Optimization of Microhaplotypes for Advanced DNA Mixture Deconvolution

Technical Summary
Bode Technology
December 31, 2025

Report submitted to the National Institute of Justice
Grant # 15PNIJ-22-GG-04393-DNAX
Project Period: 01/01/2023 to 12/31/2025
Award Amount: $607,799

Sarah Cavanaugh – Lead Researcher (Bode Technology)
Daniele Podini, PhD (George Washington University)
Andrew Feldman, PhD (NexGen Forensic Sciences)
Kenneth Kidd, PhD (Professor Emeritus, Yale University)
Jonathan M. Davoren – **Principal Investigator** (Bode Technology)
Bode Technology
10430 Furnace Rd, Suite 107
Lorton, VA 22079
Jonathan.Davoren@bodetech.com

This project was supported by Award No. 15PNIJ-22-GG-04393-DNAX, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice.

Table of Contents

Project Summary

Abstract

Detection of minor DNA components in biological mixtures has increased as molecular techniques have become more sensitive. Accordingly, mixture deconvolution has become a major concern and topic of debate in the forensic DNA community. Short tandem repeat (STR) profile data generated with capillary electrophoresis and massively parallel sequencing (MPS) are subject to inherent issues that complicate mixture deconvolution, such as PCR stutter artifacts and allele length biases. Deconvolution may be improved by sequencing microhaplotypes as they are not subject to the amplification noise artifacts and stochastic effects that impact STRs. Before microhaplotypes can be implemented in casework, the following considerations should be addressed: definition of a consistent panel of microhaplotype loci; increased population studies to determine relevant haplotype allele frequencies; incorporation of advanced sequencing technologies into forensic laboratories; development of user-friendly bioinformatic analysis and mixture deconvolution methods; and assessment of the infrastructure requirements necessary to build a searchable microhaplotype criminal database. Ultimately, implementation requires a thorough understanding of the benefits and limitations of the new marker set as well as the practical investment of time and effort put forth for its adoption by the laboratory [1].

By coupling a highly discriminatory microhaplotype MPS assay with probabilistic genotyping methods such NexGenID, a novel software platform optimized for mixture deconvolution and probabilistic genotyping of sequence data, or EuroForMix [2], a widely used open-source probabilistic genotyping software modifiable for use with microhaplotype sequence data, microhaplotype analysis may be efficiently implemented by practitioners. The proposed microhaplotype panel demonstrated high discriminatory power with combined match probabilities ranging from 9.53E-52 to 4.79E-63 and the ability to infer biogeographical ancestry. The assay proved to be sensitive down to 50 pg inputs and applicable to inhibited or degraded trace samples. Application to complex DNA mixture samples demonstrates the assay's potential to exceed minor-contributor detection when compared to STR deconvolution, help solve complex cases, increase the number of samples considered suitable for comparison, and enable retesting of cold cases where a minor contributor was assumed present but was not suitable for comparison. This effort also identified novel research areas that were technically explored to inform future solutions for the forensic community.

Goals and Objectives

In two phases, this study sought to develop, optimize and assess an wet-lab assay and MPS workflow with an analysis package for improved mixture deconvolution using microhaplotypes. Analysis of microhaplotype sequence data from complex mixtures was performed with NexGenID. These outcomes were then compared to analyses of microhaplotypes data adapted for interpretation within EuroForMix.

Research Questions

Phase I sought to determine an optimal wet-lab workflow, applicable in a forensic context to target amplify microhaplotype loci from a range a DNA sample types. This question was addressed through the following objectives: down-selection of microhaplotype loci optimal for individualization and mixture deconvolution; construction of a wet-bench target assay; haplotyping of

donor samples to obtain expanded population allele frequency data; and biostatistical assessment of the projected performance of the compiled set of microhaplotype loci.

Phase II sought to assess methods of interpretation and ultimately deconvolution of complex DNA mixtures genotyped with a microhaplotype assay. This question was addressed through the following objectives: evaluating the benefits and limitations of mixture deconvolution and probabilistic genotyping using the microhaplotype wet-bench assay with Illumina sequencing and NexGenID analysis by applying the workflow to in vitro mixtures and constructed mock evidence and comparing outcomes from NexGenID to analyses of microhaplotypes data adapted for interpretation within EuroForMix.

Summary of Project Design and Methods

Panel selection and primer design

The primary objective was to identify the most suitable microhaplotypes (MH) for a comprehensive panel for mixture deconvolution by identifying a set of microhaplotype loci with the highest possible "Ae" values (effective number of alleles), a measure of allelic variation within a given population [3]. To achieve this, a multi-step approach was employed beginning with a systematic analysis of MHs within the MicroHapDB [4] database. This database consolidates markers and population frequency data from various published research articles using allele frequencies estimated from publicly available whole genome data. Initially, the MHs were sorted based on their Ae and filtered for lengths < 270 base pairs and Ae values ≥ 5. From this narrowed-down pool of MHs, several criteria were assessed for each marker including the standard deviation of Ae values from the different populations, the range of Ae values (minimum and maximum), the presence of insertions and deletions, the combined frequency of insertions and deletions, the occurrence of homopolymers longer than 5 nucleotides within the MH span, and the BLAST search results for each marker (to avoid multicopy regions). Initially, approximately 250 markers were identified, some of which had been previously evaluated by Bode Technology and the Kidd lab [5]. For any markers that demonstrated a high Ae value but were too long, the locus was evaluated for length reduction, i.e., removal of one or more SNP sites to produce smaller amplicons. Ae values were re-calculated for modified regions, and any marker that showed a signification reduction in allelic diversity was further eliminated.

Assay primer design was completed through Illumina DesignStudio, a free web-based assay design tool that uses target regions of interest and preloaded reference genomes to generate optimal multiplex primer designs to maximize target region coverage and amplification efficiency, and in collaboration with Thermo Fisher Scientific. Primers were designed against the GRCh38 reference genome using parameters for AmpliSeq for Illumina Gene assay design in a single PCR reaction with a maximum amplicon length of 350 bp.

Allele defining SNPs annotated in NCBI dbSNP 153 (https://www.ncbi.nlm.nih.gov/snp/) with minor allele frequencies >1% were identified in the determined amplicon regions, and haplotype frequencies, Ae values, and heterozygosity for all loci were calculated based on phased SNP genotypes obtained from the 1000 Genomes Phase 3 sequence data in the UCSC Genome Browser (http://genome.ucsc.edu/) [6,7].

Population Samples

A set of 240 test samples was identified from a donor collection of nearly 500 individuals previously collected under IRB and housed at GWU. These donors represent eleven biogeographical populations. Additional donors were identified from purchased blood bank samples previously obtained for internal validation projects.

For sensitivity testing of the final assay, 2800M Control DNA (Promega Corp, Madison, WI), NA24385 (NIST RM8391), and two purchased blood samples were serial diluted to test inputs of 2 ng, 1 ng, 0.5 ng, 0.1 ng, 0.05 ng, and 0.025 ng. Each dilution was evaluated in triplicate libraries.

In Vitro Complex Mixture Sample Construction and Sequencing

Four categories of complex mixtures were constructed *in vitro* from aliquots of the population donor samples utilized in Phase I. Table 1 details the contributor ratios of each mixture constructed for each of the below described categories.

Category 1 – Limit of Detection: mixtures of donors with disparate biogeographical ancestry and in combinations of 2, 3, 4, and 5-contributors at various ratios were constructed to evaluate detection limitations of microhap analysis with the goal of deconvolving minor contributors at <5% contribution. Each mixture was constructed to target a 0.5 ng, 1 ng, and 5 ng DNA input and was processed in triplicate (total n=99). MPS technology allows for targeting of higher DNA inputs that may overload CE analysis [8], so 5 ng DNA was included to assess improved minor contributor detection in quantities that exceed the standard STR-CE input of 0.5–1 ng. Quantities below 1 ng were used to further demonstrate the expected limitations of minor contributor detection for low-level samples.

Category 2 – Number of Contributors: The second category of mixtures evaluated limitations in determining number of contributors to a mixture in the presence of allele sharing and related contributors, an issue that has been noted in the 2016 PCAST report [9] and the NIST Foundational Review [1]. In vitro mixtures were constructed containing 3, 4, and 5 contributors and included a father/child relative pair identified within the population samples housed at GWU. The additional contributors were selected from the population samples used for category 1 mixtures. A total of six mixture ratios were processed in triplicate with 1 ng and 5 ng DNA input (n=36).

Category 3 – Shared Stutter: STR-based mixtures can be challenging when the minor contributor(s) have loci that are of similar signal strength to stutter from the major contributor, leading to true alleles being indistinguishable from stutter. For category three mixtures, Bode custom software was used to engineer virtual mixtures of 3, 4, and 5 contributors from the Phase I population samples and additional known donors to have a range of allele and stutter sharing between contributors based on previously generated STR genotypes. Virtual mixtures were then constructed *in vitro* and genotyped with both STR-CE and microhap sequencing. A total of six mixture ratios were processed in triplicate with 1 ng and 5 ng DNA input (n=36).

