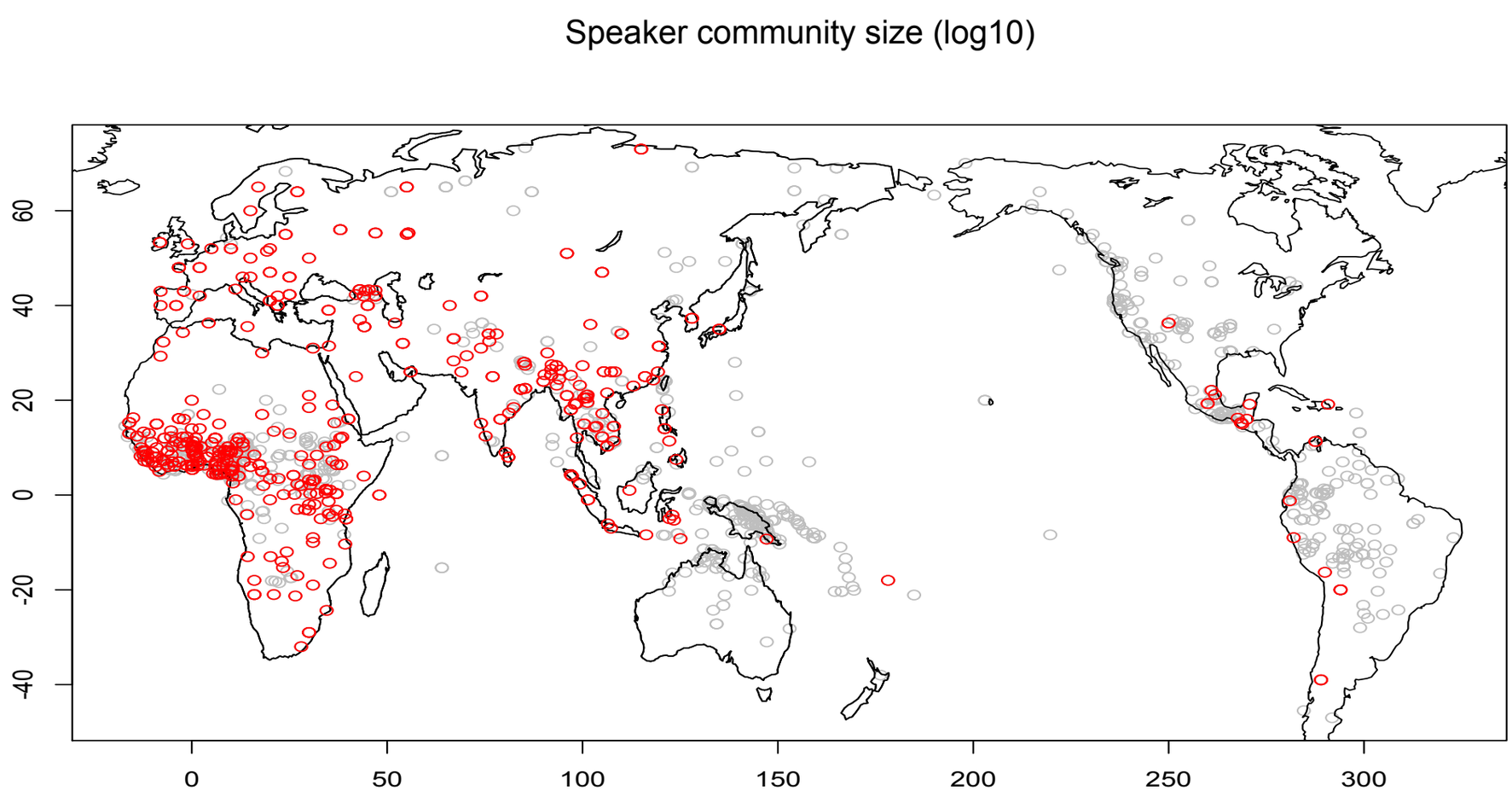
Problems

- Based on three equally-weighted WALS features
  - Consonant Inventory
  - Vowel Quality Inventory
  - Tone Inventory

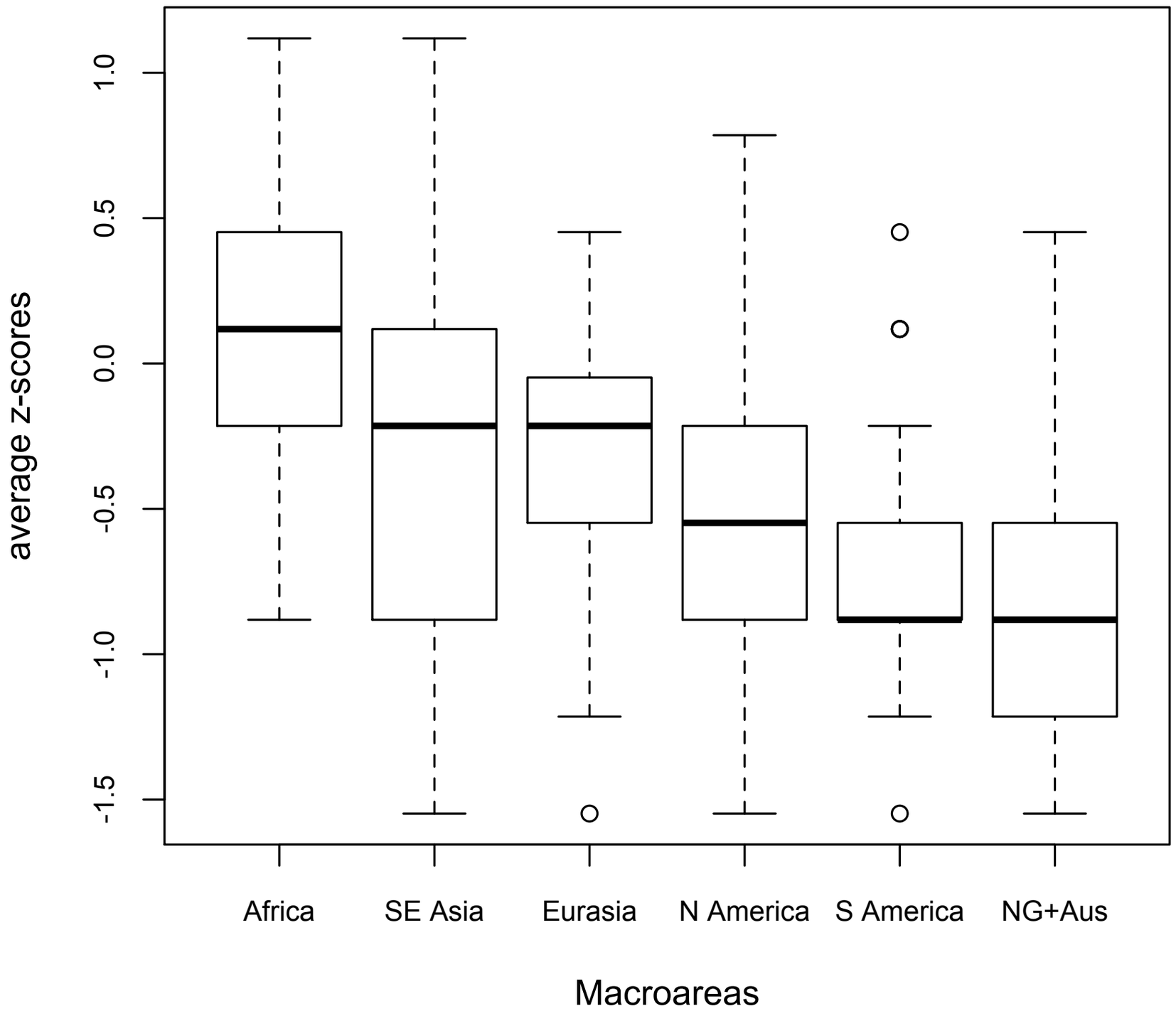
Mean
28
6
~1

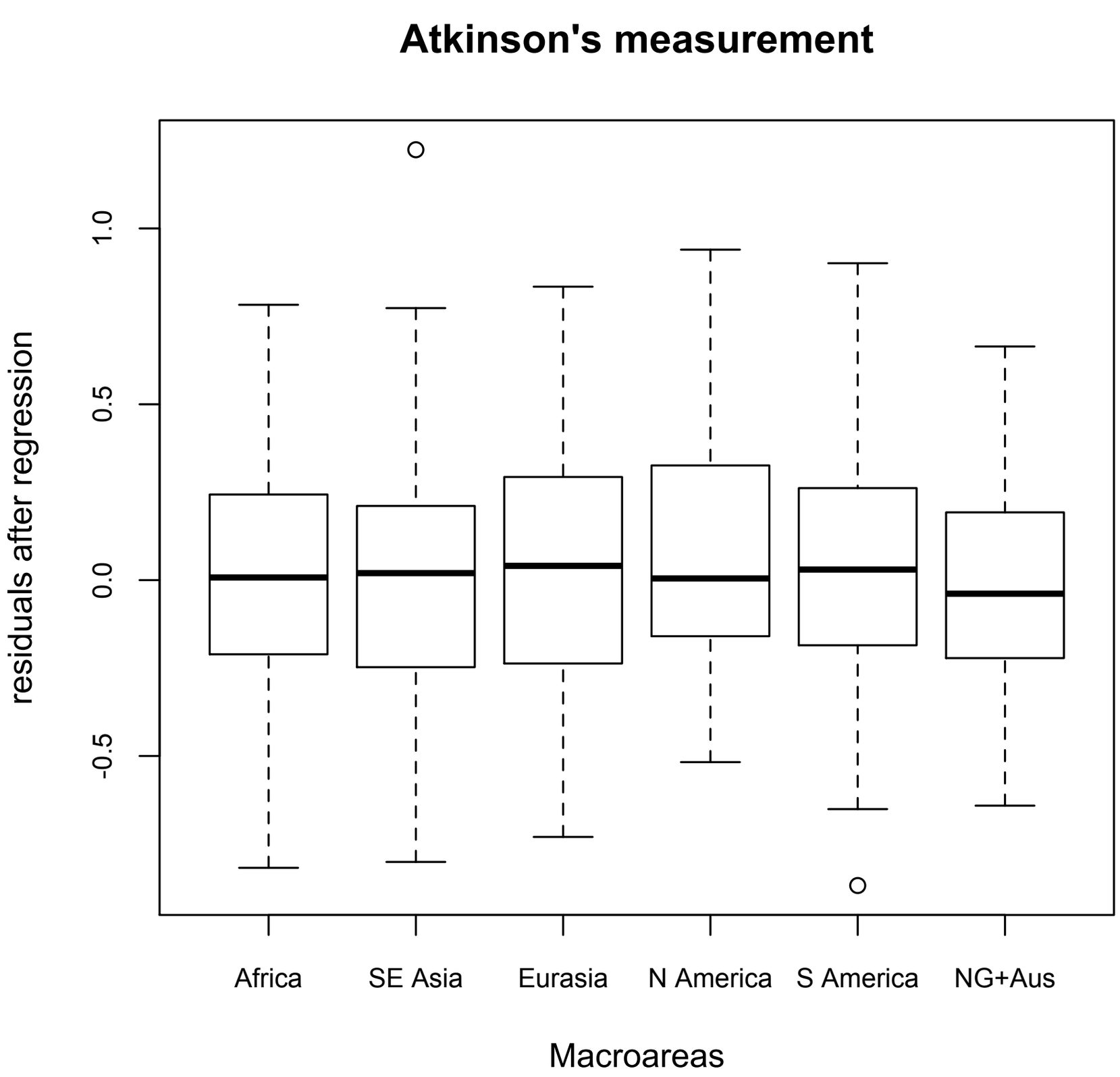
- Correlation with speaker community size
- Simplistic geographical distance measurement

