Global profile of language families supports geographical-axis hypothesis

Tom Güldemann
Humboldt University, Berlin & MPISHH
&
Harald Hammarström
Uppsala University

East-West vs North-South Human Migration

- Climate is similar along the East-West axis
- Human subsistence conditions are similar along the East-West axis
  - The same animals and plants can be utilized if you migrate East/West rather than North/South
- A theory popularized by Diamond's (1999) *Guns, Germs and Steel* holds that therefore
  - Human migration is facilitated along the East-West axis compared to the North-South axis

Can this be tested?

East-West Biased Language Families?

- Geographical extensions of language families, e.g., Indo-European, Arawak etc are the result of human migration

Two Questions:

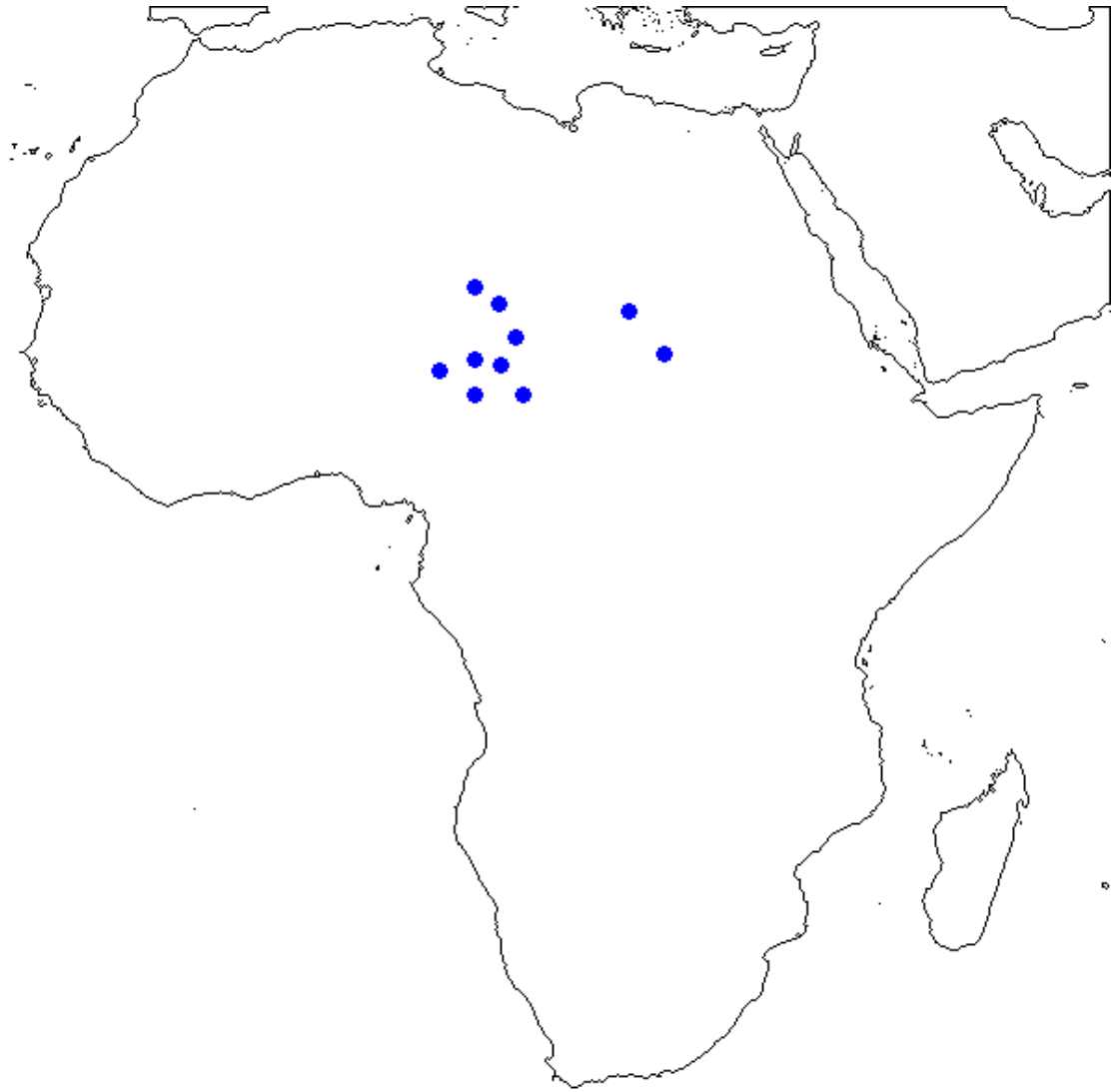
1. Are language families more horizontal than vertical in their geospatial distribution?
2. If yes, can this be plausibly explained by Diamond's theory? (Or is there some other/better explanation?)

Measuring Horizontality (Laitin et al., 2012, 10264)

1. Start with language *centre-point coordinates* and a *language family classification*
2. Estimate the shape of a language family as the joined *Voronoi-extension of its member languages* (excluding water)
3. *Slice* the shape of the family (resolution 1/8th of a lat/long degree) on the *x-axis* and on the *y-axis*
4. Take the ratio between the length of the *average* x-axis ($\overline{E-W}$) and the length of the *average* y-axis ($\overline{N-S}$)
5. The *log* of this ratio is the *horizontality* of the family

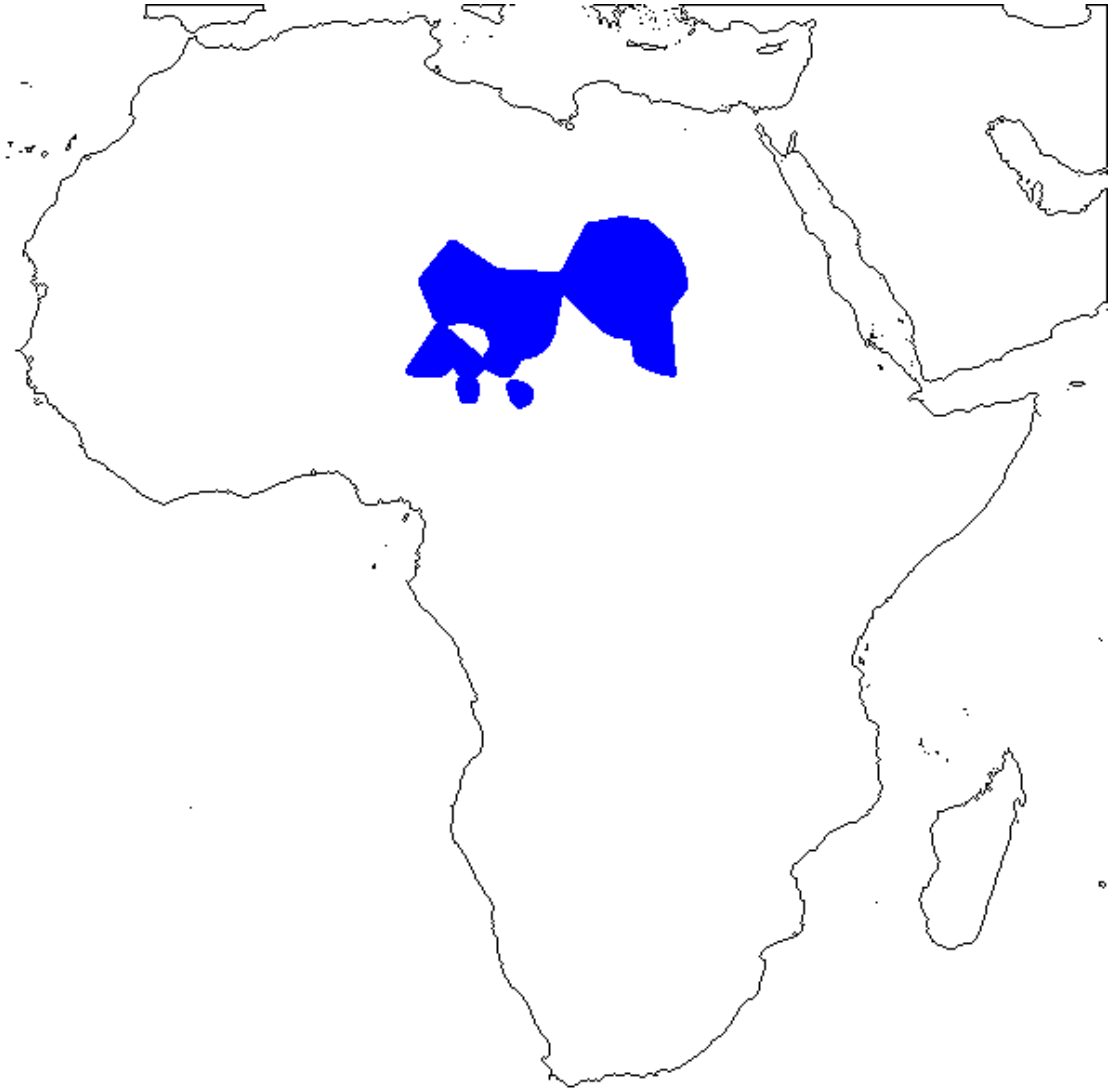
$$HOR(f) = \log \overline{E-W}/\overline{N-S}$$