Category 4 – Imbalanced Degradation: The fourth category of mixtures evaluated complications due to imbalanced degradation of donors. In vitro mixtures of 3, 4, and 5 contributors were constructed where one or more contributors were subjected to UV degradation

prior to mixture construction. A total of three mixture ratios were processed through library in triplicate with 1 ng and 5 ng DNA input (n=18).

For comparison and confirmation of expected challenges, all constructed mixtures were first processed with STR-CE analysis as follows: amplification of a 1 ng input with Globalfiler full volume reactions, capillary electrophoresis fragment separation on Applied Biosystems 3500 xl Genetic Analyzer, and data analysis with GeneMapper IDX following internally validated SOPs. STRMix v2.9 was used to deconvolve and interpret STR-CE mixtures following internally validated SOPs.

Table 1: Description of Contructed Mixture Ratios

Mixture Set	DNA Source	Number of Contributors	Mixture Codes	Mixture ratios (%)	DNA Inputs (ng)	Replicates	Total Samples
Category 1: Limit of Detection	Extracted DNA	2	Mix1; Mix2	95:5; 100:1	0.5; 1; 5	3	18
		3	Mix3; Mix4; Mix5	60:30:10; 80:15:5; 72:18:9	0.5; 1; 5	3	27
		4	Mix6; Mix7; Mix8	63:25:9:3; 40:39:16:5; 55:21:18:6	0.5; 1; 5	3	27
		5	Mix9; Mix10; Mix11	61:24:10:4:1; 36:34:24:4:2; 45:17:17:15:6	0.5; 1; 5	3	27
Category 2: Number of Contributors	Extracted DNA	3	Mix12; Mix13	76:10:14; 80:15:5	1; 5	3	12
		4	Mix14; Mix15	43:35:11:10; 50:42:6:2	1; 5	3	12
		5	Mix16; Mix17	40:21:17:15:7; 47:17:17:15:4	1; 5	3	12
Category 3: Improvement from alleles shared at stutters	Extracted DNA	3	Mix18; Mix19	76:14:10; 80:15:5	1; 5	3	12
		4	Mix20; Mix21	43:35:11:10; 50:42:6:2	1; 5	3	12
		5	Mix22; Mix23	40:21:17:15:7; 47:17:17:15:4	1; 5	3	12
Category 4: Imbalanced Degradation	Extracted DNA	3	Mix24	80:10:10	1; 5	3	6
		4	Mix25	30:30:30:10	1; 5	3	6
		5	Mix26	39:21:17:15:7	1; 5	3	6

Construction and Collection of Mock Evidence Complex Mixtures

To construct the mock touch evidence mixtures, a total of nine participants were identified. Each provided informed consent and a buccal sample to generate a reference genotype. Touch transfer DNA samples containing 3–5 contributors were constructed by having donors handle items relevant to gun crimes: handgun frames, handgun magazines, rifle bolts, and 9 mm bullets with brass cartridge cases. All substrates were decontaminated prior to handling via UV decontamination and/or bleach cleaning. Donors were instructed to wait two hours after washing their hands before handling the items with their dominant hand for 20 – 60 seconds. After handling, the 9mm bullets were loaded individually into a gun chamber and fired. Fired cartridge casings were collected and individually packaged prior to sampling.

Firearm substrate (non-casing) transfer DNA samples were collected with wet/dry nylon flocked swabbing. DNA extraction was performed using the Qiagen EZ1&2 DNA Investigator kit in 500 µl Large Volume reactions following internally validated SOP. The rinse-and-swab collection method was performed on fired cartridge casings according to Bille *et al* [10]. Cartridge casing samples collected via the rinse-and-swab method were extracted using the modified QIAamp DNA Investigator Kit method described by Bille *et al* [10]. DNA extracts were concentrated with Microcon DNA Fast flow filter units (MilliporeSigma, Burlington, MA) prior to quantification by Quantifiler Trio DNA Quantification Kit in 11 µl reaction volumes to assess DNA concentration, DNA degradation, and inhibition related to the various substrates. Donor references were amplified with 1 ng DNA. All recovered DNA from mock evidence mixture samples was targeted for amplification and library preparation with the AmpliSeq microhap assay as described below.

In addition to transfer evidence samples, a set of inhibited mixture samples was constructed. A 1 ng aliquot of NIST RGTM S8 3-persom mixture was combined with humic acid at concentrations of 50 ng, 150 ng, and 250 ng to examine amplification with the AmpliSeq reaction buffer in the presence of an inhibitor.

Library preparation and sequencing

All reference samples, sensitivity samples, constructed mixtures, and mock evidence samples were amplified using the custom microhaplotype AmpliSeq primer mix and AmpliSeq Library PLUS for Illumina prep kit following the manufacturer's recommendations for Ampliseq for Illumina Custom Panels with one primer pool. First, DNA samples were target amplified in 20 µl reactions with amplification parameters of: 99 ˚C for 2 minutes, 23 cycles of 99 ˚C for 15 seconds and 60 ˚C for 4 minutes, and a final hold at 10 ˚C for up to 24 hours. Amplicons were then partially digested with 2 µl of FuPa Reagent on a thermal cycler as follows: 10 minutes at 50˚C, 10 minutes at 55˚C, and 20 minutes at 62˚C. Next, AmpliSeq CD Index i7 and i5 adapters were ligated to the partially digested amplicon as follows: 30 minutes at 22 ˚C, 5 minutes at 68 ˚C, and 5 minutes at 72 ˚C then purified with AMpure XP beads. After a second library amplification: 98 ˚C for 2 minutes, 7 cycles of 98 ˚C for 15 seconds and 64 ˚C for 1 minute, and a final hold at 10 ˚C for up to 24 hours, libraries were purified with AMpure XP. Finally, libraries were quantified with the Invitrogen™ Qubit™ dsDNA HS assay on the Qubit 4 fluorometer and sized using the Agilent TapeStation 4120 and D1000 ScreenTapes (Agilent Technologies, Santa Clara, CA). For Illumina sequencing, libraries were diluted to 4 nM and pooled in equimolar proportions. All pools were diluted to a loading concentration of 9 pM with a 2% PhiX sequencing control, per manufacturer's recommendations. Cluster generation and 2x300 paired-end sequencing were performed on the MiSeq FGx system using MiSeq v3 (600-cycle) reagents. Libraries were pooled in groups of no more than 40 to ensure adequate depth of coverage for all loci in the sample library.

Population Samples Data Processing and Biostatistical Analysis

Genotype analysis of the population samples was first performed as follows: mapping of sequence data to hg38 Canonical reference was performed with bwa mem in Galaxy (usegalaxy.org). The resultant .bam files were further processed for microhap genotype calling using mh.jar, a JAVA-based application available by request from Thermo Fisher Scientific and adapted for the current microhaplotype assay.

Haplotype frequencies, heterozygosity, match probability, and power of discrimination were calculated using Forensic statistics analysis toolbox (FORSTAT, https://fdl-uwc.shinyapps.io/forstat/) [11]. The Ae value is used to estimate the probability that more than two different alleles will occur in a DNA mixture of two unrelated individuals and was calculated using the following formula: $Ae=1/\Sigma p_i^2$,where p_i = frequency of alleles. Informativeness (I_n) for measuring allele frequency differences among populations was calculated according to Rosenberg *et al* [12].

Genotypic linkage disequilibrium (LD) was assessed using exact log-likelihood ratio tests in Genepop [13] on the phased genotype data for 43 microhaplotype loci from 2,504 individuals across 26 populations in the 1000 Genomes Project. Rare alleles, including singletons and alleles with very low frequencies, were collapsed into a single pooled category prior to testing to reduce sparsity, and exact genotypic LD tests implemented in the Genepop framework were used, providing unbiased p-values without relying on asymptotic assumptions.

Principal Component Analysis to visualize the first three PCs was carried out in Paleontological Statistics (PAST) [14] software v4.15. Population structure was evaluated via STRUCTURE analyses run on all 1000 Genomes 26 populations samples from K=2 to K=9 in STRUCTURE v2.3.4 [15]. The program was run 20 times at each K level with 10K burn-in and 10K Markov Chain Monte Carlo (MCMC) iterations under the standard admixture model. The result with the highest likelihood of the 20 runs was selected to illustrate the results for a given K value. To assess ancestry inference from genotype results of a given donor, likelihood ratios were calculated by dividing random match probabilities (RMP) generated using allele frequencies of one population vs the RMP generated using allele frequencies of a second population (e.g., the RMP for a given 43-locus genotype if observed from the African American population was divided by the RMP of that 43-locus genotype if observed from the European population).

Additional comparative statistical analyses were performed in JMP® v18.0.1 statistical discovery software.