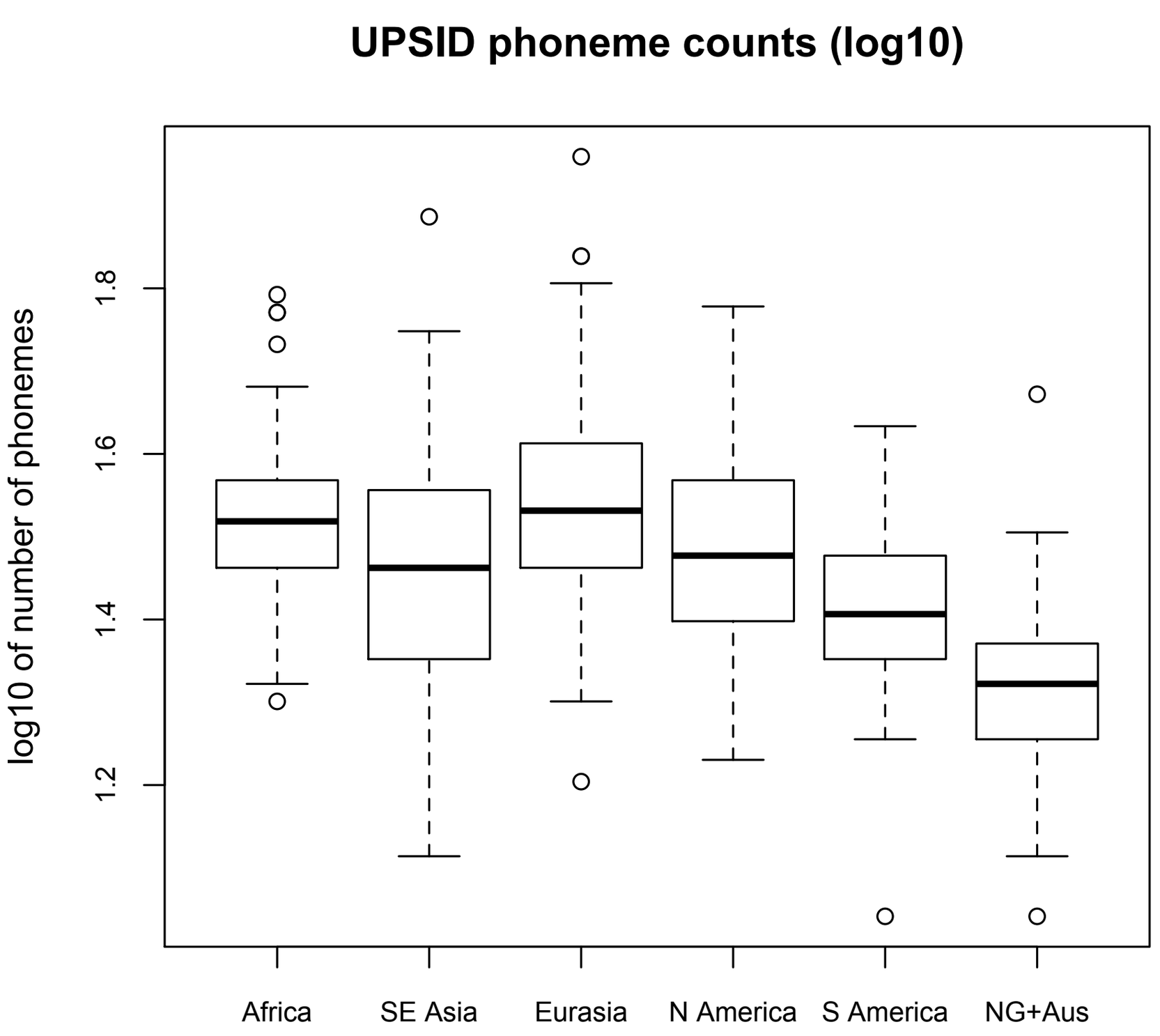

phoneme inventory size (log10)
2.0
1.8
1.6
1.4
1.2
0
2
4
6
8


Speaker community size (log10)
60
40
20
0
-20
-40
0
50
100
150
200
250
300

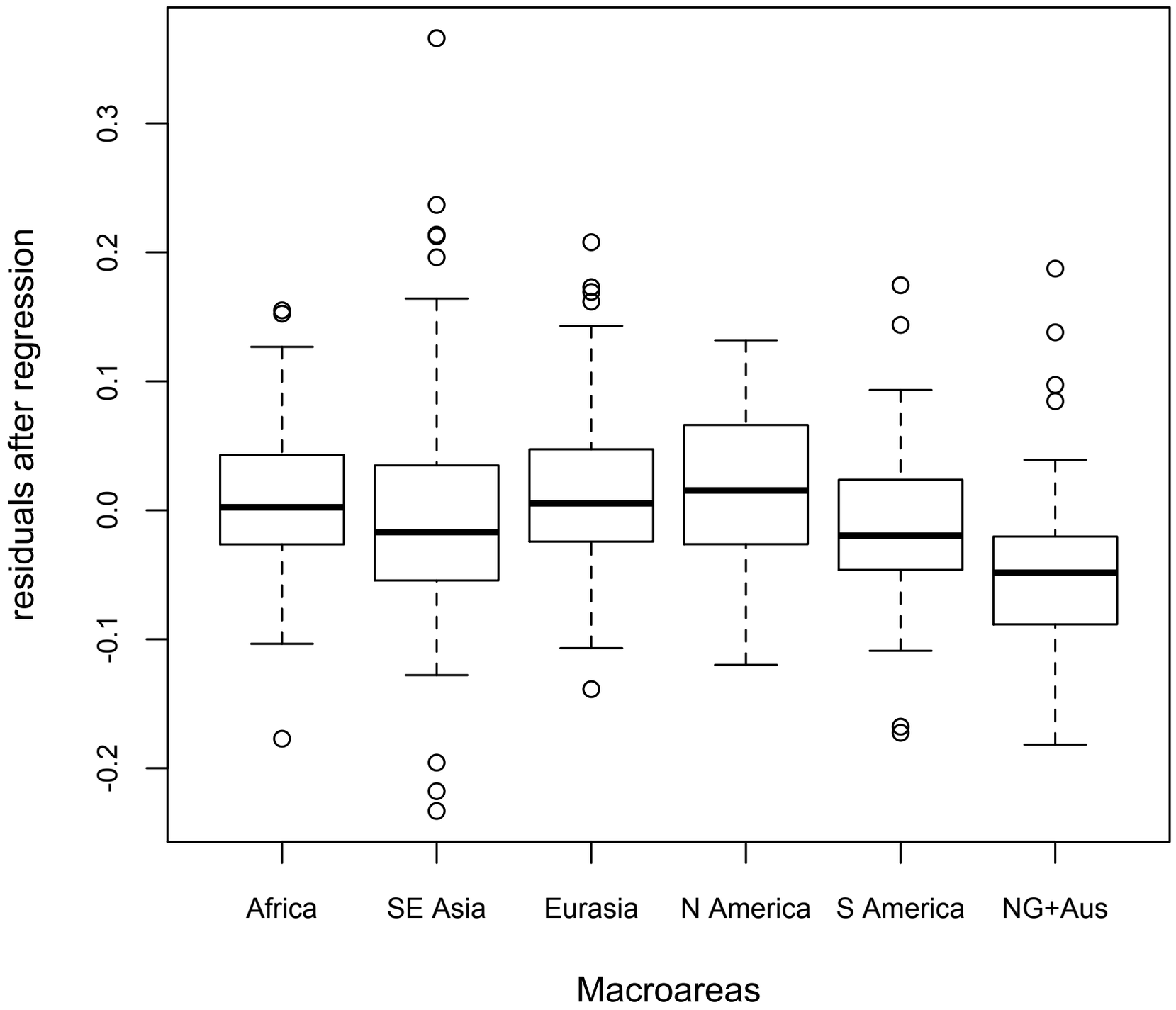
Atkinson's measurement

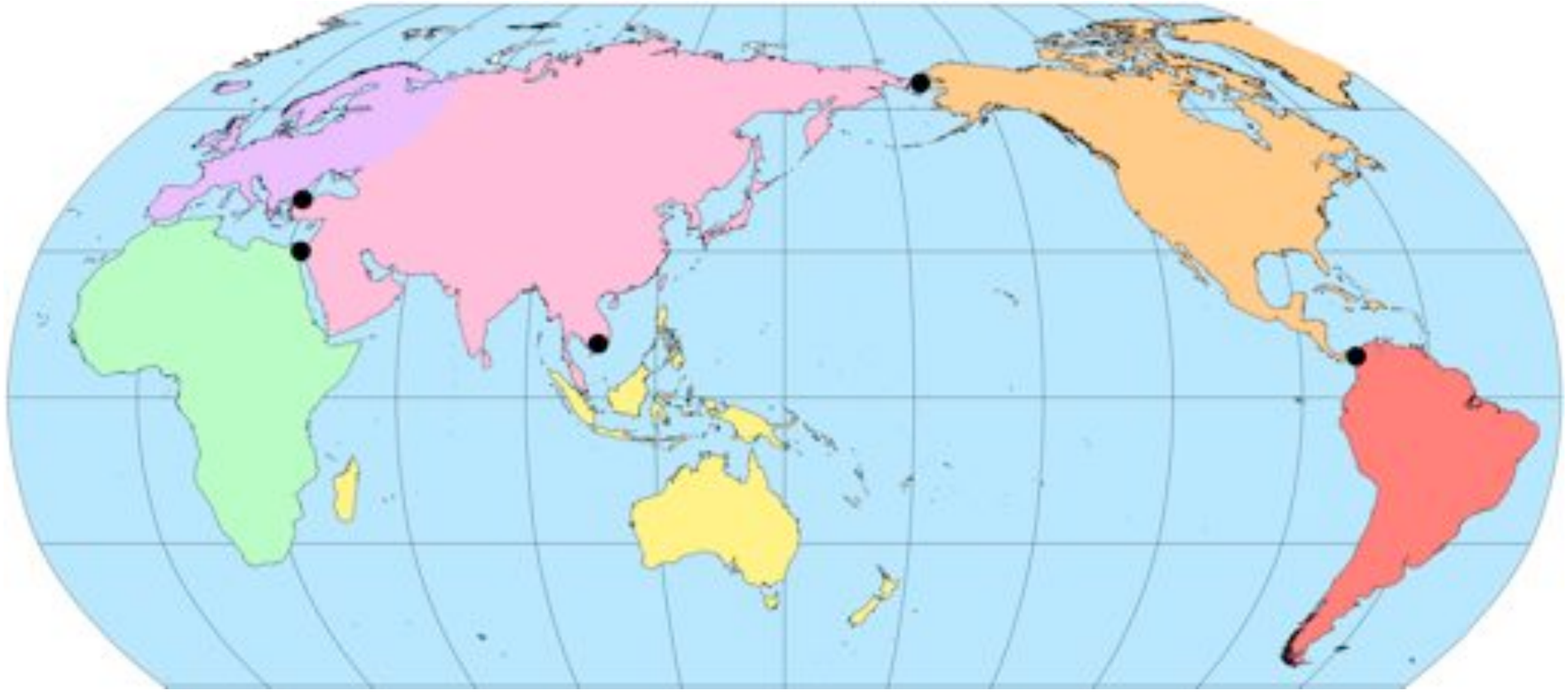


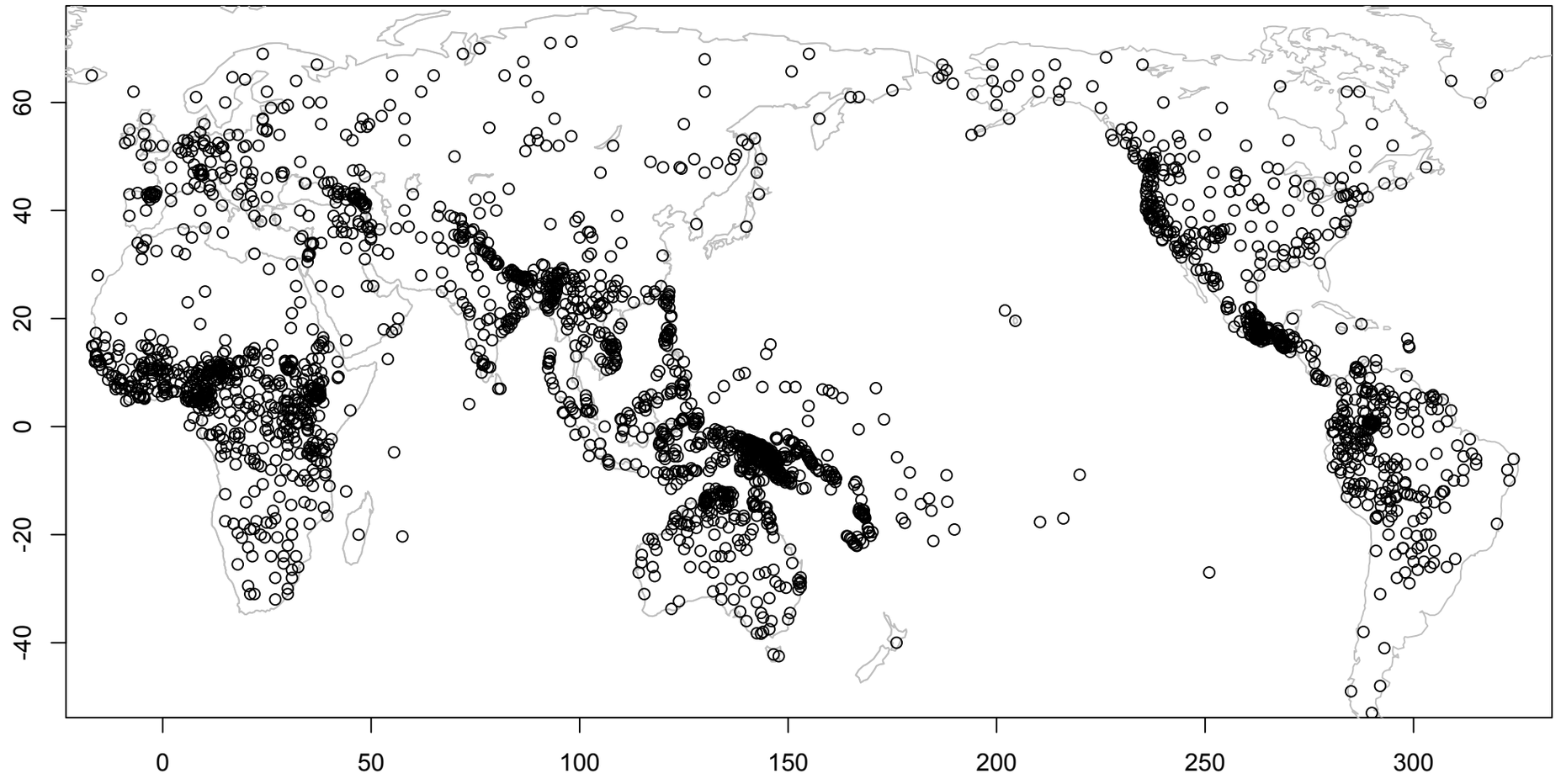

