B118 The Determination of Biogeographic Ancestry Within Hispanic Populations

Casandra Hernandez Setser, MSFS, Benbrook, TX 76132; Deanna S. Cross, PhD, University of North Texas Health Science Center, Fort Worth, TX 76107; John V. Planz, PhD, UNTHSC, Fort Worth, TX 76107; Ranajit Chakraborty, PhD, University of North Texas Health Science Center, Fort Worth, TX 76107*

Learning Overview: After attending this presentation, attendees will have a better understanding of the potential of single nucleotide polymorphisms (SNPs) to describe ancestry in such fine detail that it will separate closely related populations originating from Latin America.

Impact on the Forensic Science Community: This presentation will impact forensic science community by: (1) increasing their awareness and competence in a next generation technology that harnesses previously unused portions of the genome to obtain objective information on ancestry, and (2) allowing attendees to employ this technology in cases where standard genetic analyses are unable to provide useful information.

The STRs currently used in forensic genetics were specifically selected to have similar allele frequencies in all populations studied. However, there are instances when associating an unknown individual with a likely biogeographic ancestry (via loci with large differences in allele frequency) would be beneficial. When genotyping the Core 20 CODIS loci is not possible or when there is no CODIS hit, combining a number of ancestry informative SNPs together allows an examiner to detect the most likely ancestry associated with a sample and generate an investigative lead that is more objective than an eye-witness.[1] Therefore, it is hypothesized that a small panel of high quality SNPs is sufficient to differentiate closely related Hispanic populations into their specific biogeographic ancestry.

The previously existing Genomic Origins and Admixture in Latinos (GOAL) dataset was used computationally to develop such a panel.[2] Here, 164 unrelated samples from 5 different Hispanic populations were utilized to develop this SNP panel. Beginning with 897,336 polymorphisms genotyped within the population, PLINK was used to filter the dataset for linkage disequilibrium (LD), missingness (<10%), and minor allele frequency (>1%); 1215 SNPs were selected that had high F_{ST} values for the 4 pairwise comparisons with one country in common possible when studying 5 populations.[3] Further refinement was required when it became apparent that 71% of the 1215 SNPs could be attributed to Honduras, the only Central American population in this dataset. Therefore, only the ~10 SNPs with highest F_{ST} for the 1st and 2nd Country-in-Common (CiC) were selected for the final panel of 54 SNPs. This smaller, more balanced SNP panel achieved a similar degree of separation as the original 1215 SNP panel.

This ancestry SNP panel was compared to other SNP panels in the literature and tested for its predictive value. In comparison to the Kidd 55 and the Seldin 128, this panel of 54 SNPs produced a higher overall F_{ST} distribution.[4,5] Additionally, Bayesian method STRUCTURE and Principal Components Analysis method EIGENSOFT were employed to separate the populations.[6,7] However, it was found that this panel works best when employed as a 2-tiered analysis. The Honduras SNPs had a much higher mean F_{ST} than the other populations, so it is recommended that the panel be run for Honduras first, then for the other 4 countries. Discriminant Function Analysis (DFA) is being used to evaluate the panels to calculate centroids of each population for the prediction of unknowns.

In conclusion, this Hispanic Ancestry panel of 54 SNPs may prove useful when an individual is presumed to be Hispanic and more specific information is needed to generate an actionable investigative lead.

Reference(s):

1. Christianson, Sven-Ake. "Emotional Stress and Eyewitness Memory: A Critical Review" *Psychological Bulletin* 112, no. 2 (1992): 284-309
2. Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., … Bustamante, C. D. (2013). Reconstructing the population genetic history of the Caribbean. *PLoS Genetics*, *9*(11), e1003925. doi:10.1371/journal.pgen.1003925
3. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. a R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559
4. Kidd, K. K., Speed, W. C., Pakstis, A. J., Furtado, M. R., Fang, R., Madbouly, A., … Kidd, J. R. (2014). Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International. Genetics*, *10*, 23–32. doi:10.1016/j.fsigen.2014.01.002
5. Kosoy, R., Nassir, R., Tian, C., White, P., Butler, L., Silva, G., …Seldin, M.F. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions common in populations in America. *Human mutation*, *30*(1), 69-78. doi:10.1002/humu.20822
6. Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–59. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461096&tool=pmcentrez&rendertype=abstract
7. Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), 2074–2093. doi:10.1371/journal.pgen.0

Ancestry, Single Nucleotide Polymorphism, Population Structure

Copyright 2019 by the AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by the AAFS.