

Linguistic and genetic relationships in Northern Cameroon

Brett C. Haberstick¹, Gary P. Stetler¹, Andrew Smolen¹, John K. Hewitt¹, Zygmunt Frajzyngier²,
Eric Johnston² and Erin Shay²

¹Institute of Behavioral Genetics, University of Colorado, ²Department of Linguistics, University of Colorado

Cameroon, (Central Africa), has about 270 languages, which belong to three linguistic families. In Northern Cameroon alone, there are over 100 languages spoken, and the differences among the languages are quite significant. The object of this feasibility study was to collect DNA in the field and determine the degree of genetic diversity among individuals comprising six language groups in Northern Cameroon: Hdi (N=30), Mafa (31), Mina (29), and Gidar (30) (all Central Chadic); Peve (31) (Masa, Chadic); and Mambay (31) (Niger Congo). DNA samples were collected on buccal swabs and sent to Institute of Behavioral Genetics for analysis. Here we report the preliminary analysis of data from genotyping 20 unlinked Short Tandem Repeat (STR, microsatellite) markers across the genome with average heterozygosities greater than 0.70. Data were analyzed using the program structure (Pritchard et al, 2000, Falush et al, 2003). Results from structure indicated that the genotype of an individual can be considered to be drawn from their respective subgroup ($\alpha = 0.0491$). Furthermore, there is genetic differentiation across language groups (FST range: 0.0003-0.1229). Our ability to predict language subgroup membership based on genetic information was modest, but our results indicate that 6 disparate language groups exhibit genetic divergence and vary in genetic diversity. Whether the diversity within language groups resulted from six separate founding populations or from dispersion and isolation of the language groups from ancestral sources remains an open question. Efforts are in progress to genotype 12 Y-chromosome STR loci, and mitochondrial DNA typing is planned to address the question of ancestral origin of the genetic diversity among the six language groups. This study demonstrates the feasibility of DNA collection in these populations and our analyses suggest the potential value of more extensive genotyping of these samples and more extensive studies of the relationship between genetic and linguistic diversity in this region.

Falush D., Stephens M. and Prichard J.K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164: 1567-1587.
Pritchard, J. K., M. Stephens and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155:945-959.