Atkinson's measurement
residuals after regression
1.0
0.5
0.0
-0.5
Africa
SE Asia
Eurasia
N America
S America
NG+Aus
Macroareas


UPSID phoneme counts (log10)
log10 of number of phonemes
1.2
1.4
1.6
1.8
Africa
SE Asia
Eurasia
N America
S America
NG+Aus
Macroareas

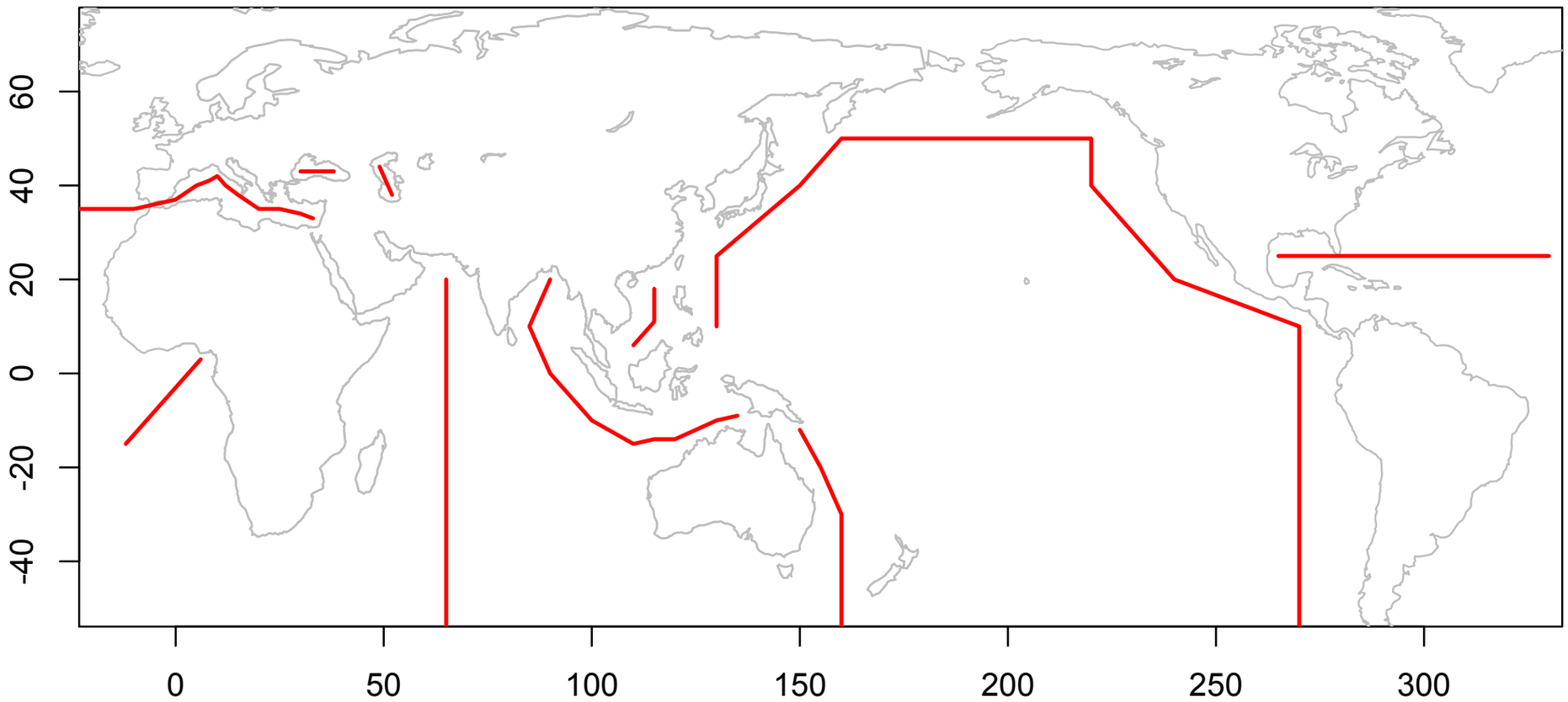
UPSID phoneme counts (log10)

