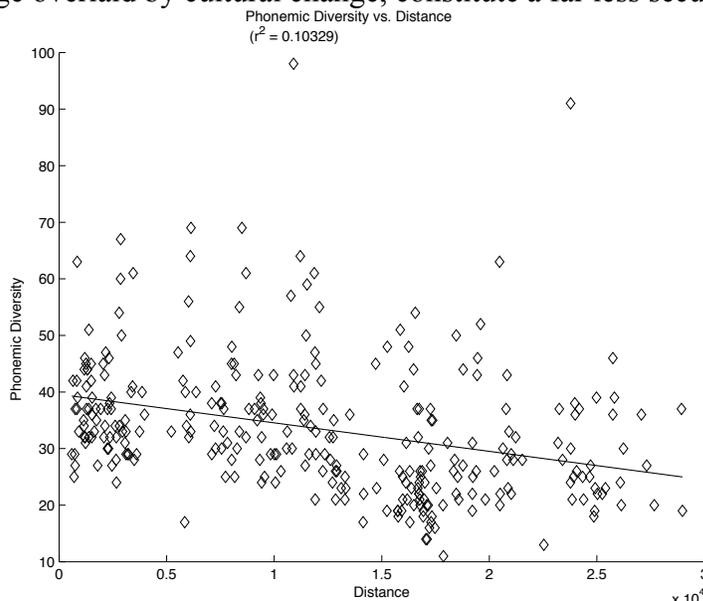


Language evolution is not like genomic evolution: Phonemic Diversity Fails to Detect Language Evolution out of Africa

Recent work [1] claims that one can track language's evolutionary history via phonemic diversity (*pd*): it asserts that as one moves away from Africa, *pd* falls, paralleling a genomic diversity-African origin link [2]. To the contrary, we have demonstrated that this hypothesized link does not hold: replacing the phonological metric in [1] with phoneme counts from the UCLA database [3], barely 10%, of *pd* variation is accounted for by distance from Africa, far less that found with genetic variation, which approaches nearly 100% [2]. We conclude that unlike genetic diversity – more easily measured, located, and tracked – language's within-species evolutionary variation over time does mirror genetic variation. Rather, the association between phonemic diversity and African origin reflects an attenuated statistical 'echo' piggy-backed on the genomic variation associated with human migration, a result consonant with other research results correlating variation in human external form, e.g., cranial measurements, with distance-from-Africa [4].

[1] modeled language diffusion as successive population waves from Africa, with phonemic diversity lower in smaller founder populations. Language diversity is also expected to be inversely proportional to distance from Africa, directly proportional to log population size (see [5]). However, other researchers have advanced arguments pro and con regarding [1]'s results; see [6]-[12]. To our knowledge, to date there has been no attempt at an exact, control replication of [1], followed by more accurate phonological analysis. We have done that here. First, we carried out a replication of [1], using data supplied by its author (population size, adjusted km distance from Africa, and 'binned' WALS phonemic data, 504 languages). We then applied statistical analysis as in [1] (including Mantel and partial Mantel correlations, using the same Arlequin 3.1 bioinformatics package as in [1]). We replicated exactly all statistical results of [1], e.g. [13], Table S3. Next, we replaced [1]'s 'binned' values for consonants, vowels, and tones (e.g., 2-4 vowels =value 1; 5-6=2; 7-14=3; etc.) with numerical values from the UCLA phonological segment inventory database [3]. We retrieved as many languages as could be matched against those from the WALS sample used by [1], roughly 80% of the original, for a total of 301 languages out of 504; the reduced sample was examined and found unbiased. Statistical tests on the new dataset following [1] showed that distance from Africa fell accounting for 25.6% of the variance in *pd* (controlled for population size), to 10% ($p < 0.001$), leaving 90% of the *pd* variation unexplained; see the Figure below. This low explained variance component is matches that found for other phenotype-African origin studies, e.g., cranial measurements in [4], who likewise noted that a complex genotype-to-phenotype mapping might attenuate statistical detection of a distance-phenotype 'signal.' In sum, the direct application of methods drawn from genetics to language evolution bear closer scrutiny: while the measurement of genetic variation comprises a staple of our understanding of how the human population diversified and changed, phenotypic properties, particularly behavioral phenotypes such as language overlaid by cultural change, constitute a far less secure 'signal' for tracking evolutionary diffusion.



- [1] Atkinson, Q. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332, 346–349.
- [2] Ramachandran S, Omkar D, Roseman C, Rosenberg N, Feldman M, Cavalli-Sforza L, 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* 102, 15942–15947.
- [3] Maddieson I. & Precoda K. 1992. *UPSID and PHONEME v. 1.1.* , UCLA, Los Angeles, CA. Data retrieved from: <http://www.linguistics.ucla.edu/faciliti/sales/software.htm>.
- [4] Manica, A., Amos, W. Balloux, T. Hanihara, T. 2007. The effect of ancient population bottlenecks on human phenotypic variation. *Nature* 448, 346-348.
- [5] Hay, J, Bauer, L, 2007. Phoneme inventory size and population size. *Language* 83 388-400.
- [6] Sproat, R. 2011. Phonemic diversity and the out-of-Africa theory. *Linguistic Typology* 15 199-206.
- [7] Rindge, D. A pilot study for an investigation into Atkinson’s hypothesis. 2011. *Linguistic Typology* 15 223-241.
- [8] Rice, K. Athabaskan languages and serial founder effects. 2011. *Linguistic Typology* 15 233–250.
- [9] Ross, B. & Donahue, M. 2011. The many origins of diversity and complexity in phonology. *Linguistic Typology* 15 251-266.
- [10] Maddieson, I Bhattacharya, T, Smith, Croft, W. 2011. Geographical distribution of phonological complexity. *Linguistic Typology* 15 267–279.
- [11] Jaeger F, Graff, P, Croft, W, Pontillo, D. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15 281–319.
- [12] Atkinson Q. 2011. Linking spatial patterns of language variation to ancient demography and population migrations. *Linguistic Typology* 15, 321–332.
- [13] Atkinson, Q. 2011. Supporting Online Material for: Phonemic Diversity Supports Serial Founder Effect Model of Language Expansion from Africa DOI: 10.1126/science.1199295.