Data Processing and Mixture Deconvolution in NexGenID

NexGenID (NexGen Forensic Sciences, Columbia, MD), provides a comprehensive solution for the analysis of microhaplotype raw sequence data and was utilized to perform haplotype determination from raw .fastq files of reference samples, sensitivity samples, constructed mixtures, and mock evidence samples as follows: cluster amplicon sequences based on locus, perform a local alignment, and identify unique alleles based on identical sequence. Analytical and stochastic thresholds are determined based on NGS and PCR parameters and applied for identification of unique alleles above noise read clusters. Both thresholds are sample-specific, driven by input DNA quantity that dictates how many templates were initially added for amplification. DNA input quantity is a user-supplied value entered at the start of raw data processing.

Following raw data processing, NexGenID enables detection/exclusion analysis using a comprehensive set of tools to assist in matching complex-sample allelic data to reference allele data from a person of interest and/or allele data from individuals in a genotype database. Here, NexGenID also provides a probability (p-value) that alleles detected above threshold are false due to noise expectations at any given locus. Default thresholds are set conservatively to favor high specificity but can be relaxed to increase sensitivity for detection/exclusion analysis. To complete a detection/exclusion test, the mixture data can be screened against a local database of known genotypes. NexGenID reports the total number of consistent alleles and assigns a measure of statistical significance to the match. In this framework, statistical significance of matches scales inversely with sample complexity (i.e., number of contributors). This novel framework quantifies the analyst's intuition that the significance of an allele match for loci heavily populated with alleles in a complex mixture is lower than for sparsely populated loci due to random chance. Detection/exclusion analysis reports the detection confidence that matched alleles are True Alleles, not noise. Any database/reference genotype meeting or exceeding statistical significance ($p<10^{-6}$) and allele confidence thresholds is deemed "Cannot Exclude".

NexGenID next provides a toolkit for genotype resolution of up to five contributors using a novel fully continuous probabilistic genotyping methodology that combines a model-weighting approach based on information theory [16] in the framework of Akaike [17-19] where the "best" model is first selected to minimize information loss with respect to the unknown data generative process and is then used to make inferences from the data. Akaike's framework effectively

prevents introducing "invented" minor contributors that "overfit" the DNA mixture data by balancing improvement in the fit of a contributor-proportion model to the data with a penalty for increasing the adjustable parameters (i.e., the number of minor contributors) in the model. Importantly, this framework also provides metrics for probabilistic genotyping that are fully analogous to Markov Chain Monte Carlo model-weighting concepts familiar to forensic laboratory analysts but fine-tuned for sequence data analysis. The model-fitting problem is formulated as a Maximum-Likelihood-Estimation (MLE) problem, and the Akaike Information Criterion (AIC) model weights are used to quantitatively rank the complete set of alternative models to infer the genotypes and compute relevant statistics for probabilistic genotyping such as the model weights. These statistics include **Likelihood Ratios** (LRs) when matching solutions to a reference, locus Genotype **Odds Ratios**, and **Genotype Confidence** values, all of which are reported in NexGenID. Only genotypes with sufficiently discriminating evidence above a genotype confidence threshold of 0.995, were included in resolved high-confidence multi-locus genotype solutions. AIC allows ranking of models with different model order (i.e., differing numbers of adjustable parameters **such as proportions of different numbers of contributors)** within a single model set to **allow inference of the best model order** given the data and the mathematical understanding of the uncertainty included for the data, rather than choosing a model order *a priori*, or using *ad hoc* algorithms. The software additionally allows the analyst to solve using a specific number of contributors.

Importantly, the NexGenID deconvolution strategy incorporates models of data uncertainty that depend on input DNA quantity, per locus sequencing read depth, and PCR assay specific amplification efficiency parameters. Understanding of the expected variation of the measured proportions for the alleles at a given locus is critical aspect mixture resolution analysis and determining confidence values and relative weights for Likelihood Ratio (LR) when comparing to a reference genotype. Once deconvolved, evidentiary weights calculated for each allele-pairing at all loci were used in familiar probabilistic genotyping-based LR hypothesis testing to compare a reference/person of interest (POI) genotype to the deconvolved solution at all loci. If high-confidence genotypes were extracted and matched against known genotypes, a modified Random Match Probability (mRMP) was computed as if the genotype were obtained from a non-mixture sample. A database match is defined as concurrence with a database entry if inferred allele-pairs match at all loci meeting deconvolution-quality acceptance criteria, *and* if individual alleles for that database entry are detected among the sequence clusters at loci *not* meeting deconvolution-quality acceptance criteria. The key features of NexGenID are summarized in Figure 1A, and an example 4-person deconvolution summary page viewed within the GUI is shown Figure 1B.

Figure 1: The key features of the NexGenID methodology (A) and a NexGenID GUI summary page for a deconvolved 4-person mixture (B).

For continuity with existing approaches, all NexGenID deconvolution RMP results and LR values reported in this program used the usual population database allele counting approach to determine the RMP for a given allele at a locus. For novel alleles detected in our analyzed population DNA samples, the canonical 5/2N (where N is the database size) was used, similar to assignments for novel mtDNA haplotypes in mitochondrial DNA analysis. An alternative approach for computing allele frequencies was also investigated. This computational method is described further in *Summary of Results PHASE II.B – INVESTIGATION OF NOVEL COMPUTATIONAL APPROACHES* (pg 23).

Mixture Deconvolution with EuroForMix

In this study, likelihood ratios were calculated using EuroForMix v4.2.5 [2]. EuroForMix is an open-source software developed for the interpretation of forensic DNA profiles, particularly mixtures. It employs a continuous probabilistic model to calculate likelihood ratios (LRs) for competing hypotheses regarding contributors to a DNA mixture. EuroForMix incorporates peak height information—or sequencing coverage in the context of MPS—to account for variability in DNA quantity. The software requires allele frequency data, mixture profiles, reference profiles, and a defined contributor number as inputs. Population allele frequency files were obtained from previous analysis based on 1000 Genomes Phase 3, while allele and coverage data for mixtures and references were extracted from the results files following raw data processing with NexGenID so that identical allelic content was input to both solvers. All files were converted to the EuroForMix-compatible format, with sequencing coverage replacing RFU values. After importing the data, the Weight-of-Evidence option was selected to compute LRs under the specified hypotheses: Hp, where each contributor was considered individually as the sole person of interest, and Hd, where all contributors were treated as unknown, with the ground truth NoC provided as the number of contributors to the mixture. Default parameters were applied, including an analytical threshold of 75 reads, Fst correction ($\theta = 0.01$), and a drop-in probability of 0.05. The Quantitative LR (Maximum Likelihood) calculation option was used, and the resulting logLR values were recorded

Summary of Results

PHASE I – MICROHAPLOTYPE ASSAY DEVELOPMENT

Optimization of Microhap Assay and Sensitivity Analysis

The final sequencing assay targets 43 microhaplotype loci selected for their optimal applicability to mixture resolution (Table 2). The insert size (i.e., DNA region between 3' primer ends) ranges from 105 bp and 273 bp with an average of 191 bp while the total amplicon sizes range from 234 bp to 350 bp. Through iterative testing of pooling concentrations and MiSeq FGx sequencing, primer pair sequences and multiplexed concentrations were modified until an optimal primer pool was determined and utilized for all sample target amplification. For sensitivity testing of the final assay, four reference DNA extracts were serial diluted to produce the following inputs: 2 ng, 1 ng, 0.5 ng, 0.1 ng, 0.05 ng, and 0.025 ng. Each dilution was target amplified and sequenced in triplicate. The assay demonstrates high intralocus balance for all loci with average allele balance of 91% (+/- 7.13%). This assay has demonstrated sensitivity with full microhaplotype profiles obtained from an input of 50 pg (average locus depths of ~8300 reads), and >90% profiles achieved from an input of 20 pg (average locus depths of ~4800 reads). However, at 50pg or less of input

DNA, allele drop-in of non-specific allelic content (e.g., PCR-induced base misincorporations) was observed above 2% of the total read depth at several loci across all reference samples. Drop-in alleles of up to 10% locus read depth were observed at a 25 pg input. The sensitivity of the sequencing chemistry enables the detection of low-level DNA template concentrations that might typically not meet detection thresholds necessary for some fluorescent-based methods such as capillary electrophoresis.

