EVALUATION OF FORENSIC RESOURCE/REFERENCE ON GENETICS KNOWLEDGE BASE (FROG-kb) FOR THE PREDICTION OF INDIVIDUAL BIOGEOGRAPHIC ANCESTRY

Kasima Brown, MFS[1]; Andrew Feiter, MFS[1]; Rachel Wiley, MFS[1]; Moses Schanfield, PhD[1]; Katherine Gettings, PhD[2]; and Daniele Podini, PhD[1].

[1]Department of Forensic Sciences, The George Washington University
[2]National Institute of Standards and Technology

Forensic DNA analyses most commonly employ Short Tandem Repeat (STR) polymorphisms to compare evidence profiles to known profiles or databases such as CODIS. However, when a DNA profile does not match identified suspects or profiles from available databases, additional DNA analyses, such as those targeted at inferring the possible ancestral origin of the perpetrator, could yield valuable information. In recent years there have been many proposed Ancestry Informative Marker (AIM) sets for use in predicting biogeographic ancestry. There is a vast range of DNA polymorphisms available with potential to be used as AIMs including Single Nucleotide Polymorphisms (SNPs), indels, mitochondrial DNA (mtDNA) sequence variations, Y-chromosome STRs and, to a lesser extent, autosomal STRs. SNPs are the most common form of genetic polymorphism and Ancestry Informative SNPs (AISNPs) have alleles associated with specific populations and can be helpful in forensic investigations when STR profiles fail to yield an identification.

FROG-kb is a freely available online tool with the primary objective of providing a web interface with the data housed in the already extensively used and referenced Allele FREquency Database (ALFRED) making it more suitable for forensic purposes. FROG-kb provides the ability to display the ALFRED data in an organized manner as well as computational tools that apply the underlying allele frequencies to generate statistics for forensic unknown samples. Multiple AISNP panels are available in FROG-kb. The purpose of this project was to evaluate the ancestry predicting potential of a 25 AISNP panel available in FROG-kb. This panel is based on a SNP assay developed at GWU, which was designed to predict ancestry among the primary U.S. populations. The analysis is performed by entering the sample genotype at each SNP locus. FROG-kb then outputs probabilities of the profile for 89 populations, sorted from highest to lowest (statistically most common to most rare, respectively). The output table consists of the populations (and biogeographic regions), the probability of the SNP profile within each population, and a likelihood ratio (LR, the probability of the SNP profile from the population in which it is highest divided by the probability of the profile in the specified population). The LR expresses how much more likely it is to observe that profile if it originated from the population in the numerator versus if it originated from the population in the denominator.

A test set of 127 samples, consisting of 31 European American, 32 African American, 32 Hispanic American, and 32 East Asian were used to develop the FROG-kb result interpretation criteria. The criteria were then used in a blind test consisting of 200 individuals from the 1000 Genomes Project. Of the tested samples, 94% were classified to the correct major U.S. population, while 3.5% (n=7) were misclassified, and 2.5% (n=5) were classified as inconclusive. These 12 were all self-reported Hispanic or Native American sample profiles.

This study has demonstrated that FROG-kb is a convenient and accurate method for predicting an individual's biogeographic ancestry and can be a valuable tool in forensic investigations.