Measuring Horizontality: The Saharan Family #1



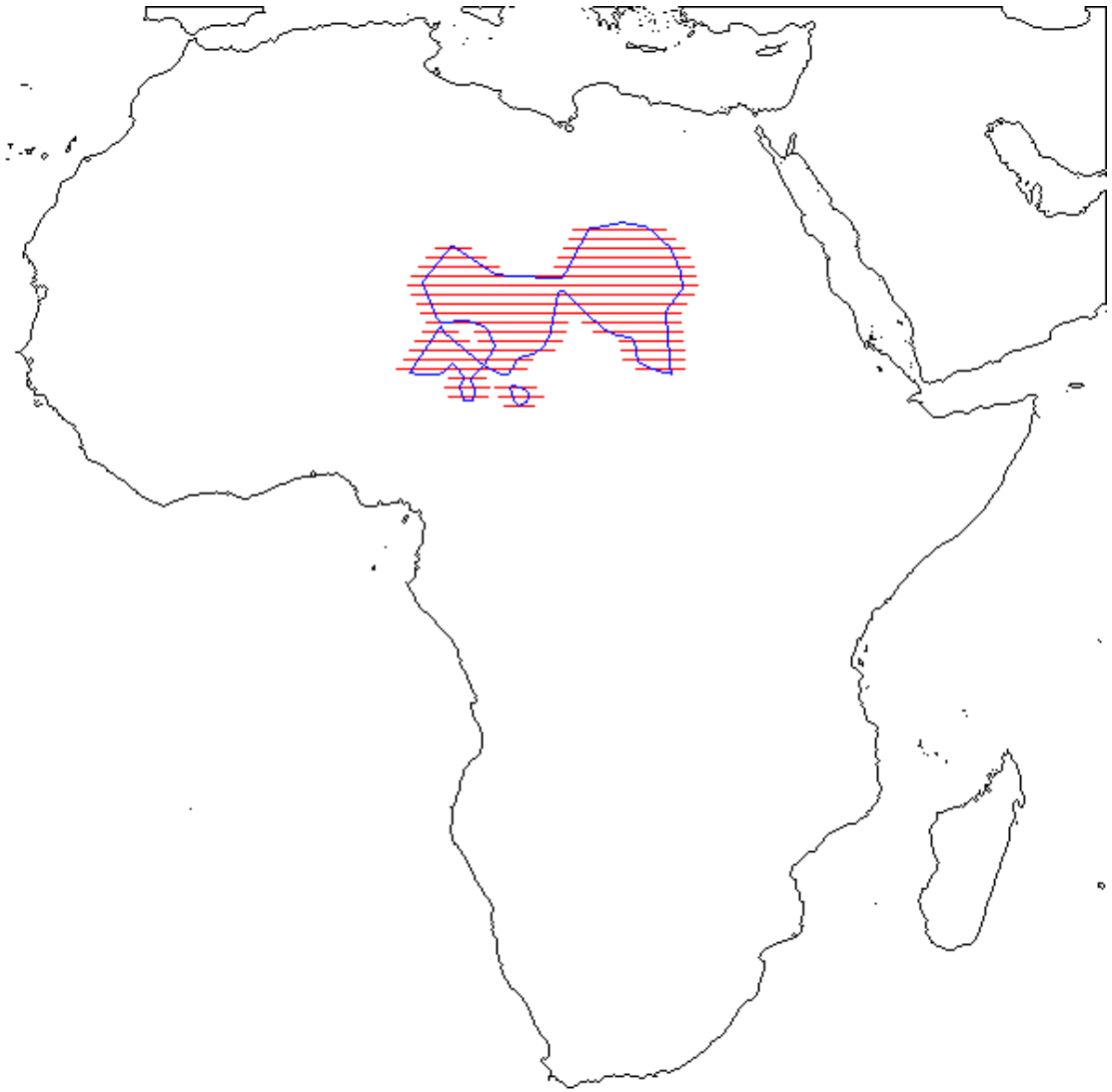
Centre-point coordinates

Measuring Horizontality: The Saharan Family #2



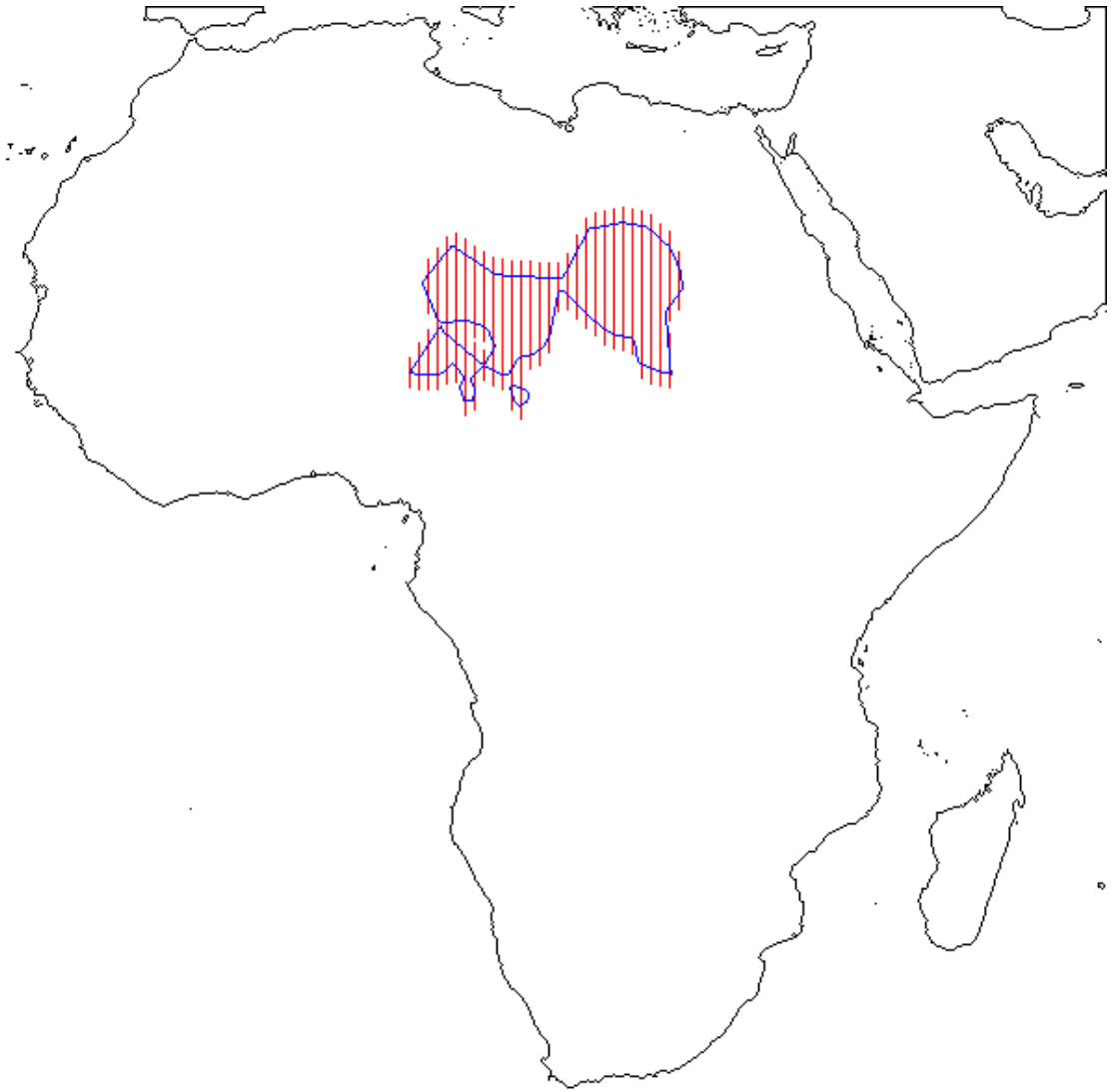
Shape from Voronoi-regions

Measuring Horizontality: The Saharan Family #3



$\overline{E-W} \approx 909.9$ kms

Measuring Horizontality: The Saharan Family #4



$\overline{N-S} \approx 647.4$ kms

Measuring Horizontality: The Saharan Family #5

$$HOR(Saharan) = \log \overline{E-W}/\overline{N-S} = \log \frac{909.9}{647.4} \approx 0.34$$

- So, Saharan has an axis bias (> 0) in the horizontal direction
- Is this a more general phenomenon?

World Dataset

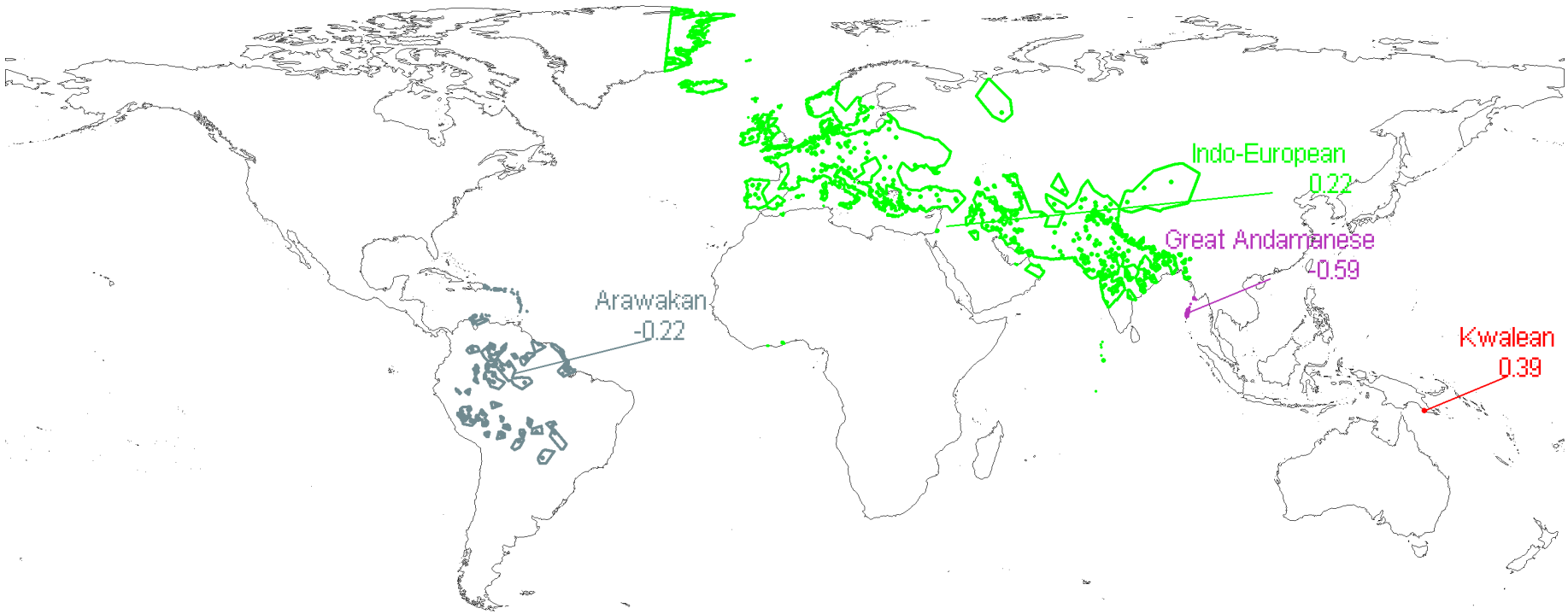