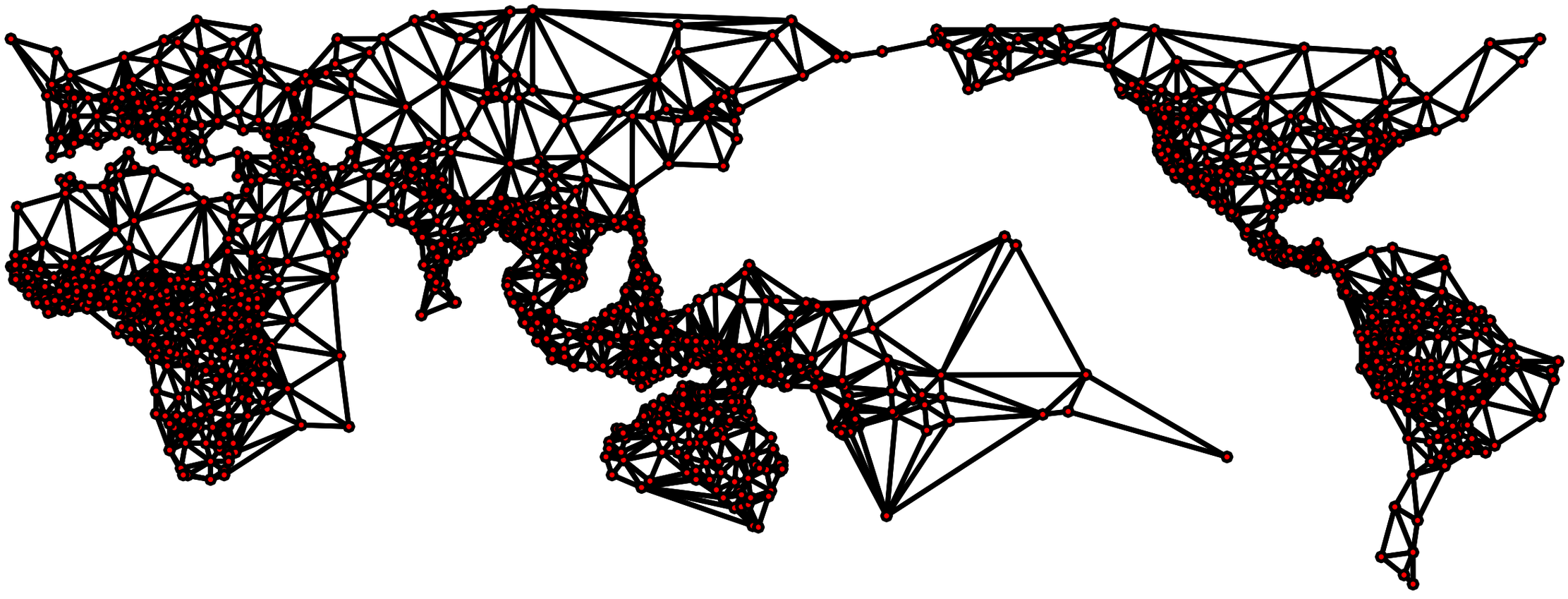


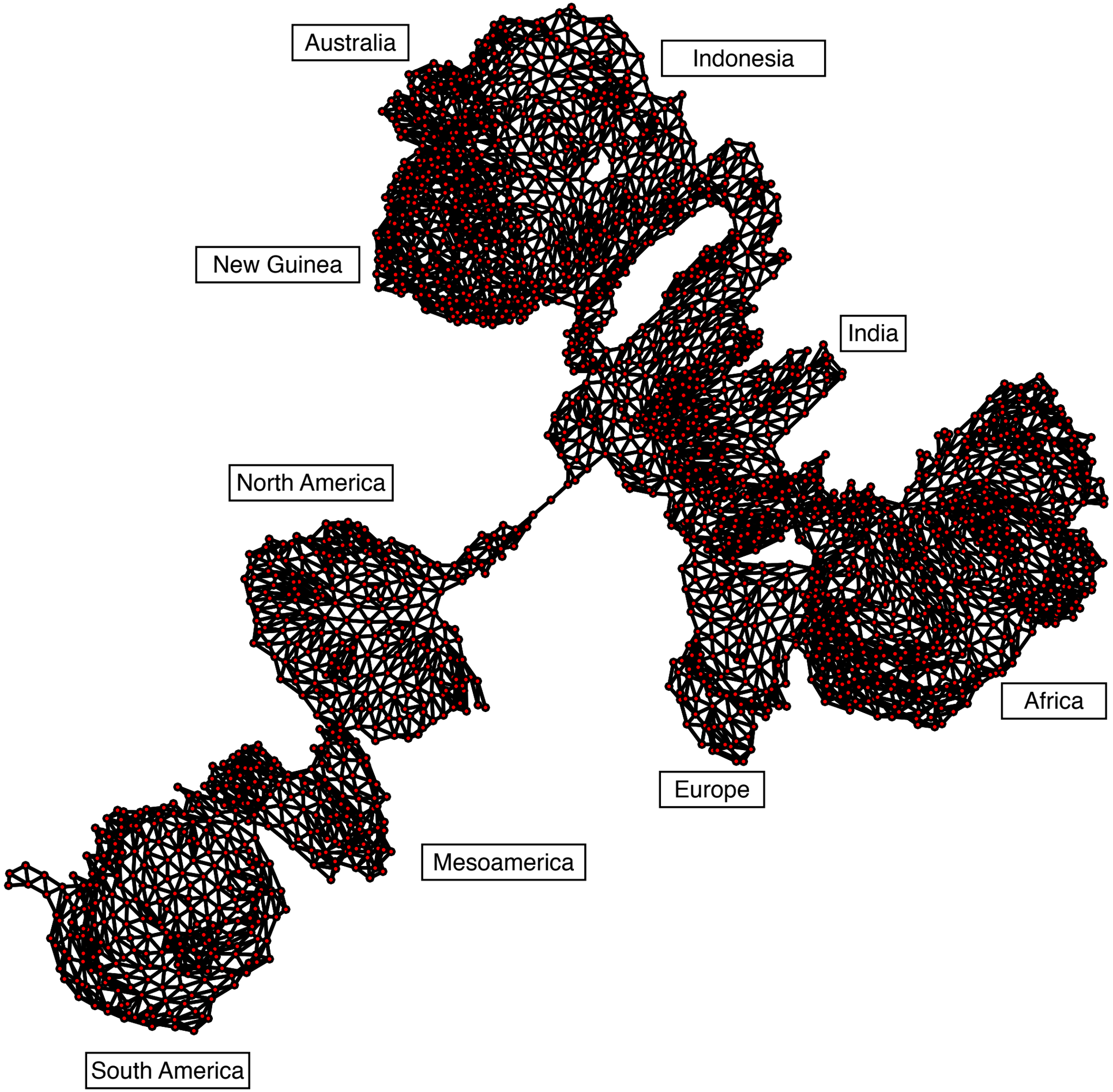



60
40
20
0
-20
-40
0
50
100
150
200
250
300

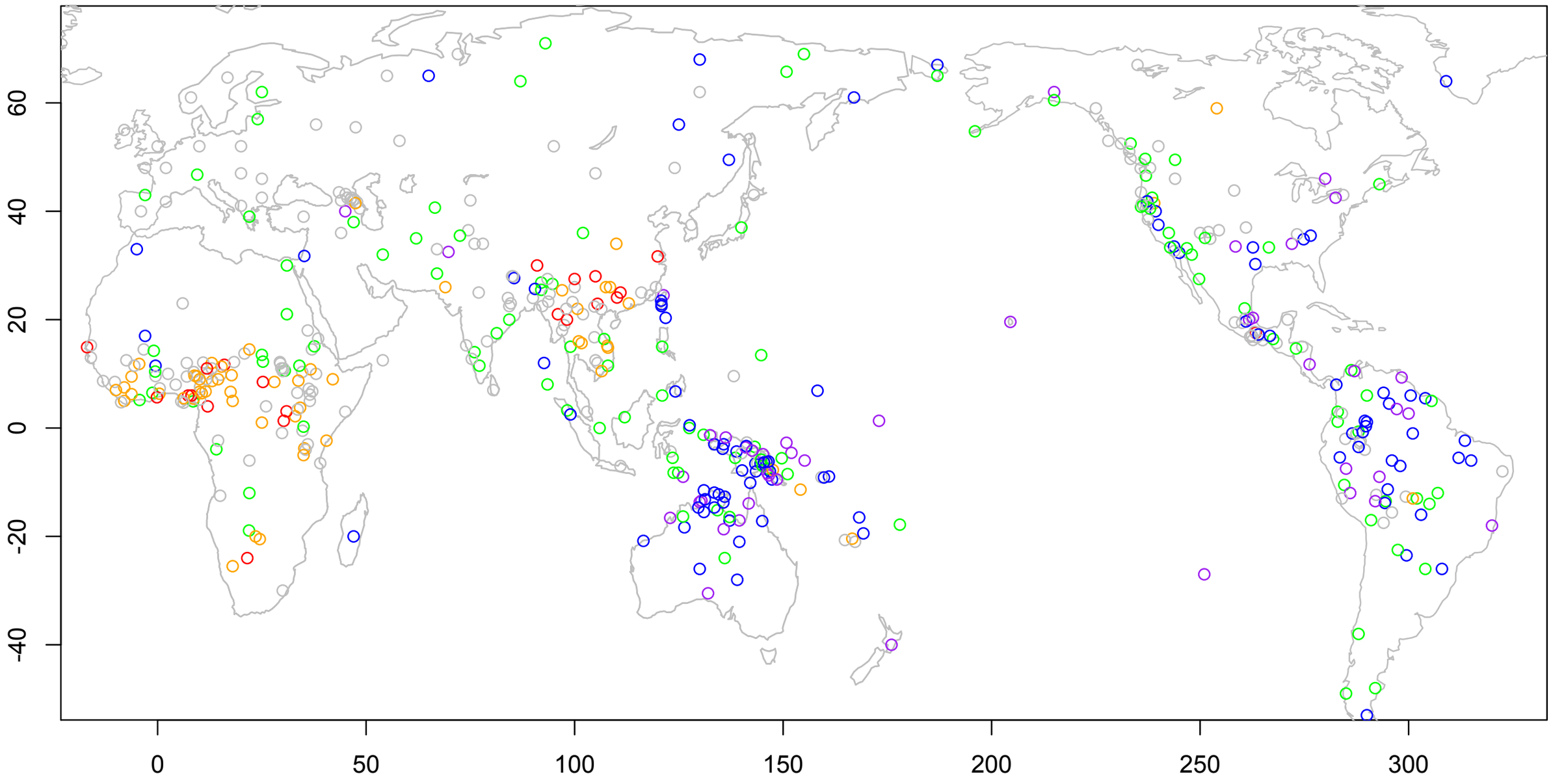
Ancient water boundaries





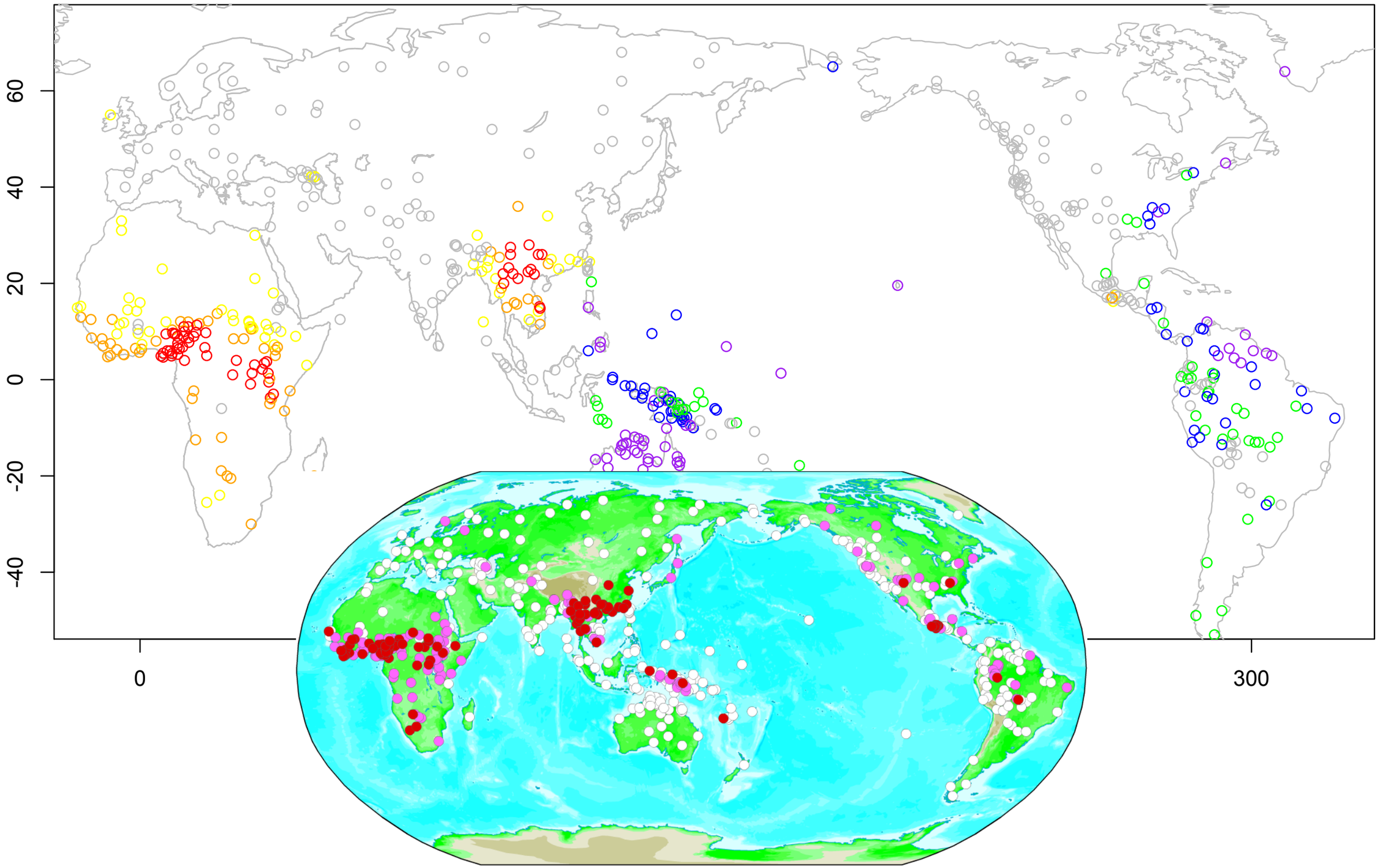

Australia
Indonesia
New Guinea
India
North America
Africa
Europe
Mesoamerica
South America