Table 2: Compiled 43-Locus Microhaplotype Assay Details

Name	Chr	Target Region Chr_Start (hg38)	Target Region Chr_End (hg38)	Number of SNPs(>1% MAF)	Total Number Observed Haplotypes (35 populations)
mh01WL-070	chr1	7556518	7556824	4	16
mh01WL-007	chr1	107376734	107376951	5	27
mh01WL-010.v1	chr1	168483999	168484220	8	64
mh01KK-212.v7	chr1	202647449	202647673	8	74
mh01WL-005.v1	chr1	212172139	212172345	7	60
mh01WL-006.v3	chr1	236518812	236519096	7	46
mh02KK-022.v1	chr2	3168640	3168934	6	31
mh02SHY-001.v1	chr2	120356630	120356859	5	15
mh02WL-002.v2	chr2	204497431	204497656	5	26
mh02KK-014.v5	chr2	227659236	227659465	9	80
mh02WL-003_v2-2	chr2	235709744	235709855	5	27
mh03WL-006.v1	chr3	2598080	2598298	7	55
mh03FHL-001.v2	chr3	11583703	11583940	5	26
mh03USC-3qC.v4	chr3	196652921	196653145	8	60
mh04WL-052.v3	chr4	2303789	2304018	8	79
mh06WL-051.v2	chr6	1708616	1708839	7	52
mh06WL-017.v2	chr6	32663093	32663303	26	170
mh06WL-008.v1	chr6	119186700	119186912	6	38
mh07WL-004.v1	chr7	13234917	13235130	7	46
mh07WL-022.v2	chr7	115873343	115873582	7	45
mh08WL-037.v2	chr8	1874214	1874486	5	26
mh08KK-137.v3	chr8	31225714	31225932	6	53
mh08WL-058.v1	chr8	124670762	124670979	4	16
mh09SHY-001.v5	chr9	28517003	28517238	7	31
mh10WL-031.v2	chr10	3825884	3826122	8	41
mh11KK-180.v5	chr11	1669475	1669756	10	40
mh12SHY-001.v2	chr12	13420513	13420730	8	60
mh12KK-201.v1	chr12	27647357	27647574	10	86
mh12SCUZJ-0392651	chr12	127034307	127034521	5	16
mh13KK-213.v1	chr13	23191400	23191586	6	13
mh13USC-13qA.v2	chr13	32979334	32979553	10	95

mh13KK-217.v2	chr13	46291751	46292029	8	51
mh13KK-218.v6	chr13	53486647	53486864	4	16
mh13WL-001.v1	chr13	94252406	94252625	9	65
mh13KK-221.v3	chr13	101106717	101106941	8	56
mh15WL-001.v5	chr15	71629608	71629893	8	71
mh15WL-031.v2	chr15	98586727	98586965	4	16
mh16WL-030	chr16	6735402	6735695	5	20
mh16KK-259.v7	chr16	83940256	83940489	7	45
mh19SHY-001.v2	chr19	7698935	7699124	5	19
mh19USC-19qA.v2	chr19	33273754	33273980	6	18
mh20WL-022.v2	chr20	1936557	1936814	5	24
mh21WL-020.v3	chr21	34035272	34035498	5	14

Biostatistical Assessment of Assay Performance

Based on global calculations of Ae for the 26 populations, the average global Ae value of the selected loci range from 6.02 to 19.09. However, there are limitations to solely using 1000 Genomes data as the populations are tightly clustered and widely spaced, geographically. Thus, an effort was made to type additional intermediate populations. Comprehensive population analyses were performed on 240 reference samples. These samples were categorized into one of twenty-three different population names, as self-described by the original donor. When synonyms were used for names (e.g., Black and African American), the individuals were grouped under the same population. Population groups with less than 10 individuals were omitted. This resulted in 9 populations added to the 26 1000 Genomes populations: 35 populations total (Table 3). Average Ae and Rosenberg's I_n were calculated for all 35 populations. The scatterplot of Ae by I_n (Figure 2) shows that all Ae values are greater than 4.5 and all but three are greater than 5. The mean and median values are quite similar and the ranges across the 35 populations are quite large and fluctuate among the 43 loci. Cumulative match probabilities for these populations range from 9.53E-52 to 4.79E-63 with combined discrimination capacity exceeding 99.999999%.

Table 3: 35 Populations Tested with 43-locus Microhaplotype Assay

World Region	Population	Names	N
Africa, SubSahara	GWD	Gambians	113
	MSL	Mende,Sierra Leone	85
	ESN	Esan	99
	YRI	Yoruba,Ibadan	108
	LWK	Luhya, Kenya	99
Admixed Africa	ACB	Afro-Caribbeans,Barbados	96
	ASW	African-American,South West	61
	Black*		35
Europe	TSI	Toscani	107
	IBS	Iberians	107
	GBR	British	91
	White*		25
	CEU	Utah Residents,N&W Eur Ancestry	99
	FIN	Finns	99
SoCen Asia	PJL	Punjabi,Lahore	96
	GIH	Gujarati,Houston	103
	ITU	Telegu,United Kingdom	102
	Keralite		27
	STU	Tamils,SriLanka from UK	102
	BEB	Bengali,Bangladesh	86
NAsia	Khanty		29
East Asia	Korean		20
	JPT	Japanese,Tokyo	104
	CHB	Han Chinese,Beijing	103
	CHS	Southern Han Chinese	105
	CDX	Dai from Xishuangbanna,China	93
	Laotian		19
	KHV	Vietnamese,HoChiMinhCity	99
Oceania	Atayal		18
America	Pima,Mexico		17
	PEL	Peruvians,Lima	85
Admixed America	Hispanic*		31
	MXL	Mexican Americans,LosAngeles	64
	CLM	Colombians,Medellin	94
	PUR	Puerto Ricans	104

* Self-identified ethnicity synonyms grouped under same popluation

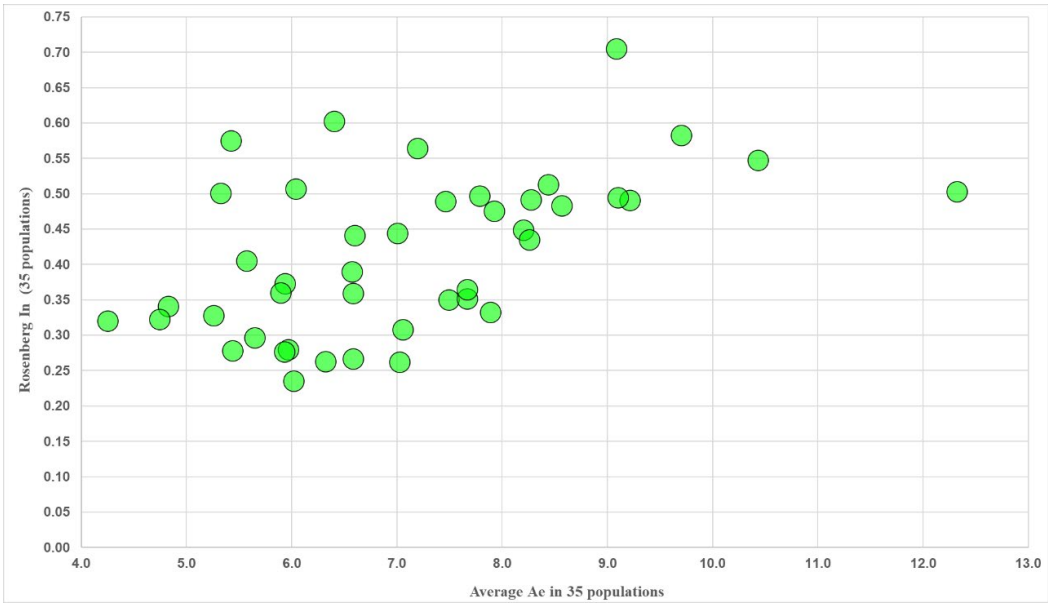

Figure 2: Plotting average Ae vs Rosenberg I_n for the 35 populations evaluated at 43 microhaplotype loci.

Principle Component Analysis (PCA) was run on the integrated 35 population data set. The first PC separates the African populations and the second and third result in a pattern of five clusters for the "Continental" groupings of populations: Africa, Europe, South Asia, East Asia, Americas. The admixed populations fall outside those five "Continental" groups most clearly visualized in the PC2 x PC3 plot (Figure 3). Also distinct is the Khanty which is the only Siberian population. STRUCTURE analysis supports the groupings of five main clusters with 5-cluster (K=5) runs producing the best results. These data indicate MHs could be informative for ancestry predictions. Likelihood ratio calculations of a given sample belonging to one population vs another population demonstrate LR >1 in support of the profile being observed in the donor's known population of origin. While the constructed microhaplotype panel is demonstrating powerful individual discriminatory power and population discrimination, genotyping of more populations is recommended to allow deeper representation of globally relevant variation, such as Middle Eastern, East African, Western Siberian, and Amerindian population groups.