Data from Glottolog 3.3 (Available `http://glottolog.org`)

- 7454 iso-639-3-like languages
- Centrepoint coordinates for all languages
- 420 Lineages (Families + Isolates)

Notes:

- Postindustrial migrations rolled back
- Languages who do not maintain a territory are not counted, i.e., those who live in symbiosis, itinerantly or dispersed among speakers of some other language which maintains the territory

Some Example Families and Horizontality



Families and Horizontality

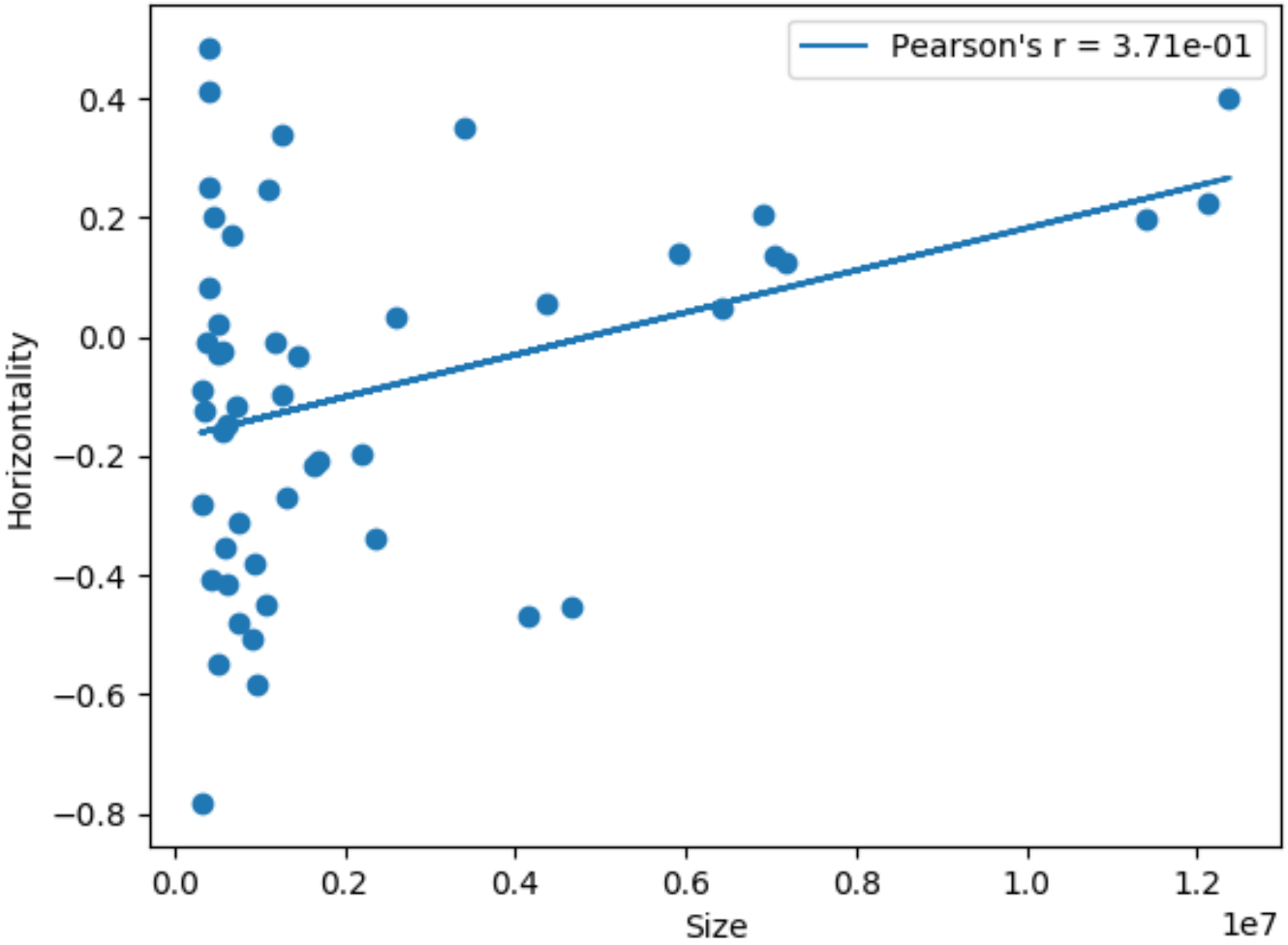
- Some families are very horizontal, e.g.
  - Kwalean 0.39
  - ...
  - Indo-European 0.22
  - ...
- Some families are very vertical, e.g.
  - Great Andamanese -0.59
  - ...
  - Arawakan -0.22
  - ...
- A lot of families are neutral

