Languages, Clicks, and Genetic Diversity

Joanna L. Mountain[1,2*], Mary Katherine Gonder[3], Brenna M. Henn[1], Holly Mortensen[3], Christopher Gignoux[1], Peter A. Underhill[2], Uma Ramakrishnan[4], Floyd A. Reed[3], Sarah A. Tishkoff[3*]

[1]Department of Anthropological Sciences, Stanford University, CA, [2]Department of Genetics, Stanford University, CA, [3]Department of Biology, University of Maryland, College Park, MD, [4]National Centre for Biological Sciences, Bangalore, India
[*]These authors contributed equally to this work.

If the relationships amongst languages of a family are well characterized, analysis of genetic variation amongst populations speaking those languages may reveal the extent to which languages evolve in the context of particular levels of gene flow. If, however, linguistic relationships are poorly characterized, conclusions may be far less clear. An optimistic goal may be to constrain the set of plausible linguistic relationships. Several questions remain regarding the nature of the relationships among the approximately 35 African languages with click consonants. The time depth of the language family/ies that include click languages is/are also controversial. However, linguists agree on at least some relationships. Here we consider the genetic diversity in light of those relationships, and assess whether the genetic data allow us to narrow the set of hypotheses regarding other linguistic relationships.

We initially examined the Y chromosome and mitochondrial (mt) DNA diversity of the Hadza of Tanzania, concluding from these data and others that the eastern African Hadza and the Jul'hoansi of southern Africa (Namibia) are genetically distant from one another, providing little evidence of contact during the last 20,000 or more years. Based on this finding, we outlined four interpretations of the data, highlighting the simplest: independent retention of clicks for tens of thousands of years. An alternative model involved other populations. Is it possible that a third population, in contact with both the Hadza and the southern African click speakers, facilitated the spread of the click component of the languages? One population that could have played such a role is that of Sandawe speakers, currently living in Tanzania within two hundred kilometers of the Hadza. Given this geographic proximity, we might expect these two click speaking populations to be genetically similar. Although a handful of linguists consider the Hadza and Sandawe languages to be distantly related, others detect no relationship. Our a priori expectation based on linguistics, therefore, is genetic dissimilarity between these two populations.

We have now generated Y chromosome and mtDNA data for the Sandawe population. These data reveal at least three phenomena. First, both data sets reveal the genetic influence of neighboring, non-click speaking populations on these two click-speaking populations. Second, both Y and mtDNA data sets suggest that the Sandawe and Hadza have remained relatively isolated from one another for the last 10,000 years. Third, comparison of these data with the limited available genetic data for southern African click speakers reveals no evidence of recent contact (within the last 20,000 years). Given these data we are convinced that these click-speaking populations have been relatively isolated from one another, even in the case of the geographically proximate pair. This finding is consistent with the linguistic evidence for deep relationships, if any, among the languages spoken by these populations. If the Sandawe served as a conduit for clicks, they did so in the distant past.