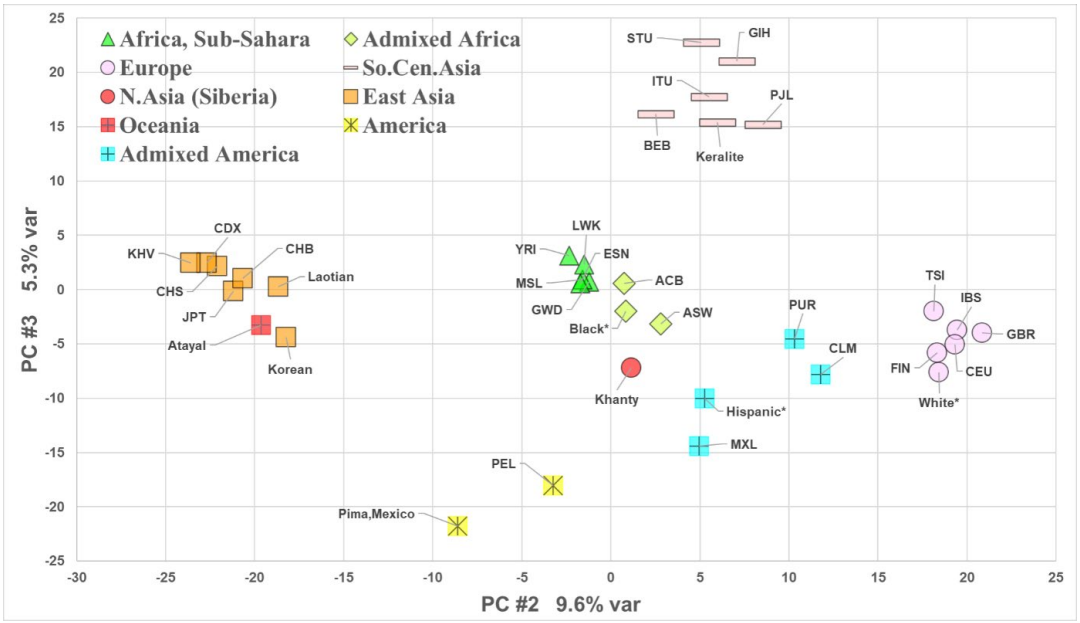

Figure 3: Principle Component Analysis (PCA) of 35 population evaluated at 43 microhaplotype loci. PC #2 vs PC#3 demonstrates clustering of the populations into the main "Continental" groupings with admixed populations PEL and Pima Mexico falling outside a main grouping.

To properly apply genotype frequency match probability calculations, an assumption of no LD is applied to forensic loci. LD refers to non-random associations among alleles across loci and is influenced by evolutionary and demographic processes such as drift, recombination, and population history. Thus, genotypic LD was assessed for the 43-locus microhaplotype panel using exact log-likelihood ratio tests in Genepop, both at the population level and globally across populations. After Bonferroni correction for 903 pairwise comparisons, 13 locus pairs showed statistically significant global LD signals, of which only one pair was located on the same chromosome. Despite statistical significance, the observed LD patterns raise concerns regarding biological plausibility. The two loci located on the same chromosome are situated ~9 Mb apart, a distance at which LD is unexpected. Most significant associations involved loci on different chromosomes and were not consistently detected across populations; even the single same-chromosome locus pair showed LD in only two populations. These inconsistencies, together with residual sparsity, sampling variance, demographic heterogeneity, multiple-testing effects, and potential artifacts introduced by allele pooling, suggest that many detected LD signals likely reflect statistical noise rather than true biological dependencies. In multi-allelic systems like microhaplotypes, LD testing is complicated by high-dimensional and sparse contingency tables caused by many low-frequency alleles, which undermine the validity of standard asymptotic tests. Nevertheless, it cannot be excluded that some of the detected LD signals are genuine. Additional analyses—including alternative LD metrics, sensitivity analyses of allele-pooling thresholds, demographic modeling, and cross-dataset validation—are required before any locus pairs can be confidently interpreted as being in true linkage disequilibrium. If confirmed, such LD could be mitigated by pruning (i.e., removing the less informative of the two loci found to be in LD) or explicitly accounted for in downstream statistical analyses.

PHASE II.A – MIXTURE DECONVOLUTION APPLICATION TO MICROHAPLOTYPES

Expanded Testing of Complex Mixtures for Deconvolution using Microhap Loci and NexGenID

Microhap implementation will only be considered by casework practitioners if paired with probabilistic genotyping that reduces interpretation time and overcomes human-induced variation in interpretation of complex mixtures [20]. Thus, user-friendly bioinformatic analysis and mixture deconvolution methods that have been tested with real-world mixtures are needed. With *in vitro* constructed DNA extract mixtures, the deconvolution capability of this microhaps assay was assessed based on accuracy of estimated/observed contributor proportions, accuracy of allele pairing per locus at 99.5% confidence, and LR calculations for both true contributors and non-contributors.

A total of 189 constructed mixture samples containing between 2 and 5 contributors were processed in triplicate with the 43-locus microhaplotype assay using total DNA inputs of 0.5 ng, 1 ng, and 5 ng, sequenced on the MiSeq FGx and analyzed with NexGenID (n=716 amplifications). NexGenID estimated the correct number of contributors (NoC) for 84% of mixture. Proportion estimate observations for individual contributors (Figure 4) demonstrated significant variance from the expected, constructed proportions ($p<0.001$); however, this variability is not unexpected for such constructed mixtures as the DNA extracts were aged and likely exhibited forms of degradation not detected with Quantifiler Trio QC that impacted sampling accuracy and amplification efficiency of individuals. Inaccurate NoC estimation and thus exclusion of true minor contributors was observed when the original mixture contained a minor contributor with less than 2% proportion. In these instances, the mixture data best fits a model with support for one less donor than the expected number. For example, five-person mixtures presented the greatest challenges in resolving minor contributors (e.g., 61:24:10:4:1 or 36:34:24:4:2), occasionally not supporting the detection of the lowest contributor primarily as a consequence of drop out rather than software limitations. While alleles for such contributors were identified in the mixture data, often true alleles were sequenced at read coverages below the minimum allele threshold and may be classified as noise. Such samples typically did not demonstrate any loci with total allele counts to support an additional contributor (e.g., no loci with 9 or more alleles in an expected 5-contributor sample), and if forced to deconvolve for 5 contributors, the AIC information-theoretic framework applied by NexGenID returned statistical metrics indicating model overfitting with a 5th contributor.

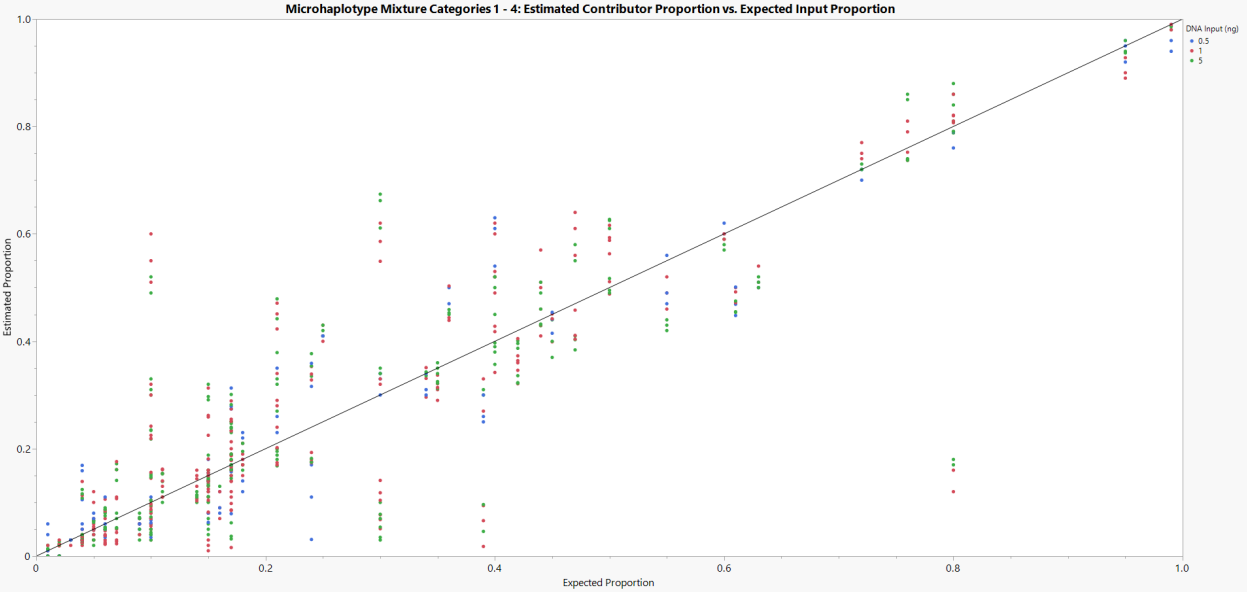

Figure 4: Comparing the expected donor proportion in constructed mixtures to the estimated proportion following data processing and mixture deconvolution with NexGenID. All mixtures analyzed are represented. Data points are colored based on the total DNA input of the sample into amplification.

In total across all mixtures, 797 of 841 expected contributors were correctly detected with detection significance exceeding the threshold of < 1E-06 and an average of 83 alleles observed in the mixture data. Of those donors detected, 757 contributors demonstrated LogLRs >1, indicating support for their corresponding genotype's inclusion to the sequenced mixture profile. When true minor contributors could not be excluded from the mixture, their LogLRs ranged from 0.963 to 89.905 (Figure 5). The impact of original DNA input can be seen as the range of LogLRs shift with increasing ng inputs. Note for both within a mixture and between replicates with different input DNA quantity, the LogLR variability is higher for lower input DNA quantity. An identified limitation of this assay through review of category 1 mixtures, specifically 2-person mixtures, is an unexpected increase in noise alleles, often related to overamplification of single base errors in short homopolymer stretches within the amplicon target region. Continued optimization of noise models will increase ability to identify noise reads.