Atkinson's Assessment of Inventories



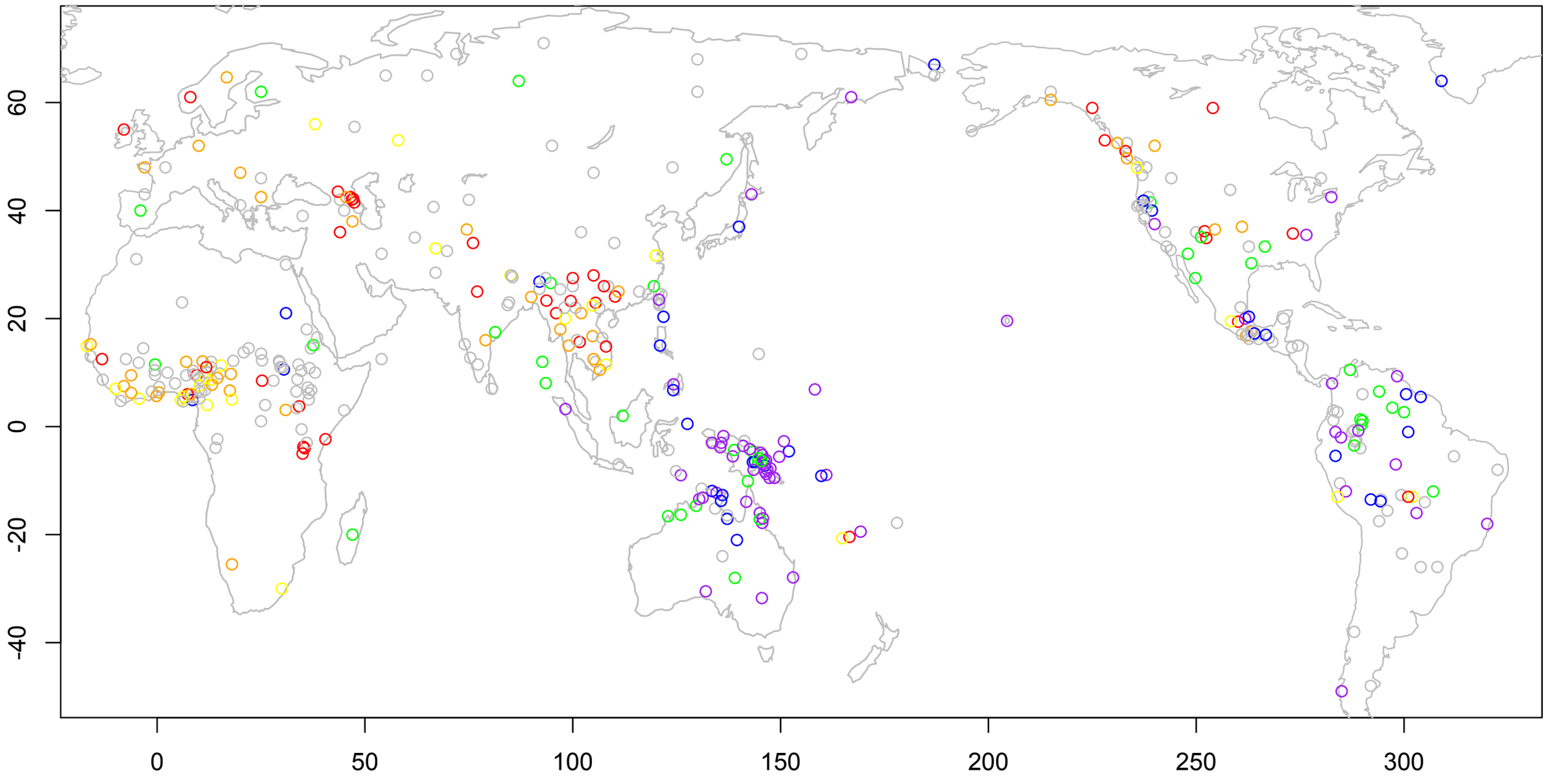
Atkinson's Assessment of Inventories

Raw numbers



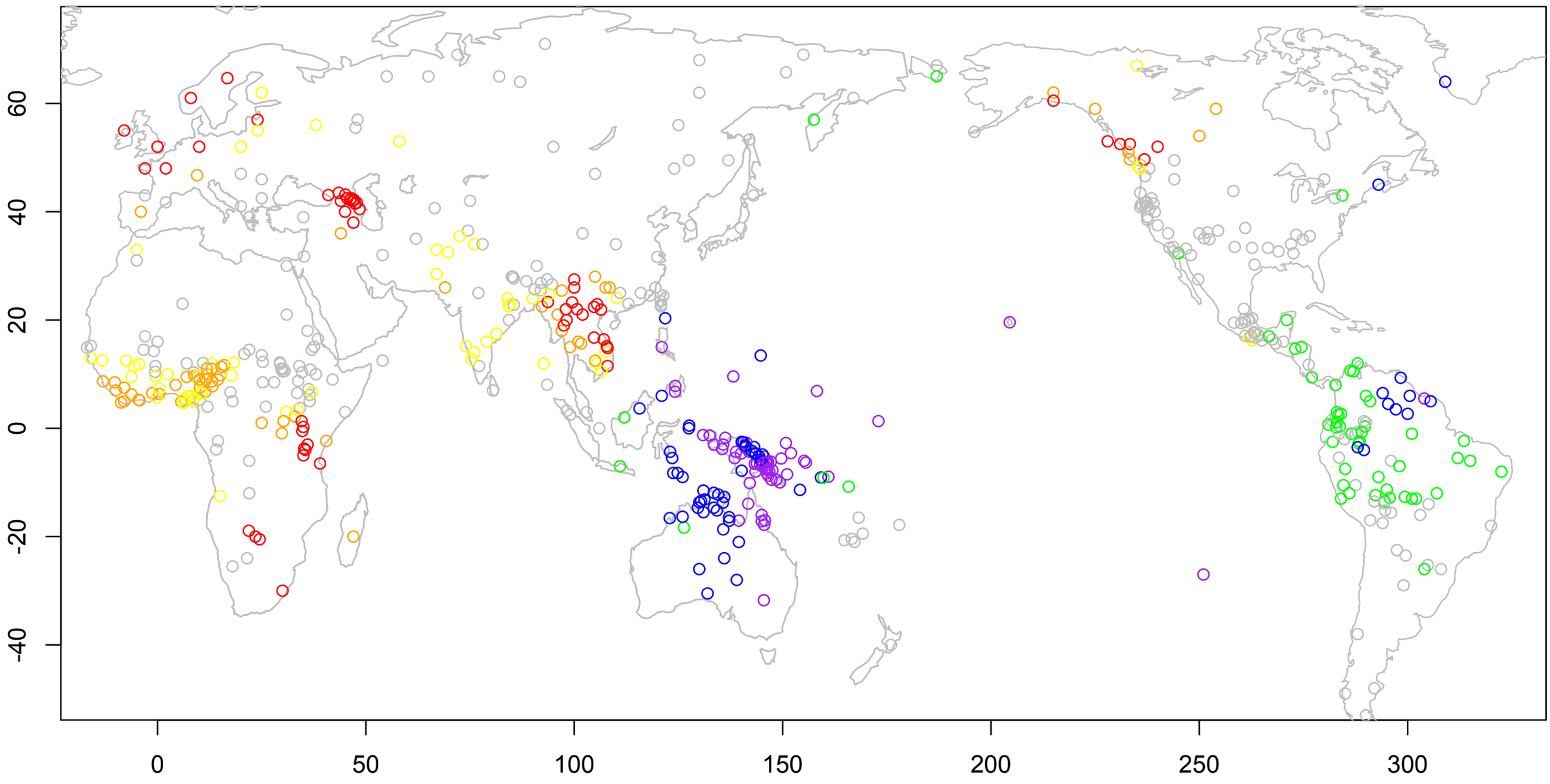
UPSID Inventory size

Raw numbers

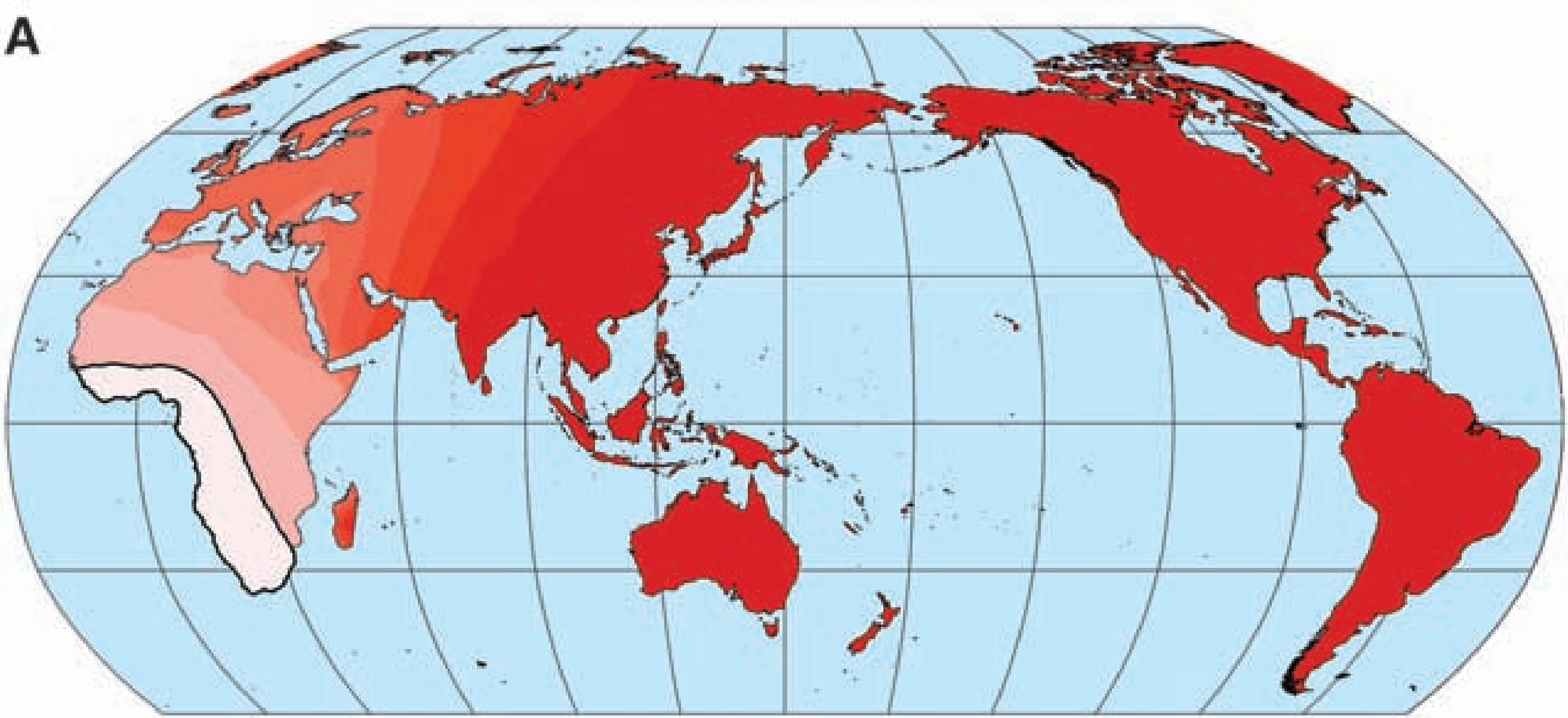


UPSID Inventory size

