B1 Complex DNA Mixture Analysis: Massively Parallel Sequencing (MPS) of Rare Single Nucleotide Polymorphisms (SNPs)

Martha Petrovick, PhD, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA 02420; Natalie Damaso, PhD, Oak Ridge Institute for Science and Education/FBI Laboratory, Woodbridge, VA 22191; Tara Boettcher, BS, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA 02421; Josh Dettman, PhD, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA 02420; Darrell O. Ricke, PhD, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA 02421; Philip Fremont-Smith, MS, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA 02421; Kyleen E. Elwick, PhD, Oak Ridge Institute for Science and Education/FBI Laboratory, Quantico, VA 22135; Mark F. Kavlick, PhD, Quantico, VA 22135; James M. Robertson, PhD, Reserve Support Unit, Quantico, VA 22135; Eric Schwoebel, PhD, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA 02421*

Learning Overview: After attending this presentation, attendees will understand how using MPS of rare SNPs allows the matching of more than three references to complex mixtures of DNA, including those obtained from touch DNA.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing experimental evidence from two independent laboratories that demonstrates the power of using MPS of rare SNPs for the analysis of complex mixtures.

DNA mixtures from three or more contributors have proven difficult to analyze using the current state-of-the-art method of Short Tandem Repeat (STR) amplification followed by Capillary Electrophoresis (CE). This is due to multiple issues, including the sharing of alleles between different contributors, the production of confounding stutter products during amplification, and the small amount of DNA recovered from some contributors, causing allele drop-out. The discriminatory power of STRs lies in the fact that they are highly polymorphic, even though the individual alleles are not necessarily rare. This feature has enabled forensic scientists to streamline their STR analysis workflow to tens of loci in the genome. In contrast, there are only two alleles at each SNP locus used in this study, and the presence of each minor allele in an individual is relatively uncommon. Therefore, more SNP loci are required to produce a unique individual signature, and many more SNP loci are required for successful mixture analysis.[1] Although the SNP loci used in this study are also affected by allele sharing and drop-out, the drop-in rates for single nucleotides are relatively low, there are no stutter artifacts, and many more loci can be analyzed simultaneously, producing statistical power that enables detection of individual contributors in complex mixtures.

SNP loci were selected based on low fixation index (F_{ST}) values and low global minor allele frequencies. In addition, certain SNPs were chosen from the literature for other purposes (biogeographic ancestry, surname and phenotype prediction, identification).[2-7] This panel of 18,147 loci was submitted to Ion Torrent™ AmpliSeq™ White Glove Design service, and primers were successfully designed for 14,934 loci (14,731 amplicons) in a single-tube amplification format. The resulting panel of primers was tested by sequencing DNA from 178 individuals, and 13,917 of these primers (approximately 93%) produced amplicons with adequate sequence coverage at default quality scores. Within the large primer panel containing 13,917 primer pairs, only a subset of loci is appropriate for analysis of DNA mixtures. The panel to be used for mixture analysis was selected from the larger primer panel based on specific criteria ($F_{ST} \leq 0.08$, minor allele frequency ≤ 0.3, number of reads within a ten-fold range, strand bias ratio >0.5, minor allele ratio for homozygous major loci ≤ 0.005, no Mendelian errors, distance of $\geq 500{,}000$ base pairs between loci). Primers meeting these criteria created a mixture panel targeting 2,655 loci, of which 2,311 are used for mixture analysis, while an additional 344 loci were included for identification and biogeographic ancestry prediction.

A custom software platform, IdPrism, was developed with a simple user interface and is composed of modules that identify sequence variants (minor alleles), as well as perform identification, mixture analysis, determine familial relationships, phenotypes, and Bio-Geographic Ancestry (BGA).[8-13]

Two independent laboratories conducted experiments with both controlled, laboratory generated mixtures, and realistic touch samples. A dilution series of single-source samples demonstrated the sensitivity of MPS, where reference profiles can be identified 100% of the time with 20pg of input and 50% of the time with 5pg input. In addition, the minor contributor can be identified at a minor to major ratio of 1:100 for 1st degree relatives. Several mixtures were prepared from saliva specimens to simulate complex forensic mixtures (e.g., two- to ten-person mixtures) in various proportions and total template amounts to identify potential limitations of the method. Results from these mixtures demonstrate the ability to resolve up to ten-person mixtures with as low as 1ng total input DNA as well as 1st degree relatives. In addition, results from touch samples demonstrated that as many as ten people can be identified in a single mixture, depending of the detection threshold that is applied. Even at a stringent threshold of Probability of Random Man Not Excluded P(RMNE) of 10×10^{-9}, 35 touch samples with four or more individuals were correctly identified as contributors.

Reference(s):

1. Voskoboinik, L. and A. Darvasi. Forensic identification of an individual in complex DNA mixtures. *Forensic Sci Int Genet*, 2011. 5(5): p. 428-35.
2. Kidd, J.R. et al. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investig Genet*, 2011. 2(1): p. 1.
3. Kidd, K.K. et al. Expanding data and resources for forensic use of SNPs in individual identification. *Forensic Sci Int Genet*, 2012. 6(5): p. 646-52.
4. Walsh, S. et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet*, 2013. 7(1): p. 98-115.
5. Pakstis, A.J. et al. SNPs for a universal individual identification panel. *Hum Genet*, 2010. 127(3): p. 315-24.
6. van Oven, M. et al. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum Mutat*, 2014. 35(2): p. 187-91.
7. Flaquer, A. et al. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet*, 2008. 16(7): p. 771-9.
8. Isaacson, J. et al. Robust detection of individual forensic profiles in DNA mixtures. *Forensic Sci Int Genet*, 2015. 14: p. 31-7.
9. Ricke, D.O. et al. GrigoraSNPs: Optimized Analysis of SNPs for DNA Forensics. *J Forensic Sci*, 2018. 63(6): p. 1841-1845.
10. Ricke, D.O. et al. Estimating Individual Contributions to Complex DNA SNP Mixtures. *J Forensic Sci*, 2019.
11. Helfer, B.S., P. Fremont-Smith, and D.O. Ricke, 2019.
12. Ricke, D.O. et al., 2018.
13. Ricke, D.O. et al., 2017.

DNA Mixtures, Touch DNA, DNA Sequencing

*Presenting Author