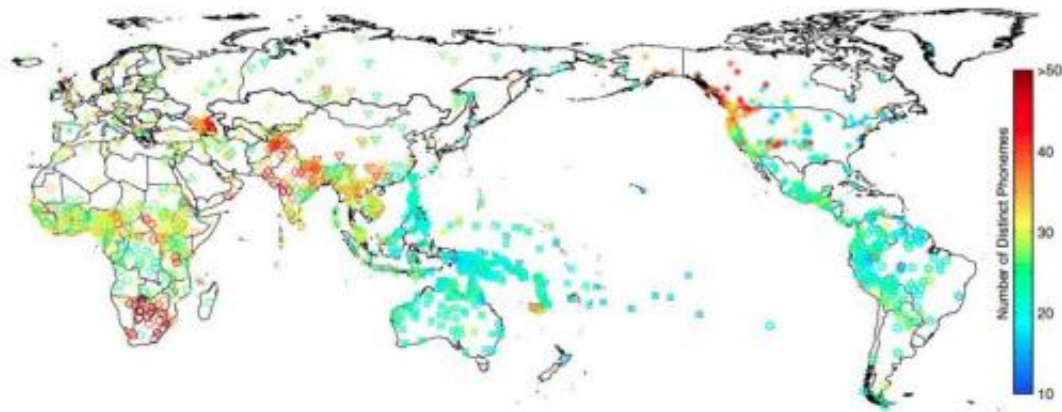# Probing the deep history of human genes and language

January 20 2015, by David Orenstein

Diversity of differences. Researchers analyzed distinct sounds — phonemes — in more than 2,000 languages around the world alongside genetic markers from more than 200 populations to uncover geographic patterns of how languages differ.

Brown University evolutionary biologist Sohini Ramachandran has joined with colleagues in publishing a sweeping analysis of genetic and linguistic patterns across the world's populations. Among the findings is that geographic distance predicts differentiation in both language and genes.

Producing new insights into the evolution and development of human populations around the globe is no easy task, but scientists can draw on multiple sources of data to do it. In a new study, Sohini Ramachandran

and colleagues at Stanford University and University of Manitoba analyzed troves of data on genetics and distinct sounds in [language](#) —phonemes—to discern important patterns.

Among the findings published in *Proceedings of the National Academy of Sciences*, is that genes and languages both vary more as geographic distance increases. The analysis showed there are distinct geographic patterns, or axes, of the greatest differences. The data also reflect how languages and genes evolve differently, for instance among isolated populations.

Ramachandran, assistant professor of ecology and evolutionary biology, discussed these and other insights with writer David Orenstein.

Why are language and genes sometimes combined in studies of populations?

Fields that study the human past, especially ancient human history, have to draw on multiple disciplines and lines of evidence in order to confirm and calibrate observed signatures in data, since we can't truly know all events in human history. Because language is inherited 'vertically' [from parents to children] like genes, and also changes 'horizontally' based on contact among populations, many researchers in genetics interpret analyses of DNA from different populations in the context of the languages the study populations speak.

This kind of interdisciplinary work is what initially drew me to studying [human evolution](#).

In this study what did you find was similar between languages and genes and what was different?

We saw that axes of differentiation in both our linguistic and genetic dataset corresponded, meaning that differences in both datasets of very different types of markers were geographically distributed quite similarly.

One very interesting contrast we saw between languages and genes had to do with isolated populations: an isolated population loses genetic diversity rapidly, as individuals marry within the [population](#); in contrast, we saw a range of variation in linguistic markers for languages that are geographically isolated (have few neighboring languages). Some languages that are isolated lose complexity and others gain complexity and innovate new sounds. This makes me wonder whether contact among populations homogenizes their languages in some way so people can understand each other.

We found that linguistic markers do not hold signatures of the human expansion out of Africa, which is not surprising due to the rate at which languages changes and can be influenced by neighboring languages.

Tell us more about that difference between what genes and languages showed regarding human origins in Africa?

To be precise, genes tell us that the people living today with the most [genetic diversity](#) currently live in Southern Africa (like the San bushmen) and that modern humans emerged in Africa, but we don't know where the geographic origin of our species was precisely based on genetic data. The language analysis did not reveal this African origin because language changes in a complex way, much differently from genes where we have a good sense of the mutation process. In my conversations with different linguists, including those at Brown who generously listened to me present our ideas multiple times, the rate at

which language mutates, and which linguistic markers are more likely to change than others, seems to be an open question.

You found geographic axes, or directions, of difference in language and genetics. What might they tell us about human evolution and history?

These axes, which look for directions along which a dataset is most differentiated, tell us about axes along which humans likely did not migrate a great deal. For example, migration north/south in Africa would mean moving across climate regimes; we also know populations are quite different across latitudes in Europe and we see that for both our language datasets and genetic datasets.

What do your findings tell us about how we can use genes and language, either together or separately, for population studies?

We learn more from using both data types together and analyzing them using similar methods than we would have learned from either type alone. One signal we saw loud and clear in this study is how much geographic distance affected our ancestors' genes and languages; geographic distance predicts differentiation in both data types, underscoring that there are still deep signatures of ancient migrations in our genomes and cultures today.

  More information: "A comparison of worldwide phonemic and genetic variation in human populations," by Nicole Creanza et al. *PNAS*, www.pnas.org/cgi/doi/10.1073/pnas.1424033112

Provided by Brown University