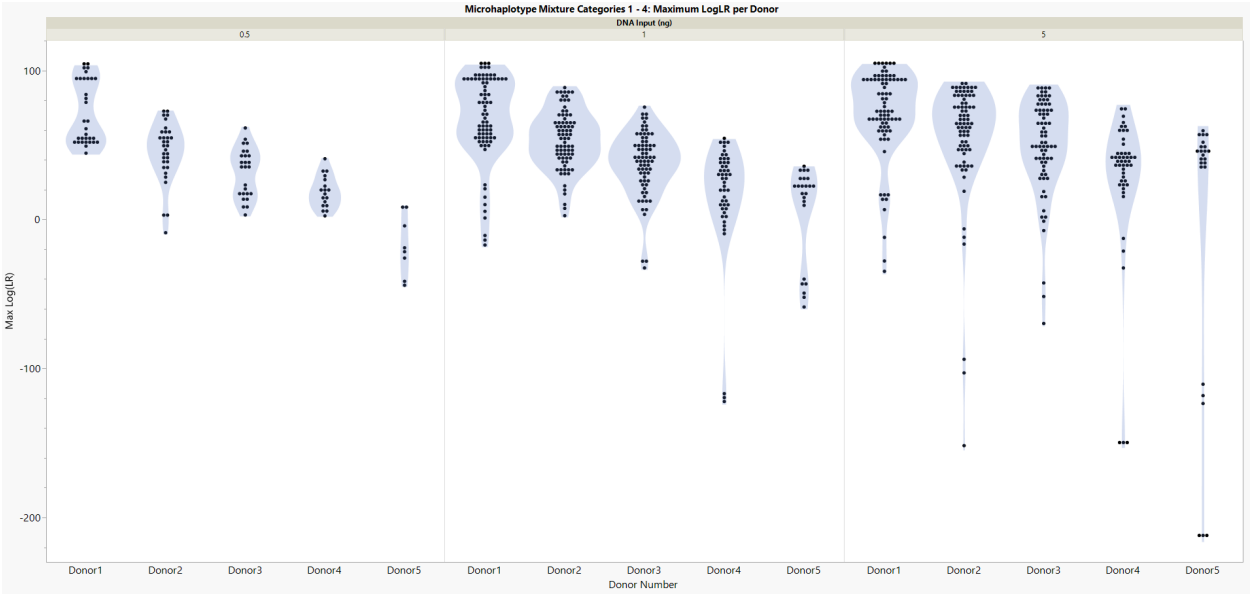

Figure 5: Violin plot of maximum LogLR obtained for all constructed mixtures based on donor order, separated by total DNA input for amplification: 0.5 ng, 1 ng, and 5 ng. Donor 1- Major contributor, added at highest proportion and decreasing to Donor 5 – added at lowest proportion in 5-person mixtures.

Evaluation of category 2 mixtures demonstrated the ability of NexGenID to correctly determine the number of contributors and deconvolve mixtures containing first degree relatives when the most minor contributor exceeds 2% contribution. This mixture set contained a father/child contributor pair in which the child was consistently the major contributor and the father was either the minor (3 and 4 person mixtures) or 2nd minor (5-person mixtures). Proportion estimates were consistently lower than expected for the father while the child's proportion was overestimated. In 13 of 36 mixtures, this impacted overall allele detection from the father contributor and subsequently deconvolution model allele pairing, resulting in low to no support (LogLR<0) for inclusion of the father's genotype to the mixture data. However, in the remaining 23 samples, the father's haplotype alleles were sufficiently detected with calculated LogLRs ranging from 4 to 43 to support inclusion of the father's genotype to the mixture data.

Category 3 mixtures were designed to assess the impact of stutter elimination with microhaps and NexGenID analysis on minor contributor deconvolution when mixutres may exhibit a range of sharing between major STR allele stutters and minor contributor allele. Issues caused by stutter artifacts are eliminated when using microhap markers, but this effect has not been quantified across a range of mixtures. Review of category 3 microhaplotype mixture results again demonstrated accurate detection and deconvolution of all true contributors when the most minor contributor exceeds 2% contribution. Specific improvements over STR analysis are discussed in the proceeding section. Additionally, this set of mixtures was evaluated for comparisons of non-contributor relatives as POI. The majority of non-contributor relative comparisons generated LogLRs <0, except one example where a first degree relative generated LogLR >3 when the true contributor was present in the mixture at a proportion of ~2% (70% of the donor's alleles were detected) and a full sibling of the non-contributor POI. Across all mixture ratios, no non-relative, non-contributors generated a LogLR > 0. Category 1 non-contributor test LogLR ranges are provided as example in Figure 6.

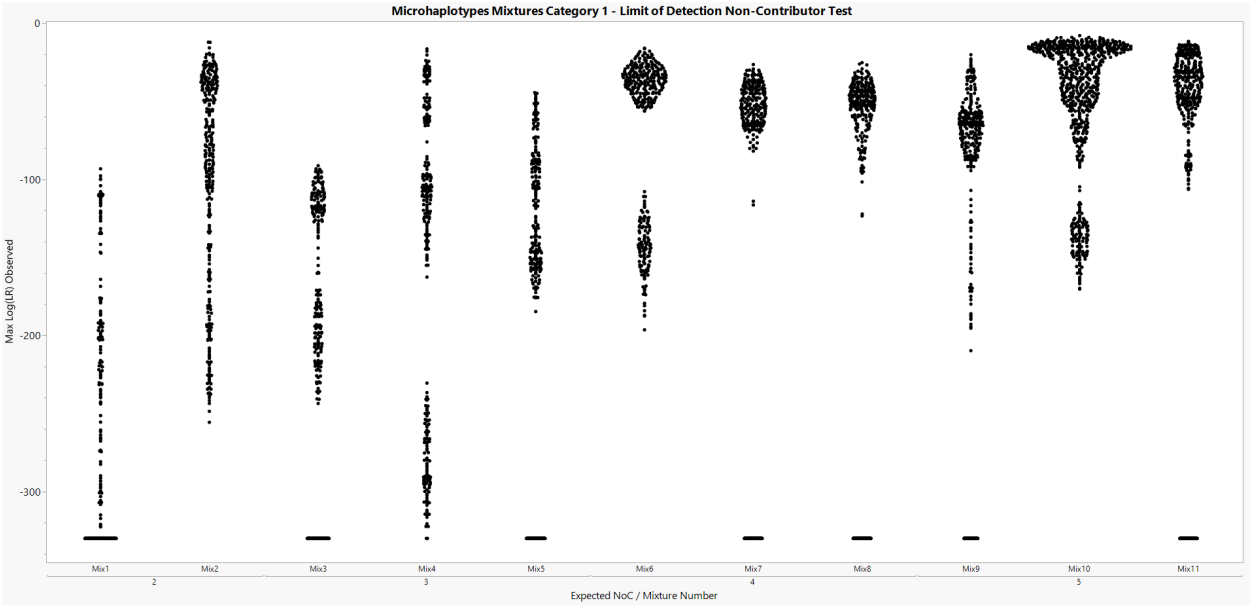

Figure 6: Non-contributor test performed for all Category 1 mixtures. No non-contributors of the 44 donors stored in NexGenID database generated a maximum Log(LR) exceeding 0 (LR=1). LRs equal to 0 (Log(LR) = ∞) are graphed as the minimum observed number.

Evaluations of category 4 mixtures highlight another current limitation of NexGenID. When mixtures of contributors with imbalanced degradation are processed with STRs, major/minor contributor assignment and genotype compilation can be impacted as shorter loci are preferentially amplified. This set evaluated the hypothesis that similar preferential amplification is not observed with the microhap assay when imbalanced degradation is present as alleles do not vary by size and amplicons are <300 bp. The imbalanced degradation had no effect on estimating NoC and full genotypes of all known contributors were detected in the mixture data. However, imbalances in allele coverage depth across loci impacted contributor proportion estimates; thus, leading to inaccurate deconvolution solutions. While the amplicon sizes of this assay range from 234 bp to 350 bp, UV degradation did not demonstrate a correlation between loci coverage and amplicon size. NexGenID would be improved by incorporating a degradation model capability.

Comparison of Microhaplotype Deconvolution with NexGenID to STR Deconvolution with STRmix

Comparison of LogLRs generated from microhaplotype genotyping to LogLRs generated from STR genotyping demonstrate ability improve inclusion strength for minor contributors. Mixtures constructed specifically to assess STR deconvolution vs MH when donors exhibit alleles that may fall into stutter positions emphasize this improvement (Figure 9). Minor contributors who cannot be excluded but produce LogLRs in a moderate/low support range can indicate strong support when the mixture is additionally analyzed with microhaplotype loci.