The mean of all is -0.044 – close to neutral!

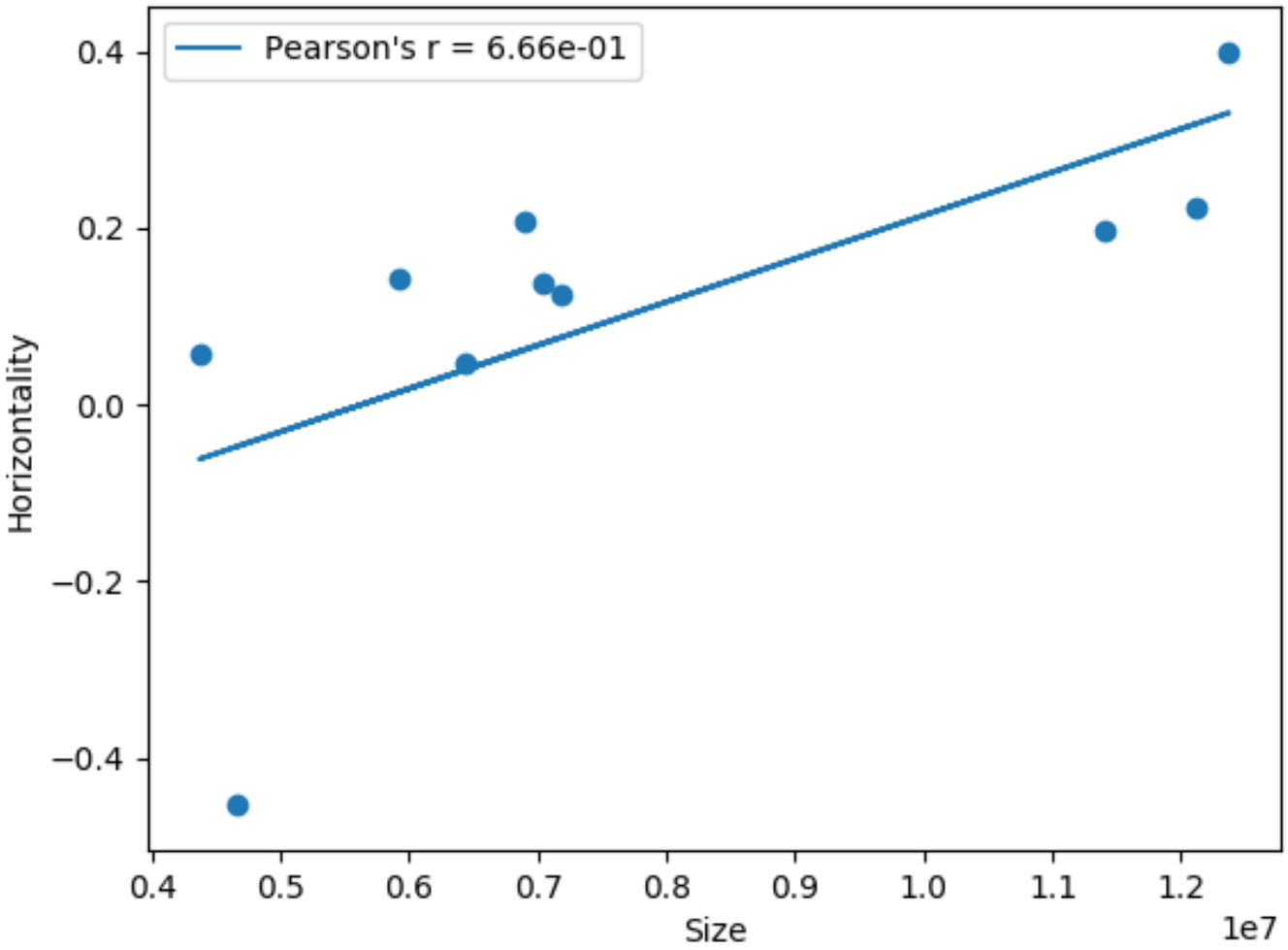
It Doesn't Work, Or ...?

- Most language families are very small
- Out of 420 families only 44 ($\sim$ 10%) are bigger than Sweden
- The small families will dominate the global mean
- On short distances, climate is "the same" in either east-west or north-south direction
- What if we look at only big families?
  - Size of a family: Simply the size of its shape (calculated at 1/8 of a degree)
  - Example: Saharan is 1 264 682.9 km2
- The bigger the family, the more horizontal?

