## B51 Microhaplotypes (MHs): A Possible Biogeographic Ancestry Predictor for Forensic DNA Evidence

*Kelly E. Long, BSc*, The George Washington University, 7308 Snowden Court, Springfield, VA 22150; and Katrina F. Maddela, BS*, 4888 MacArthur Boulevard, NW, Washington, DC 20007*

After attending this presentation, attendees will understand MHs and the potential that MH analysis has in predicting the biogeographic ancestry of an individual.

This presentation will impact the forensic science community by suggesting how MHs can enhance mixture interpretation and ancestry prediction capabilities, potentially providing investigators with an extra tool for identifying and/or eliminating suspects.

MHs are short regions (<250 base pairs long) that have two or more Single Nucleotide Polymorphisms (SNPs) with three or more allelic combinations. Given the short distance between SNPs in MHs, recombination is highly unlikely. Furthermore, the mutation rate is significantly less than that of Short Tandem Repeat (STR) markers, which makes them a better candidate for familial testing. Contrary to STRs, the amplification of MHs is not affected by polymerase slippage; given the absence of a short sequence repeated multiple times, this feature eliminates the presence of a common STR analysis artifact: stutter. With conventional SNP typing methods, such as Sanger sequencing, TaqMan®, and SNaPShot®, haplotypes cannot be determined. Using Massively Parallel Sequencing (MPS) methods, the *cis/trans* relationship between SNP alleles within the same amplicon (i.e., haplotype detection) can be accomplished. MPS-based MH analysis can be a valuable forensic tool for parentage testing and mixture deconvolution, particularly in mixtures where the minor contributor (or contributors) is in the stutter range of the major. The small amplicon size also makes MHs a good candidate for typing degraded DNA samples. Although initially selected based on the highest effective number of alleles across populations, MH allele frequencies vary significantly in different populations. The goal of this project was to evaluate if a selected panel of MHs, together with individualization, could also effectively predict biogeographic ancestry.

A panel of 33 MHs was used on 98 individuals who self-identified as European-Americans. Samples were sequenced on an Ion S5 MPS platform using a 530 chip. The initial library preparation was conducted manually but the templating step was carried out on an Ion Chef™. Allele frequencies from a database of 58 different populations were used to calculate the Random Match Probability (RMP) of each profile in each individual population. A biogeographic prediction was performed by evaluating the raw RMP values for each population and whether or not the highest RMP was in a European population. A second method to evaluate biogeographic inference was by calculating a Likelihood Ratio (LR) determined by dividing the probability of the highest (as in most common) RMP in question by the RMP of the population chosen for comparison.

Of the 98 individuals, only nine were found to have their highest RMP in a non-European population. All nine were within one order of magnitude or less. For these nine individuals, the highest RMP was calculated in a Middle Eastern population. European populations and Middle Eastern populations are generally considered admixed.

The LR was calculated for all 98 individuals by placing the European-American RMP in the numerator and RMPs of Yoruba, African-American, Cambodian, San Francisco Chinese, and Arizona Pima populations in the denominator as proxies for the major United States populations. The value of the LR obtained represents how much more likely it is to observe that profile if the individual is of European descent versus one of the other populations.

The average LR was ~3 $E^{18}$ for European-Yoruba or -Pima, only a magnitude greater for European-Cambodians, but three orders greater for the European-San Francisco Chinese at $E^{21}$. The European-African-American was the smallest range from 556 to 5.5 $E^{14}$ while the European-San Francisco Chinese was the greatest from 208 to 6.3$E^{23}$. The European-Pima and European-Yoruba LRs were similar from the 10,000s to $E^{20}$. For the 98 samples, 94 LRs were above the arbitrary level of confidence of 1,000. The lowest value in the four remaining was 208 (European/-San Francisco Chinese), and the second lowest value was 556 (European-African-American). These four profiles also had high frequencies in other non-European populations, suggesting they may have originated from admixed individuals.

Further research is necessary to identify more effective ancestry informant MHs. Yet, these results support the hypothesis that MHs can be used for biogeographic ancestry prediction, which can provide investigators with useful information in cases in which the STR profile from crime scene evidence does not match any suspects or database profiles.

**Microhaplotypes, Forensic, Ancestry**