In general, the Log(LR) was higher when analyzed with microhaplotyes, especially for the Donors 1 and 2 in higher NoC mixtures. However, the corresponding STRmix Log(LR)s were also high for these contributors. If microhaps are utilized as a complementary assay to improve mixture deconvolution, the increasing benefits are minimal with large numbers. The benefit is observed

when STRs generate Log(LR)s below the range for very strong support. According to the ENFSI Guidelines for Evaluative Reporting in Forensic Science [21], -5 > LogLR > 5 provide very strong support for excluding or including a person of interest. Here, we observe instances of minor contributors to 3-person, 4-person, and 5-person mixtures where the STR Log(LR) gave weak support for inclusion or even no support (Log(LR)<0) of the profile to the mixture, but the ability to capture more alleles with greater allelic diversity using MHs produced Log(LR)s >10. The specific contributors exhibiting this improvement in Figure 7 represent mixtures from mixture category 2 and category 3. For example, Donor 2 of Mix13 generated logLR <0 with STRs, likely impacted by the presence of first-degree relatives in the 3-person mixture where Donors 1 and 3 were related. When using microhaps, the logLR of this contributor was calculated as > 60. Three additional mixtures from category 2 also demonstrated increased LogLRs for the non-related contributors when analyzed with microhaplotypes. From category 3 where contributors were selected based on allele sharing in stutter positions, four of the mixtures demonstrated increased LogLR calculations for the lowest proportions contributors (Donor 4 and Donor 5) through microhap analysis.

Figure 7: Scatterplot comparing the determined maximum LogLR for each contributor when using microhaplotypes (y-axis) and GlobalFiler STRs (x-axis). Mixture samples from categories 1, 2, and 3 are included. The LogLR ranges of -5 to 5 and -10 to 10 have been highlighted for STR values along the x-axis. Points are colored/shaped based on the contribution order of the donor: Donor 1- Major contributor, added at highest proportion and decreasing to Donor 5 – added at lowest proportion in 5-person mixtures. When sample falls above the line, Log(LR) is higher for a given contributor when using microhaps to analyze the mixture. Points below the line indicate Log(LR) was higher for a given contributor when the mixture was analyzed with STRs and STRmix.

In one instance, degraded major contributor showed no support of inclusion with a STRmix Log(LR) = -3 while with microhaps and NexGenID analysis, Log(LR) of 21 indicated support for inclusion of the genotype to the mixture.

These results demonstrate how highly discriminatory microhaps with high Ae value will be highly useful to deconvolution of minor contributors in complex forensic DNA mixtures, with the potential to exceed minor contributor detection in comparison to STR deconvolution.

Comparison of EuroForMix Deconvolution to NexGenID Deconvolution

The purpose of this task was to evaluate the consistency of results obtained with NexGenID by comparison to a well-established probabilistic genotyping system. EuroForMix is grounded in a continuous statistical framework that incorporates quantitative signal information into a likelihood-based inference paradigm. The software models the DNA mixture process using Bayesian principles, with inference primarily performed via maximum likelihood estimation (MLE) rather than full posterior sampling. Observed allele coverages are assumed to follow a gamma distribution, reflecting stochastic variation in PCR amplification, with expected values expressed as functions of contributor-specific DNA quantities, mixture proportions, and locus-dependent amplification parameters. The framework explicitly accounts for artifacts such as allelic dropout and degradation, with dropout probabilities typically modeled as functions of read coverage and analytical thresholds. However, unlike NexGenID, EuroForMix does not explicitly model sequencing-platform–specific noise artifacts, representing a limitation that would require thorough validation prior to routine implementation with sequencing data.

Likelihood ratios generated using EuroForMix were generally consistent with those produced by NexGenID, although a modest upward trend was observed (Figure 8). This difference is attributed to NexGenID's explicit modeling of sequencing noise, resulting in a more conservative interpretation framework (further discussed below in *Comparing LR Values Across Specimen Samples and Mixture Deconvolution Solvers* (pg. 25). While an extension of EuroForMix designed for sequencing data (MPSproto) is now available, it was not used in this study and will be the subject of future work. Importantly, this component of the project demonstrates that (1) genotype calls generated through NexGenID can be successfully exported and interpreted using independent probabilistic genotyping software, and (2) data produced by this

microhaplotype multiplex are compatible with multiple probabilistic genotyping frameworks already available to the forensic community.

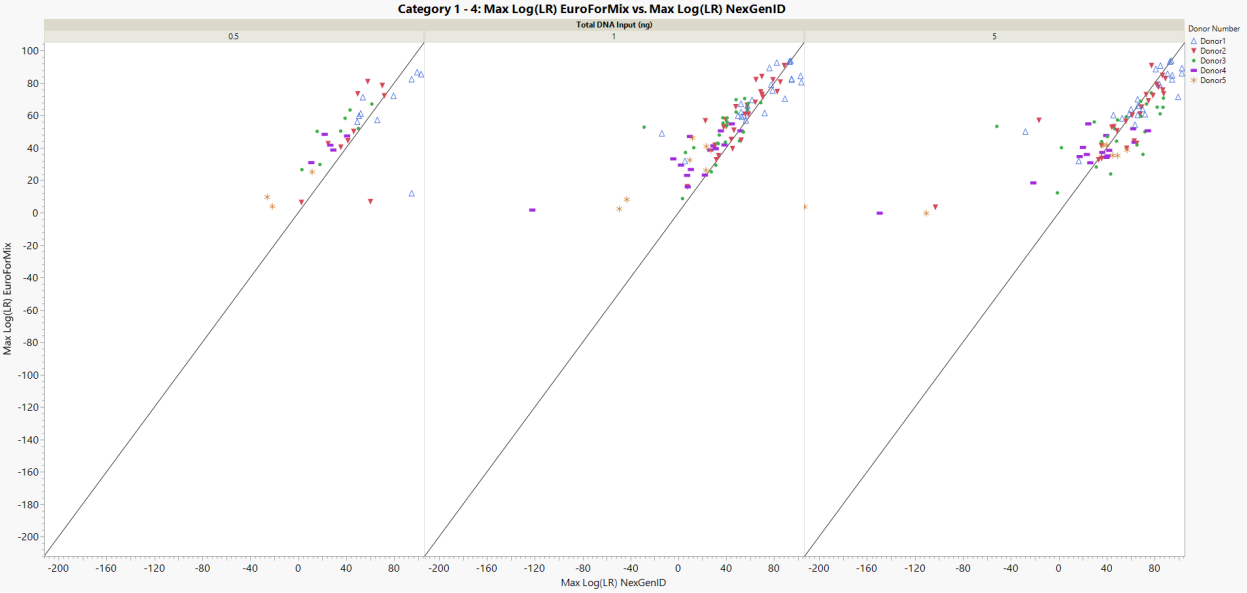

Figure 8: Scatterplot comparing the determined maximum LogLR for each contributor when using microhaplotypes with NexGenID (x-axis) and with EuroForMix (y-axis). Mixture samples from categories 1, 2, 3, and 4 are included. Points are colored/shaped based on the contribution order of the donor: Donor 1- Major contributor, added at highest proportion and decreasing to Donor 5 – added at lowest proportion in 5-person mixtures. When sample falls above the line, Log(LR) is higher for a given contributor when using EuroForMix to analyze the mixture. Points below the line indicate Log(LR) was higher for a given contributor when the mixture was analyzed with NexGenID.

Application of microhaplotype deconvolution to low-level, degraded mock evidence

Mock evidence samples were constructed to mimic complex trace evidence mixtures often observed in gun crime cases. Table 4 summarizes the results metrics for these samples including the total DNA obtained after sampling and extraction, library quantification results, total good quality sequencing reads obtained, and the best fit NoC model (observed number of contributors). Overall, DNA was successfully recovered and genotyped with the microhaplotype assay and analyzed in NexGenID for 16 of 18 mock samples and 2 of 3 inhibited samples.

While sample were constructed with known "touch/transfer" contributors, DNA deposition is highly variable, and so recovery of all contributors was unknown. These samples demonstrated recovery of contributors who deposited DNA and did not consistently correlate with the order in which they handled the items. For example, the major contributor was consistently the same donor, regardless of his place in handling order or length of time of handling. In a number of samples, a couple of a contributor's unique alleles may have been observed below AT, but there was little to no evidential support for inclusion of their genotype to the mixture after deconvolution, Log(LR)s <0. Additionally, secondary transfer from unknown contributors is hypothesized in several instances where detected alleles could not be attributed to any of the expected handlers. These results may be interpreted as a reflection of true number of contributors who "contributed DNA" to the sample, regardless of the number of individuals who held the item.

True contributors generated Log(LR)s >0 for all mixtures that were deconvolved, and no non-contributors generated Log(LR)s >0. Only three of the mock samples indicated the presence

of the expected number of contributors (deconvolved into the correct NoC). The remaining indicate NoC of one or two fewer, except the handgun frames where additional contributor contamination was observed. The additional contributors were consistent between the two replicates, but a source could not be identified. Inhibited sample consisted of a controlled 3-person mixture mixed prior to amplification with 50ng, 150ng, and 300ng humic acid suspension. The 300 ng humic acid sample failed to amplify. All donors detected and generated high Log(LR)s for the successfully amplified samples 50ng and 150ng HA.