Size vs Horizontality: 50 Largest



Size vs Horizontality: 10 Largest



Size-Horizontality Correlation

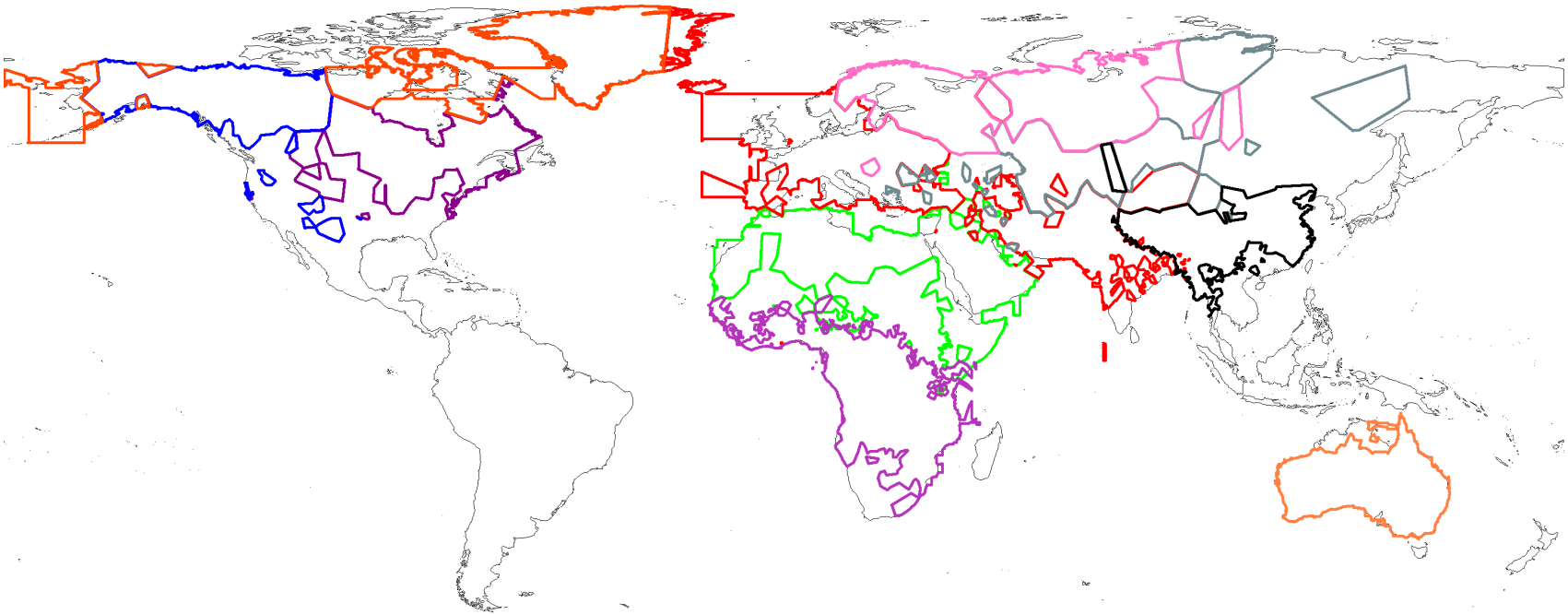
Linear regression:

- Taking only the 50 largest families gives a modest $r \approx 0.37$ but highly significant $p < .001$ correlation
- Taking only the 10 largest families gives a stronger $r \approx 0.66$ but somewhat less significant $p < 0.01$ correlation

The 10 Largest Families in Numbers

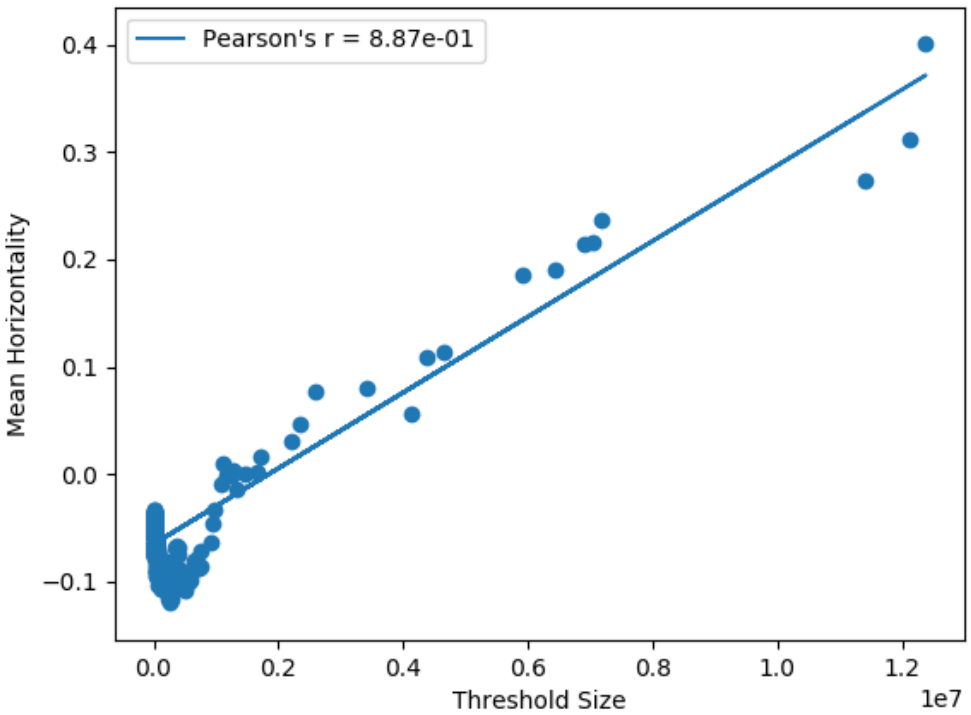
Family	Size	Horizontality	# lgs
Afro-Asiatic	12374703.72	0.40	370
Indo-European	12114637.25	0.22	478
Atlantic-Congo	11403296.52	0.20	1434
Turkic	7179819.42	0.12	43
Uralic	7044678.11	0.14	48
Pama-Nyungan	6894737.95	0.21	248
Sino-Tibetan	6432281.60	0.05	488
Algic	5922007.24	0.14	45
Athabaskan-Eyak-Tlingit	4662996.31	-0.45	45
Eskimo-Aleut	4369194.26	0.06	12

The 10 Largest Families Graphically



Afro-Asiatic, Indo-European, Atlantic-Congo, Turkic, Uralic, Pama-Nyungan, Sino-Tibetan, Algic, Athabaskan-Eyak-Tlingit, Eskimo-Aleut

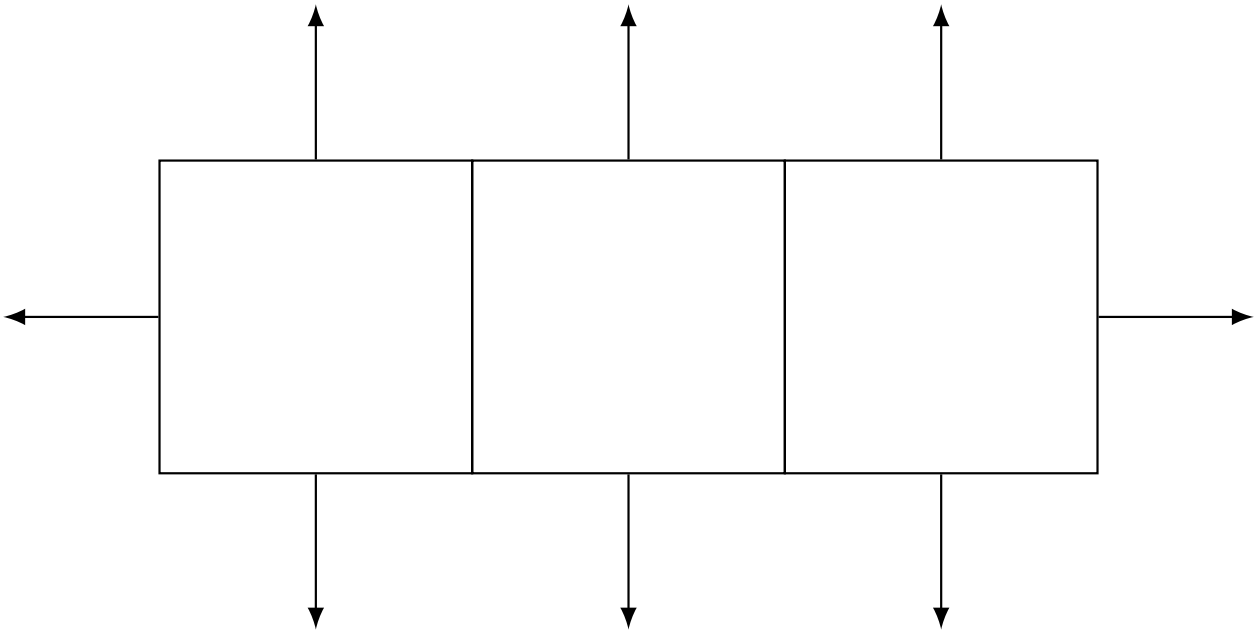
Threshold Size vs Mean Horizontality



- If we filter out small families successively, taking the mean of the
  - the 420 biggest families
  - the 419 biggest families
  - ...
  - 2 biggest families
  - 1 biggest family
- The mean of the remaining families increases smoothly