Raw numbers



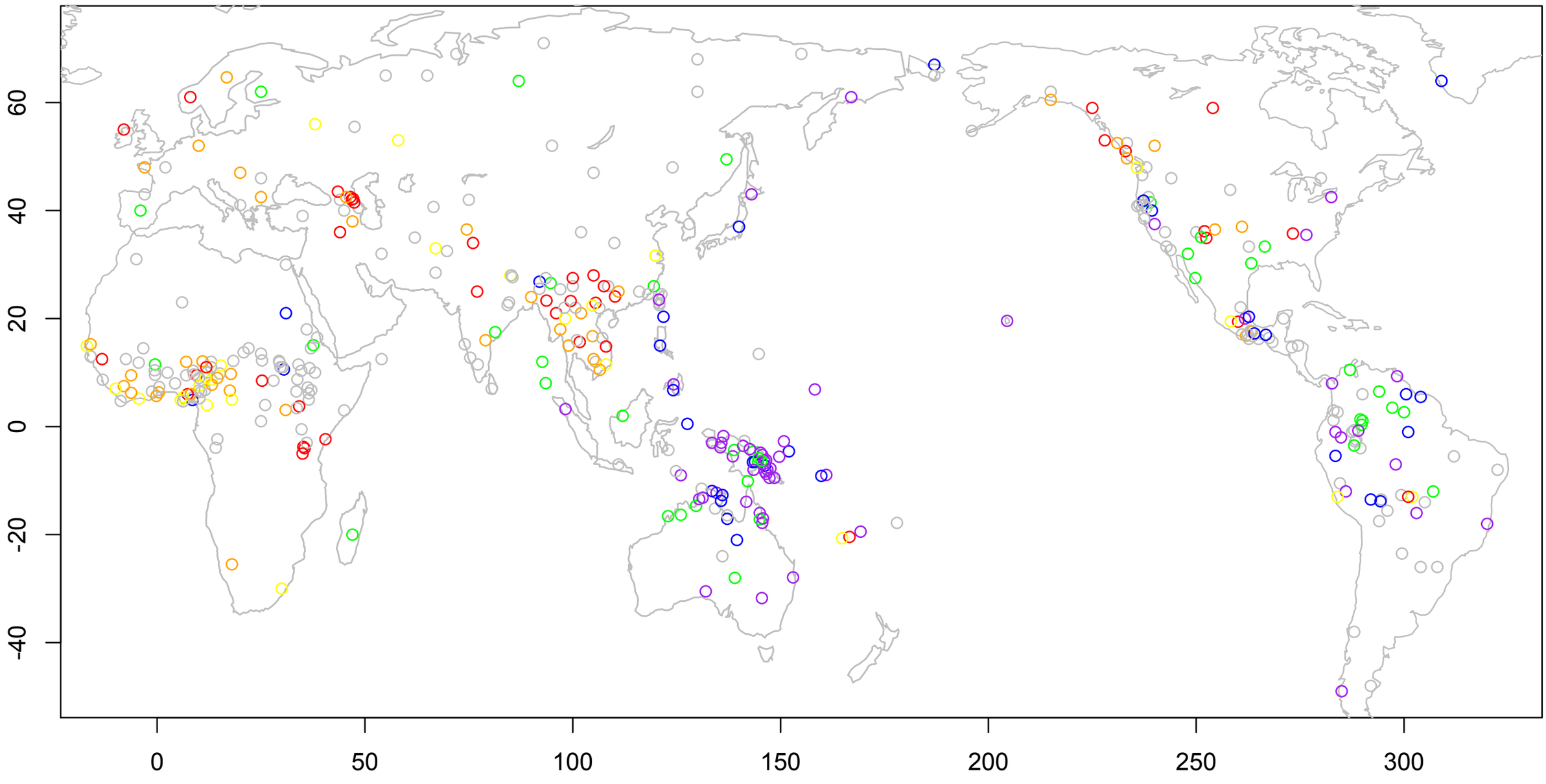
Out-of-Africa Gradient

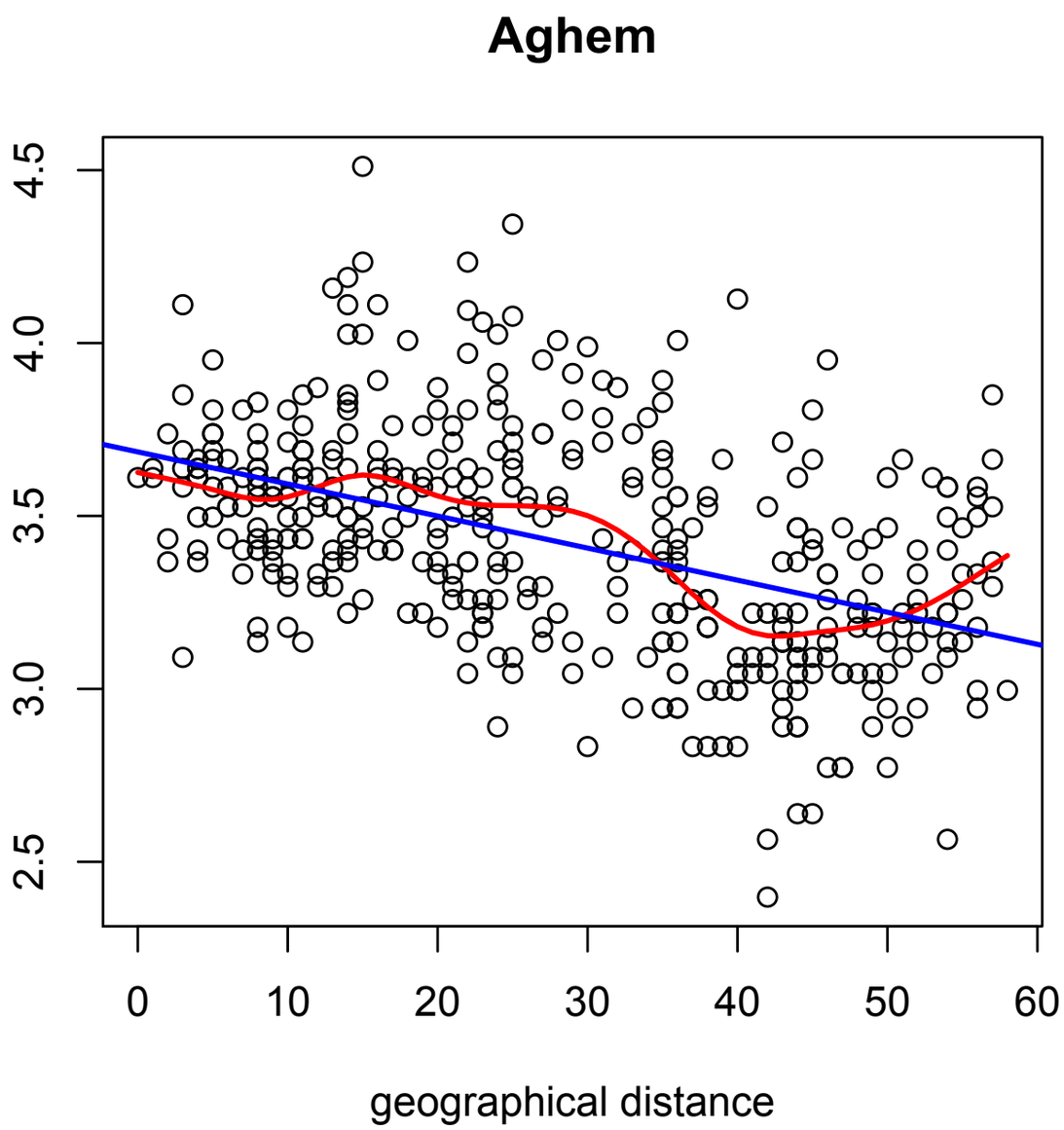


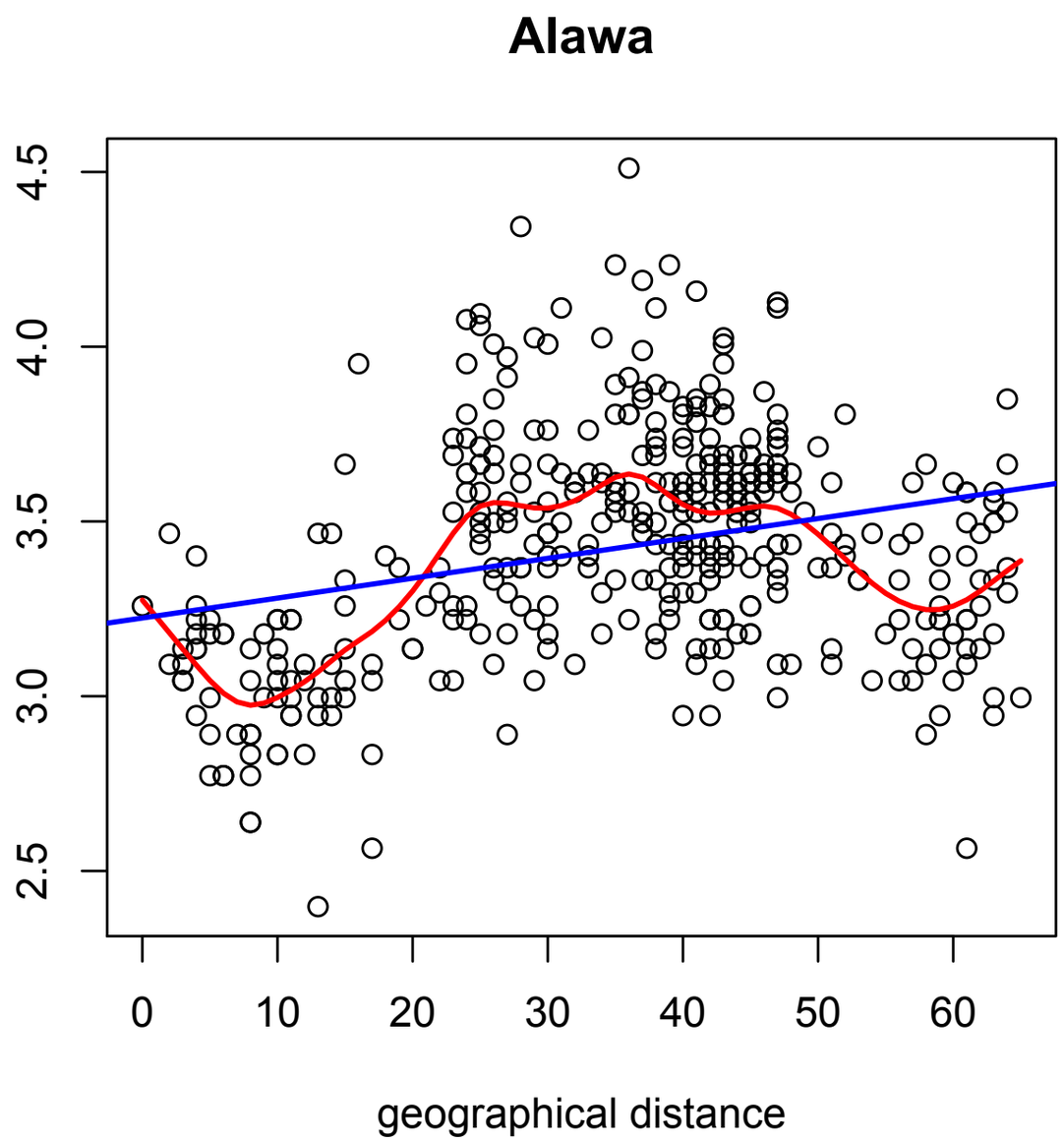
Atkinson, Quentin D. 2011. Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. Science 332: 346-349.

