A5 Population Identifiability From Forensic Genetic Markers: Ancestry Variation in Latin America

Cris E. Hughes, PhD, Department of Anthropology, 109 Davenport Hall, 607 S Matthews Avenue, Urbana, IL 61801;and Bridget F.B. Algee-Hewitt, PhD*, Stanford University, Rosenberg Lab, Dept of Biology, Gilbert Bldg, Rm 109, 371 Serra Mall, Stanford, CA 94305-5020*

After attending this presentation, attendees will better understand how the ancestry content of forensic genetic markers can contribute to the study of Latin American variation and how it can assist forensic anthropologists in choosing population-specific methods when inferring other biological profile parameters.

This presentation will impact the forensic science community by demonstrating that forensic Short Tandem Repeats (STRs), such as the Combined DNA Index System (CODIS) markers, are valuable resources for population structure analyses at micro-regional levels, specifically in Latin America. This presentation also quantifies and confirms the agreement between ancestry/admixture patterns produced from a small panel of forensic genetic markers and a large gold-standard dataset.

Ancestry information content in panels of forensic STRs, including the CODIS loci, has been most recently addressed by Algee-Hewitt et al., who demonstrated that forensic STR markers with high individual identifiability carry a non-trivial amount of information on ancestry.[1] These findings support the legacy of population research in anthropological genetics that uses forensic genetic markers in the study of population history and the analysis of contemporary variation. The observed link between individual and population identifiability also has implications for forensic anthropology. These markers have both research and practical importance: the individual identification profiles often represent the only source of genetic information for understudied populations and they provide another source of biological information on admixture and ancestry that is of concern for, broadly, casework logistics and, specifically, individual skeletal case analysis. While small sets of forensic STR loci have been used to reveal latent structure, estimate ancestry, or generate proportions of admixture, the appropriateness of these applications using these markers has not been thoroughly studied. The goal, here, is to determine the compatibility of ancestry/admixture estimates generated from forensic and non-forensic STR markers and to articulate the value of forensic loci for different levels of population identifiability in forensic anthropology.

Genetic data representing Latin America variation are used to evaluate the utility of forensic microsatellite markers for fine-grained population research. This region is of special forensic relevance because of the humanitarian crisis at the United States-Mexico border and the challenges that the identification of undocumented migrant fatalities pose for forensic anthropologists. Moreover, Latin America is composed of highly admixed populations with varied patterns of ancestry. Genotypes were sourced from the Wang et al. dataset, which includes 240 admixed individuals from 13 Latin American populations genotyped for 678 microsatellite markers.[2] Within this "Full" dataset, 9 of the traditional 15 forensic loci were identified, and tetranucleotide STR markers that recapitulate key properties of the CODIS loci. Drawing on this sample of nine CODIS and the identified CODIS-like markers, ten unique "Test" datasets, which contain 15 STRs each, were generated to reflect the traditional size of a forensic panel. Parallel STRUCTURE analyses are run for all datasets. K=2 solution, including Native American and European parental populations, is the preferred model.

As the purpose of this analysis is to determine if forensic STR panels display structure patterns and confer

Copyright 2017 by the AAFS. Unless stated otherwise, noncommercial *photocopying* of editorial published in this periodical is permitted by AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by AAFS.

ancestry information in amounts statistically similar to traditionally large, microsatellite datasets not used in forensic identity testing, the Full dataset was designated as the non-forensic, gold standard. A consensus solution is obtained from the STRUCTURE results for each of the ten Test datasets (15 loci). It is compared against the results generated using the Full dataset (678 loci). The similarity statistic computed between these two sets of admixture-cluster results is high (0.86). Cross-classification using the hard (indigenous) cluster assignments from STRUCTURE is X^2 significant (R^2=0.55) and classification error is remarkably low (4.24%). Significant and strong positive correlations exist between the indigenous component estimates from the Test and Full datasets when partitioned by subpopulation ($0.66 \leq p \leq 0.96$, $p < 0.001$).

Results of two-sided paired *t*-tests for mean indigenous cluster assignments between the Test and Full datasets are significant for 3 of the 13 Latin American samples. The average of the mean differences in indigenous cluster assignments is 6%. Kolmogorov-Smirnov testing identifies a significant distributional difference for only one sample. The cumulative analyses indicate that the CODIS/CODIS-like markers tend to overestimate the amount of indigenous ancestry, when K=2; however, this overestimation is small, systematic, and corrected when the STRUCTURE analysis includes African parental, K=3. Analysis of Variance (ANOVA) reveals significant subpopulation mean differences for the Test (R^2=0.81) and Full (R^2=0.92) datasets. Post-hoc comparisons find fewer among-population, microregional differences for the Test dataset, indicating that forensic markers provide a lower-resolution, more conservative picture of Latin American variation.

Reference(s):

1. Algee-Hewitt Bridget F.B., Edge Michael D., Kim J., Li Jun Z., Rosenberg Noah A. Individual Identifiability Predicts Population Identifiability in Forensic Microsatellite Markers. *Current Biology.* 2016: 26:935-942.
2. Wang S., Ray N., Rojas W., Parra M.V., Bedoya G., Gallo C., Poletti G., Mazzotti G., Hill K., Hurtado A.M. et al. Geographic Patterns of Genome Admixture in Latin American Mestizos. *PLoS Genet.* 2008: 4:e1000037.

Forensic Genetics, Ancestry, Latin America

Copyright 2017 by the AAFS. Unless stated otherwise, noncommercial *photocopying* of editorial published in this periodical is permitted by AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by AAFS.