Growth Dynamics

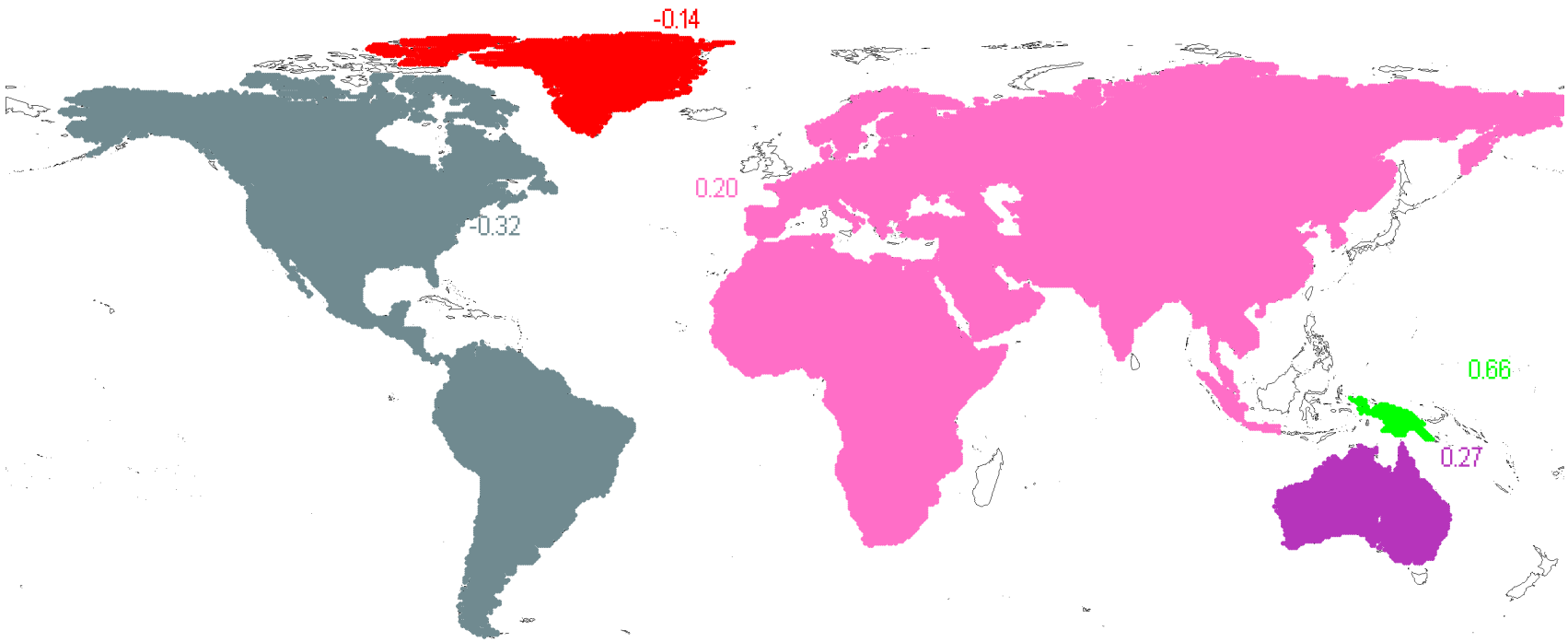
- We do not know the actual growth dynamics but
- If the model is correct, the measure we use *underestimates* the actual horizontal bias
  - If a family already has a horizontal profile
  - There is a broader frontier for expansion in N-S than E-W direction
  - Random walks out of the family would favour expansions N-S over E-W
  - Horizontal bias is needed to *maintain* the horizontal profile
  - And we observe that it even increases its horizontal profile if it expands



Climate or Landmass Bias?

Landmass bias:

- Maybe the landmass itself is horizontally biased?
- People do not like to live near the poles, making the playground more horizontal



Testing Climate vs Landmass this

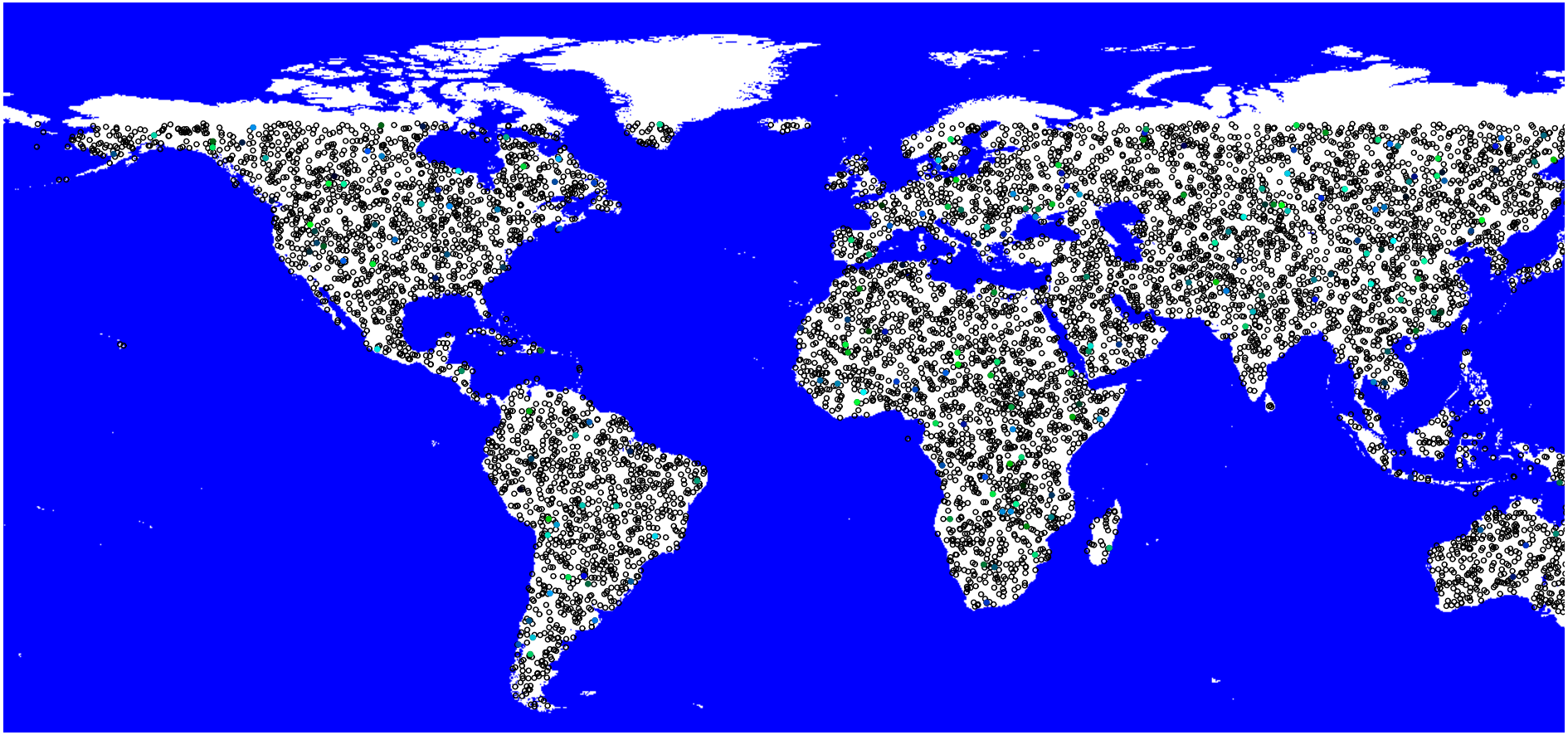
- Simulate *random* family growth on the actual landmass
- If the horizontality found in the actual world beats the horizontality found in 100 random worlds, we conclude the landmass bias is not sufficient to account for the horizontality, leaving Diamond's theory the winner