UPSID Inventory size

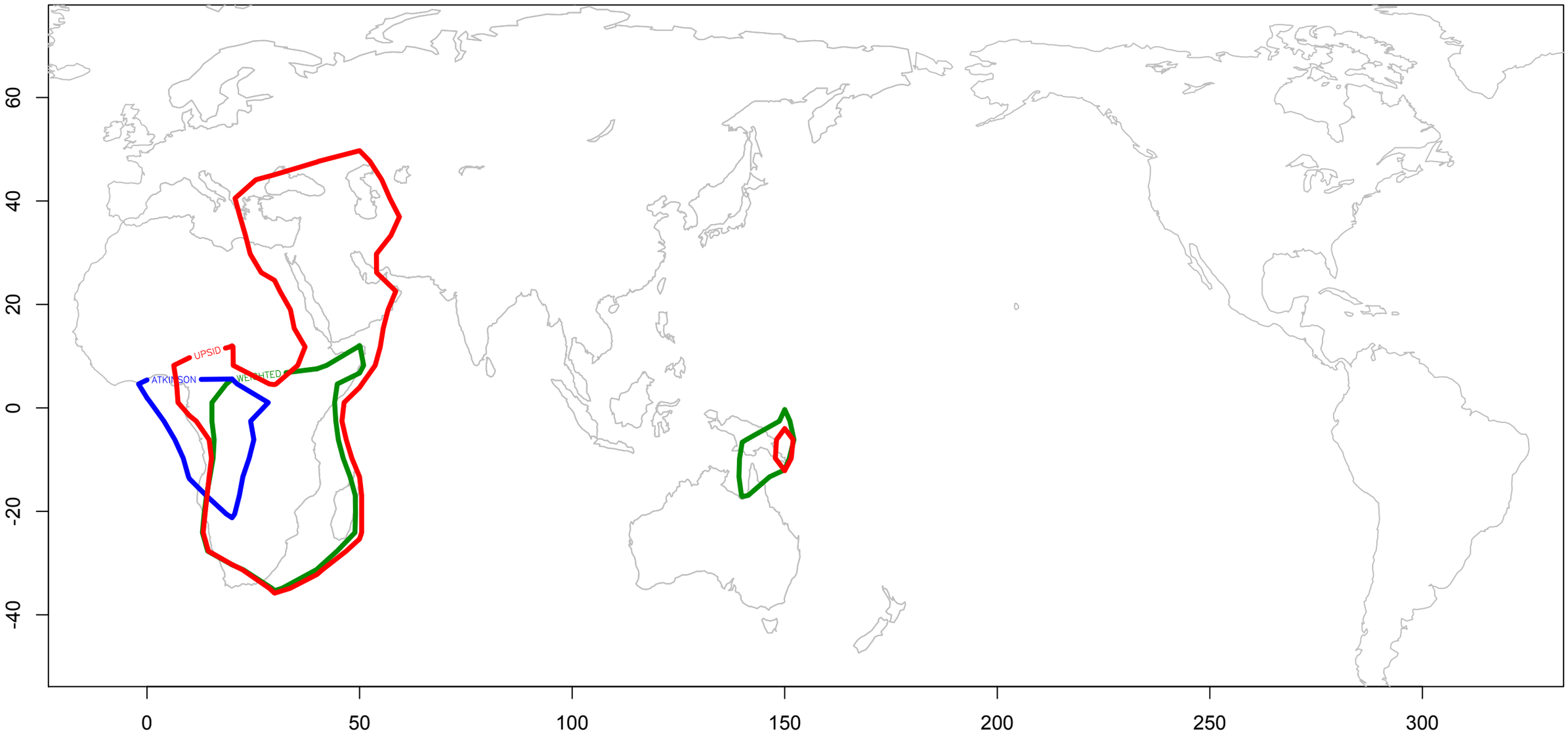
Raw numbers



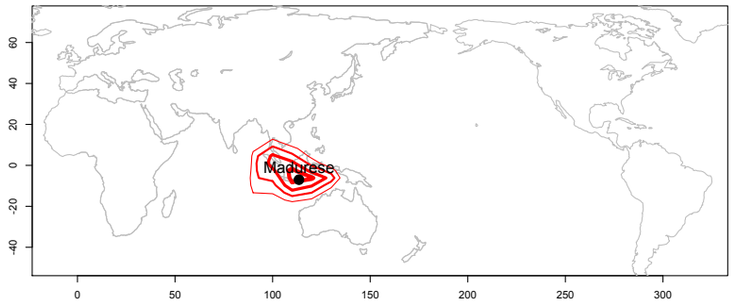

Aghem
4.5
4.0
3.5
3.0
2.5
0
10
20
30
40
50
60
geographical distance


Alawa
4.5
4.0
3.5
3.0
2.5
0
10
20
30
40
50
60
geographical distance

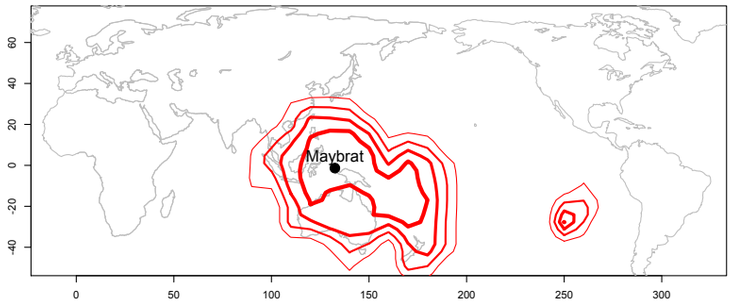
'Origin' of phoneme inventory size



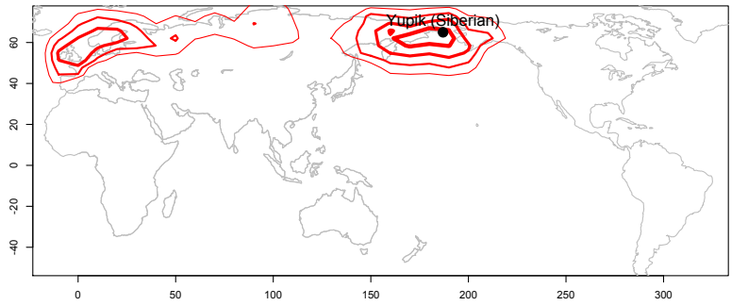
WALS 9



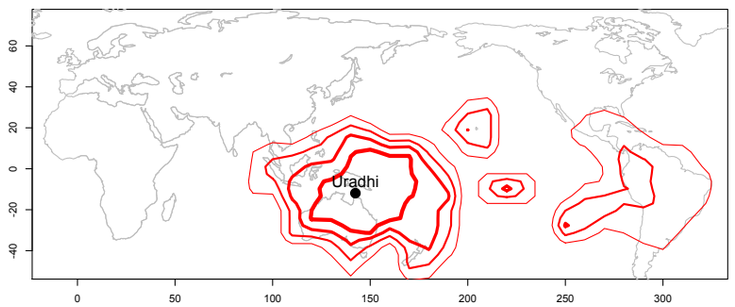
WALS 10



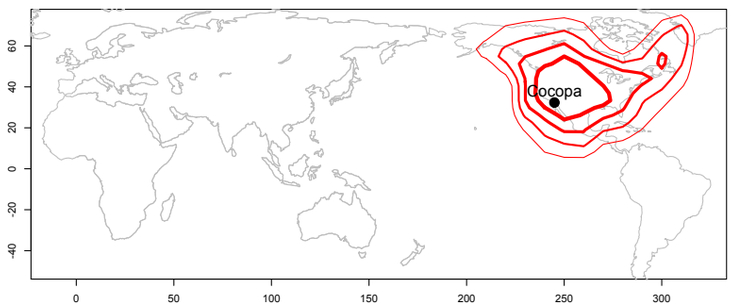
WALS 12



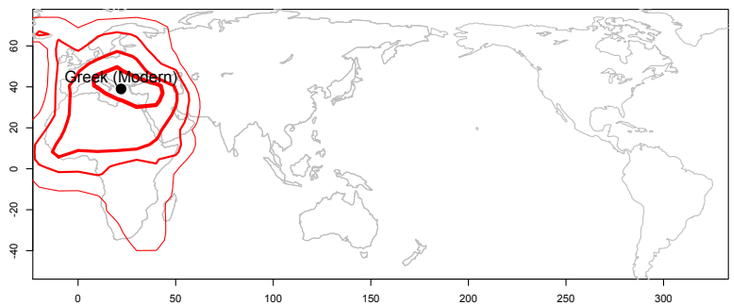
