B41 The Development of an Innovative Massively Parallel Sequencing (MPS) Panel of Microhaplotypes for Improved Biogeographic Ancestry Inference

Sathya Prakash Harihar, The George Washington University, Department of Forensic Science, Washington, DC 20007; Fabio Oldoni, PhD, The George Washington University, Washington, DC 20007; Aishwaryaa Subramanian, The George Washington University, Washington, DC 20052; Leena Yoon, Tysons, VA 22102; Sharon C. Wootton, PhD, South San Francisco, CA 94080; Robert Lagacé, BS, Thermo Fisher Scientific, South San Francisco, CA 94080; Ryo Hasegawa, BS, Foster City, CA 94404; Joseph P. Chang, BS, Thermo Fisher Scientific, South San Francisco, CA 94080; Moses S. Schanfield, PhD, The George Washington University, Department of Forensic Sciences, Washington, DC 20007; Kenneth Kidd, PhD, Yale University School of Medicine, New Haven, CT 06520; Daniele S. Podini, PhD, Washing State University, Department of Forensic Science, Washington, DC 20007*

Learning Overview: After attending this presentation, attendees will be able to understand the potential of using microhaplotype (MH) markers for ancestry inference.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating the usefulness of an innovative massively parallel sequencing (MPS) assay of 74 MH loci for enhancing biogeographic ancestry prediction capabilities.

With the advent of high-throughput sequencing technology new genetic markers called microhaplotypes (MHs) have become available to the forensic community. MHs are loci characterized by the presence of two or more single nucleotide polymorphisms (SNPs) within a short distance from each other (< 300 nucleotides) and associated in three or more allelic combinations.[1] Standard Sanger sequencing method is unable to determine the *cis/trans* relationship among SNP alleles within the same expanse of DNA. MPS instead allows distinguishing the parental haplotypes at any given locus within the same amplicon by specific clonal sequencing of each individual DNA strand. This process allows determining the haplotype phase of the targeted SNP alleles within each MH locus. Key characteristics of MHs include multi-allelic nature, absence of stutter peaks, small amplicon size, and lower mutation rates than conventional short tandem repeat polymorphisms (STRPs).[2] All of these features contribute to making MHs promising candidates for different forensic applications including ancestry inference, mixture deconvolution, and human identification. The aim of this study was to evaluate the performance of a recently developed MPS-based MH assay for the prediction of biogeographic ancestry of individuals using specifically phased-inferred allele frequencies reported on the Allele Frequency Database (ALFRED) database.[3,4]

A novel panel targeting 74 MH loci and totaling 230 SNPs was developed on the Ion Chef™ and Ion S5™ MPS (Thermo Fisher Scientific) platform.[5] A set of 20 unknown African Americans (AAs), European-Americans (EAs), East Asian Americans (EAAs), South West Hispanics (SWHs) and one Mexican Pima (MP) test-samples were selected and tested for evaluating biogeographic origin prediction. Statistically phased-inferred allele frequencies of the selected 74 multi-SNP loci were extracted from ALFRED and further utilized to calculate the random match probability (RMP) of the unknown test-samples in each relevant population. The Log10 of the RMP was further calculated for all test-samples by utilizing ALFRED allele frequencies from the 74 MH loci across a worldwide set of 26 different populations representative of African (7), European (9), Asian (6) and Native American (4) population clusters. Overall, the biogeographic ancestry of the 80-unknown test-population samples was correctly predicted using the ALFRED MH allele frequencies. Among the populations studied, the Log10 RMP calculated for each test-sample was found to average significantly higher in the corresponding population of origin while SWHs averaged equally high in European and Native American populations, as expected. Average values of Log10 RMP for unknown AAs, EAs, EAAs, SWHs test-samples ranged between -69 and -78, -63 and -70, -62 and -65, -76 and -83 across African, European, East Asian and European/Native American populations, respectively while Log10 RMP of MP test-sample ranged between -56 and -60 in NA populations.

These findings indicate that the MPS panel of 74 MH loci is an effective forensic DNA tool, which provides valuable information on the biogeographic origin of individuals complementing the accuracy of current available forensic DNA-based ancestry prediction assays.

Reference(s):

1. Pakstis AJ, Fang R, Furtado MR, Kidd JR, Kidd KK. Mini-haplotypes as lineage informative SNPs and ancestry inference SNPs. *European Journal of Human Genetics* (2012) 20(11): 1148-1154.
2. Kidd KK, Speed WC, Pakstis AJ, Podini DS, Lagacé R, Chang J, Wootton S, Haigh E, Soundararajan U. Evaluating 130 microhaplotypes across a global set of 83 populations. *Forensic Science International: Genetics* (2017) 29:29-37.
3. Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK, ALFRED: an allele frequency resource for research and teaching, *Nucleic Acids Res*. (2012) 40D:D1010–D1015 (Database issue).
4. https://alfred.med.yale.edu
5. Oldoni F, Hart R, Long K, Maddela K, Cisana S, Schanfield M, Wootton S, Chang J, Lagace R, Hasegawa R, Kidd K, Podini D, Microhaplotypes for ancestry prediction, *Forensic Science International: Genetics Supplement Series* (2017) 6:e513-e515.

Microhaplotype, MPS, Ancestry Prediction