Randomly Generated Families

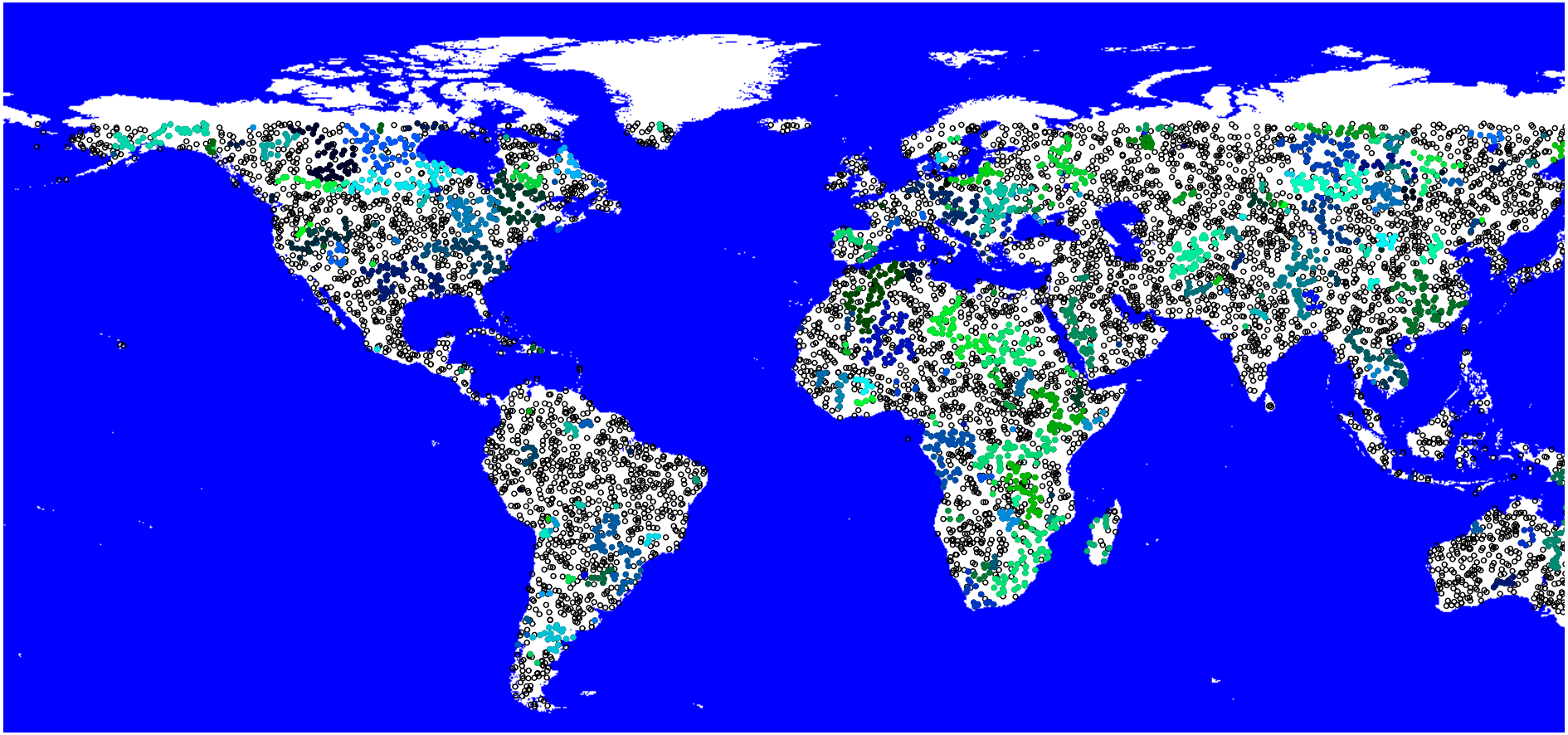
Simplistic Model:

- Assume that people do not like to live south of 50 degrees S nor north of 65 degrees N
- Randomly distribute ca. 8000 language locations on the landmass (in 65N - 50S)
- Randomly pick 420 homelands out of the 8000 possibilities
- Each of the 420 homelands is destined to grow to the number of languages corresponding to a specific family in the real world
- Iterate until growing is saturated (= the real world number of languages have been reached):
  - for each family, pick the geographically nearest (free) language (in any direction) and incorporate it