WALS 27



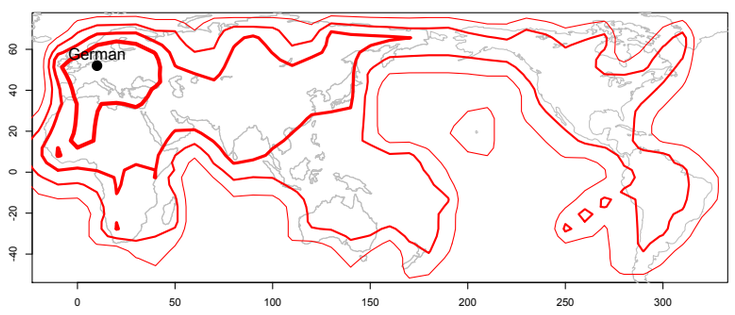
WALS 30



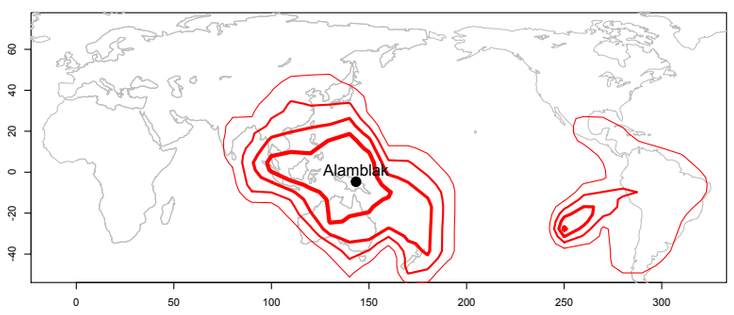
WALS 34



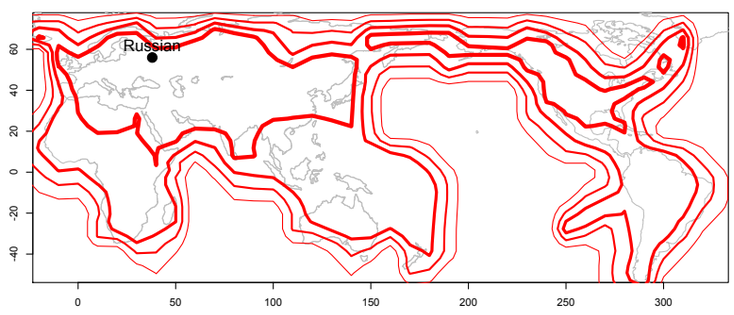
WALS 41



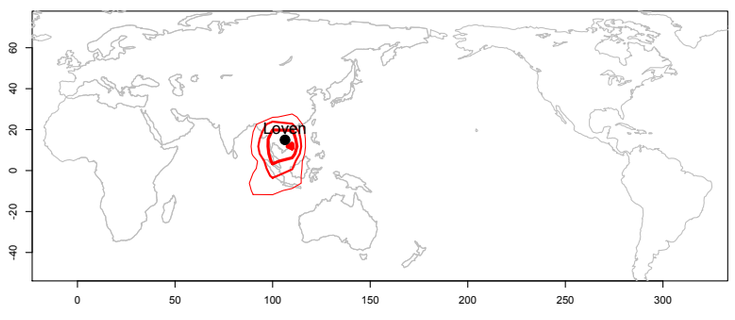
WALS 47



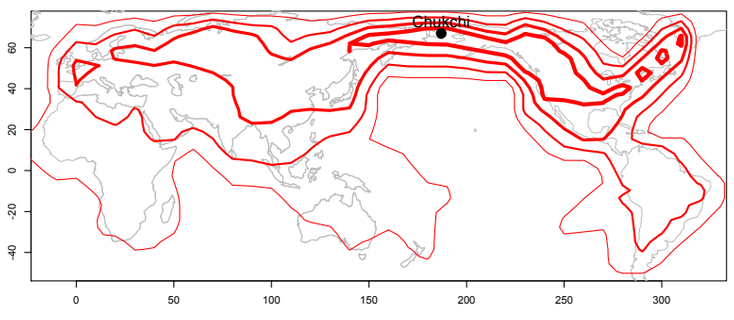
WALS 49



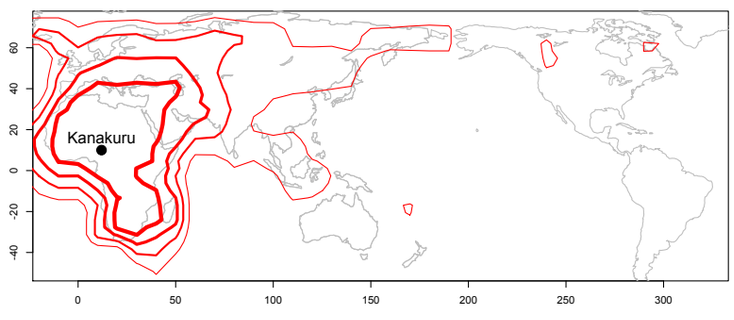
WALS 55



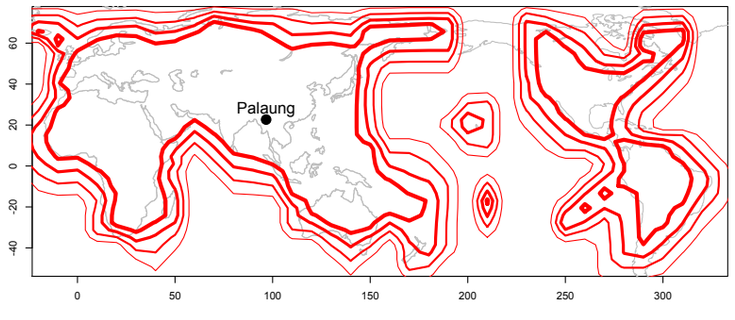
WALS 59



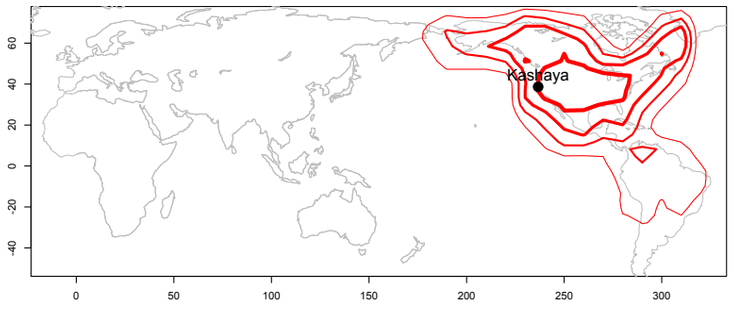
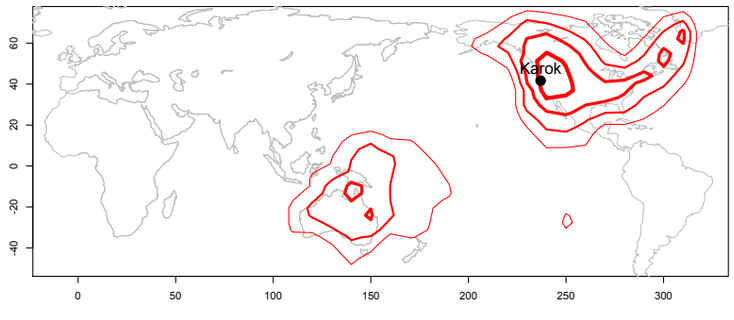
WALS 65



WALS 67



WALS 77



WALS 107