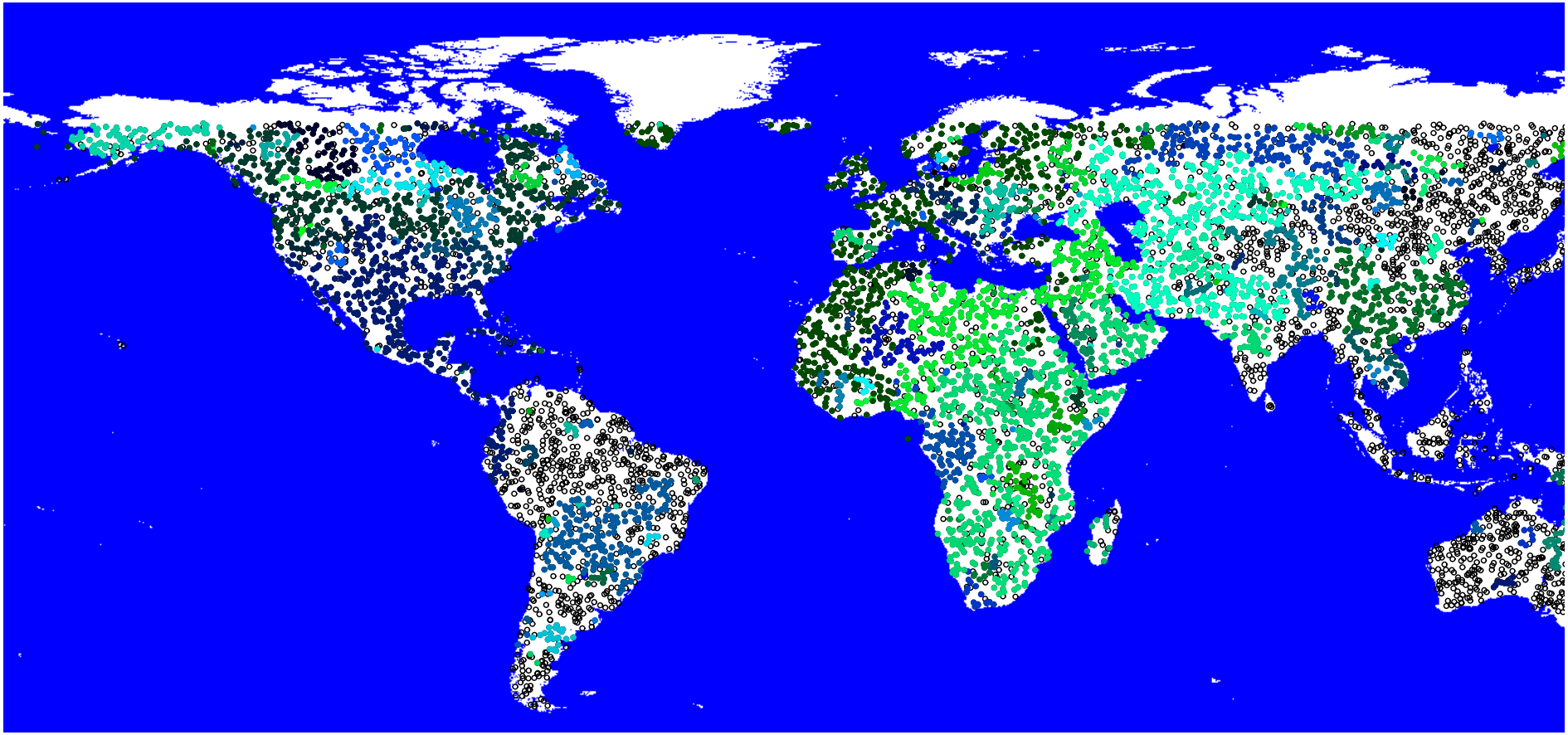
Example: Iteration 0



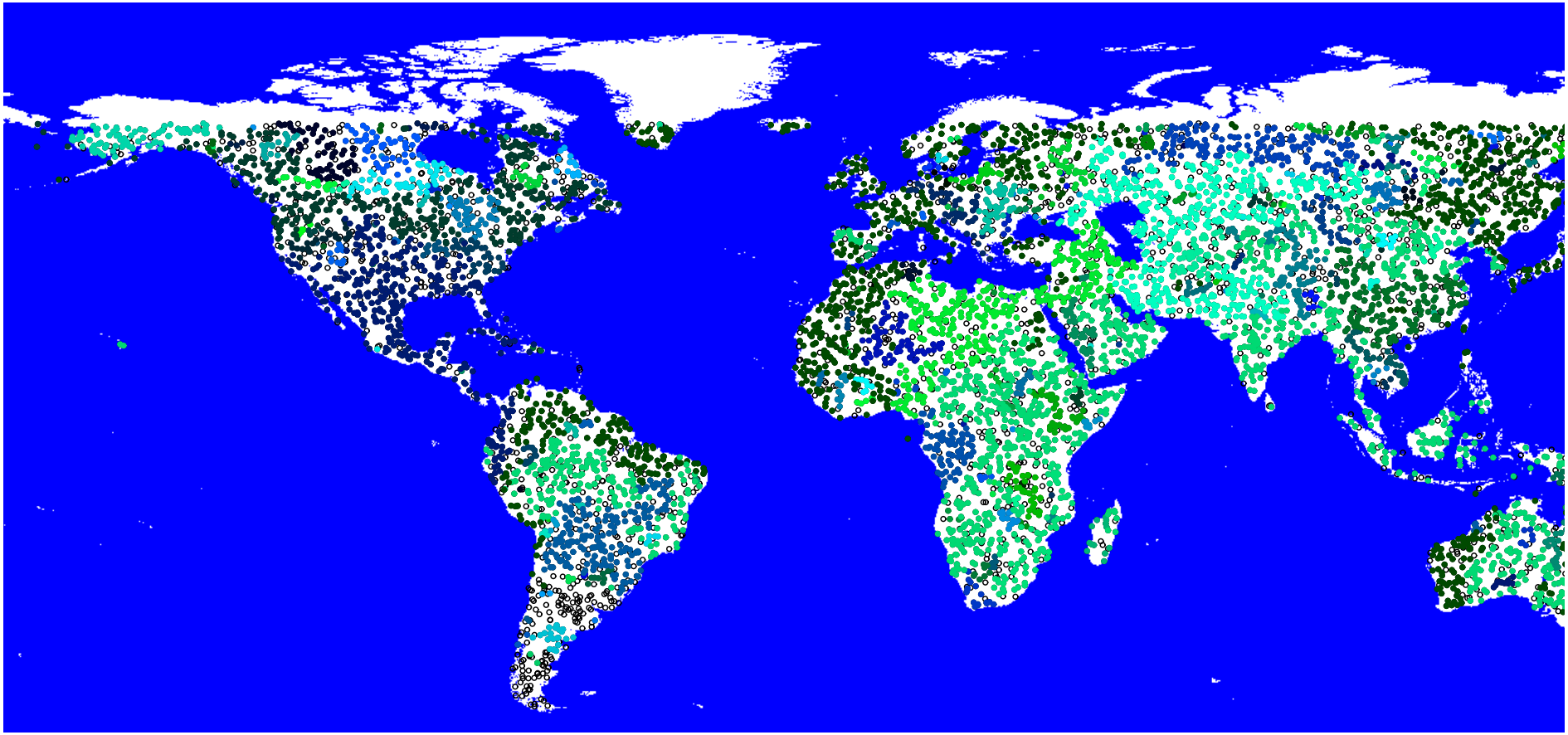
Example: Iteration 50



Example: Iteration 500



Example: Iteration 1434 (Final)



Random Worlds and Horizontality

- All (small and large) families in 100 random worlds:
  - The mean horizontality for all families in one random world ranged -0.25 – 0.10
  - Mean across all random worlds is -0.071 (real world has -0.044)
- The 10 largest families in 100 random worlds:
  - The mean horizontality for the 10 largest families in one random world ranged 0.042 – 0.112
  - The mean across all random worlds is 0.064 (real world has 0.108)

Large Families in Random Worlds

- Plot size versus horizonality and fit a regression line to the datapoints corresponding to the 50/10 largest families.
- Random worlds do not exhibit the strong size-horizontality relationship exhibited in the real world
  - The fit (Pearson's r) for the regression line for the random worlds is smaller than the real world whether for the 50 largest ($p < .01$) or the 10 largest ($p < .01$)
  - The slope of the regression line for the random worlds is less steep than the real world whether for the 50 largest ($p < .01$) or the 10 largest ($p < .01$)
  - In the random worlds, the regression line may start somewhat higher or lower in horizontality than in the real world, but in either case, always proceeds with a less steep slope.

Related Things We Checked

- Does it work for subfamilies?

  $=>$ *Yes, though subfamilies tend to be smaller*

- What about other family classifications?

  $=>$ *Results highly similar for the Ethnologue 21ed classification*

- Does it work if the shape of a family is approximated by its rectangular bounding box?

  $=>$ *It works less well*

- Does it work if the size of a family is measured in the number of languages rather than geospatial size?

  $=>$ *It works much less well, it really needs to be geospatial*

- Is there a difference between language families with hunter-gatherers vs agricultural vs pastoral peoples?

  $=>$ *No, though not many large non-agricultural families*

Discussion & Conclusions

- The general idea seems to work
- Are there more predictions?
  - Linguistic
    - Large convergence areas should also have a horizontal bias (cf. Güldemann 2010, 2018 for Africa)
    - "Longitude spread constraint" as counterpart of "latitude spread potential": When populations do move North-South over long distances, they are forced to interact more with local populations, which should yield more linguistic change (cf. Güldemann and Hammarström in press)
  - Non-linguistic
    - Non-linguistic features of human populations (cultural anthropology, genetics, etc.)
    - Features of non-human populations

Thank you

Diamond, J. (1997). *Guns, germs and steel: the fates of human societies*. London: Cape.

Güldemann, T. (2010). "sprachraum" and geography: Linguistic macro-areas in africa. In Lameli, A., Kehrein, R., and Rabanus, S., editors, *Language and Space: An International Handbook of Linguistic Variation Volume 2: Language Mapping*, volume 30/2 of *Handbooks of Linguistics and Communication Science*, pages 561–585. Berlin: Mouton de Gruyter.

Güldemann, T. (2018). Areal linguistics beyond contact, and linguistic areas of afrabia. In Güldemann, T., editor, *African Languages and Linguistics*, pages 448–545. Berlin: DeGruyter Mouton.

Güldemann, T. and Hammarström, H. (in press). Geographical axis effects in large-scale linguistic distributions. In Crevels, M., Muysken, P., and Hombert, J.-M., editors, *Language dispersal, diversification and contact*, page 25pp. Oxford: Oxford University Press.

Laitin, D. D., Moortgat, J., and Robinson, A. L. (2012). Geographic axes and the persistence of cultural diversity. *Proceedings of the National Academy of Sciences*, 109:10263–10268.