Review of sequencing results shows the trade-offs in detection sensitivity of MPS for low-input degraded samples. Microhaplotype panel performance is more influenced by DNA input than DNA degradation. Sequence data was processed with NexGenID in which all sources of uncertainty and statistical variation are modeled from first principles based on the input DNA quantity specified by the analyst. As the software bases the analytical and stochastic thresholds on the total DNA input for amplification, this value needs to be carefully assessed. Because the target amplicons are sized closer to the large autosomal target of Quantifiler Trio, it is recommended to evaluate the total input of amplifiable DNA using the large autosomal quantification value. Using the small autosomal value over estimates the amount of amplifiable DNA, resulting in a lower analytical threshold determination. Subsequently, overamplified noise reads were evaluated as "true alleles", overestimating NoC, and decreasing the ability to deconvolve true minor contributors. The tradeoff, however, comes in that alleles belonging to true minor contributors may then fall below the determined analytical threshold.

Table 4: Mock Evidence Summary Metrics

Sample Name	Mixture Substrate	Quantifiler Trio Degradation Index	Total DNA Input (ng, Small Autosomal)	Total DNA Input (ng, Large Autosomal)	Library Quant ng/µl	Total Good Reads	Expected Number of Contributors	Observed Number of Contributors
9mm_casing_3pA	9mm brass casing	UND	0.0745	0.0048	FAILED			
9mm_casing_3pB	9mm brass casing	4.25	0.525	0.1235	0.475	374,243	3	2
9mm_casing_3pC	9mm brass casing	5.43	0.489	0.09	0.47	312,142	3	2
9mm_casing_4pA	9mm brass casing	2.57	0.845	0.3285	0.981	527,949	4	2
9mm_casing_4pB	9mm brass casing	4.25	0.55	0.1295	0.327	194,456	4	2
9mm_casing_4pC	9mm brass casing	5.41	1.15	0.2125	0.917	441,462	4	3
9mm_casing_5pA	9mm brass casing	8.62	0.56	0.065	0.514	251,199	5	2
9mm_casing_5pB	9mm brass casing	4.28	1.58	0.369	1.84	604,720	5	2
9mm_casing_5pC	9mm brass casing	3.65	1.24	0.3395	2.18	380,674	5	3
rifle_bolt_3p	Rifle Bolt	1.89	0.945	0.499	2.07	428,281	3	3
rifle_bolt_4p	Rifle Bolt	3.21	0.555	0.173	0.86	201,389	4	4
rifle_bolt_5p	Rifle Bolt	1.70	2.34	1.375	3.42	527,805	5	2
magazine_3p	Handgun magazine	4.10	1.12	0.2735	1.64	195,006	3	3
magazine_4p	Handgun magazine	2.83	0.3125	0.1105	0.78	170,918	4	3
magazine_5p	Handgun magazine	2.19	2.955	1.35	3.33	318,006	5	4
handgun_frame_3pB	Glock handgun frame	1.57	2.60	1.65	1.52	522,518	3	5
handgun_frame_4pB	Glock handgun frame	1.26	3.79	3.01	5.25	1,093,781	4	5
handgun_frame_5p	Glock handgun frame	1.26	3.11	2.46	0.791	FAILED		
3p_HA_300ng	Humic Acid Inhibited		1	1	FAILED			
3p_HA_150ng	Humic Acid Inhibited		1	1	1.71	767,268	3	3
3p_HA_50ng	Humic Acid Inhibited		1	1	3.95	740,769	3	3

In comparisons between deconvolution analysis using NexGenID vs EuroForMix, ten contributors across six samples generate LogLRs >0 (4 logLRs >5) with EuroForMix while NexGenID calculates LogLR<0 (all but 2 are LogLR < -7). For these donors, 40-90% alleles were captured, with <20% falling below AT. For example, a minor contributor to the 5-person in magazine had 78 alleles detected at 2% contributor proportion (2.11E-15 detection significance). A logLR = -23 was for this contributor in NexGenID, but a logLR = 16.6 was calculated with

EuroForMix. Visual review of this sample suggests this donor is exhibiting degradation that EuroForMix is better able to interpret as it incorporates degradation in the deconvolution model.

PHASE II.B – INVESTIGATION OF NOVEL COMPUTATIONAL APPROACHES

Over the course of this effort, two novel research areas were identified in response to optimization of microhaplotype deconvolution analyses within NexGenID. Technical exploration of these areas is discussed below, and the subsequent findings may inform future solutions for the forensic community should SNP-based microhaplotype loci find routine implementation.

Investigating an Alternative Approach to Compute Population Microhaplotype Allele Frequencies

As described in the Methods, NexGenID RMP and LR calculations used the usual population database allele counting approach with the 5/2N (where N is the database size) frequency assumption for rare allele observations. For many loci in our microhaplotype panel only a single instance of an allele was present in the 1000 genomes database, and thus assigning the appropriate random match probabilities to such alleles involves a high degree of uncertainty. This uncertainty problem is similar to that encountered assigning random match probabilities in Y-STR haplotype analysis, many of which are singletons or under-represented in Y-STR population databases. Given NexGenID's full sequence clustering approach, versus analysis of the fixed set of known polymorphic sites, the technology is highly sensitive to detecting novel alleles that will improve mixture resolution and discrimination.

To investigate a potential solution to assign allele frequencies for rare or novel alleles more rigorously and/or conservatively given limited available population data, we tested model fitting methods that use the available data to estimate the full joint probability distribution over the variant sites based on directed acyclic graph (DAG) probability decompositions. The DAG model is trained using the pairwise probabilities of occurrences of polymorphisms at the variant sites observed in population database data. The rationale for decomposing the probability in this way is that there are far more exemplars of the pairings than of the complete haplotypes, and the pairwise frequencies used to determine the joint probability decomposition model can thus be estimated with higher certainty than using haplotype counting of the rarer types. Once the graphical model is trained on the pairwise data, a specific haplotype's frequency can be inferenced, even if it was never previously observed. This pairwise model of the joint probability distribution over the variant sites is called a Chow-Liu (CL) decomposition and the tree model is optimally derived using maximum likelihood methods [22].

We investigated the use of the CL decomposition to not only assign a frequency to novel alleles but also as a tool to rationally derive the most conservative frequency value for an allele based on the available data. We considered a concept of implementation where complete haplotypes, inclusive of all amplicon sites, are used for the purpose of achieving high resolvability of the DNA mixture, but for the purposes of assigning the random match probabilities, we use ***only those variant sites*** for which the under-sampled allelic database contains statistically supportable information. In this approach, those sites for which there is insufficient statistical support are ignored (treated as untyped) when determining the allele frequency. The RMPs are computed via counting database sequences including only these truncated haplotypes per the usual approach, or using the truncated CL DAG tree model which is pruned based on information theoretical method described below.

Implementing the above concept requires a statistical method to determine which variant sites have sufficient "support" and which do not. For this purpose, we joined information theory with CL decomposition modeling of the joint probability distribution. The problem is cast as determining the optimal number of parameters in the CL tree model that avoids "over-fitting" in the information theory sense. This problem can be solved mathematically using the Akaike Information Criterion (AIC). It can be shown that for models with a variable number of adjustable parameters (in our case variant sites), the "best' model in the AIC information theory sense, is the one that minimizes the information lost when that model is used to approximate the unknown "true" generative process giving rise to the observed data. This is the model for which the AIC value obtained by maximum likelihood fitting of the model to available data is minimized. The unknown "true" model in our case is the population joint probability distribution of all haplotype frequencies at our locus but for which we only have the limited sample in our databases.

An example result from this investigation is shown in Figure 9 below. The nodes represent the variant sites in the haplotype and each edge connecting them has the mutual information value between those two nodes. In the plot (left), the edges in the full CL tree decomposition model for locus mh02KK-014 (shown in the inset, right) are pruned successively, edge by edge based on the mutual information (MI) value (from least MI to highest MI). The minimum AIC value of the CL model fit is obtained after pruning the tree until it only contained the highest four MI edges. This procedure eliminated the four least informative polymorphic sites from the tree and identified the sites with "sufficient statistical support". Allele frequencies are then computed based on only the remaining sites using the truncated CL model or by counting DB entries. The comparison of the truncated model-computed frequency values versus the database frequency values determined via counting using only the four high-information-content polymorphic sites is shown Figure 10. Our investigation revealed that this information-theory-based approach worked well across all microhaplotype loci in our panel. These results suggest that this procedure can be considered as a methodology to assign highly conservative (higher) frequencies to alleles, i.e., values that are the most favorable to the defense in criminal casework, and that are derived using only the rigorously defined most informative sites in the database to compute them.

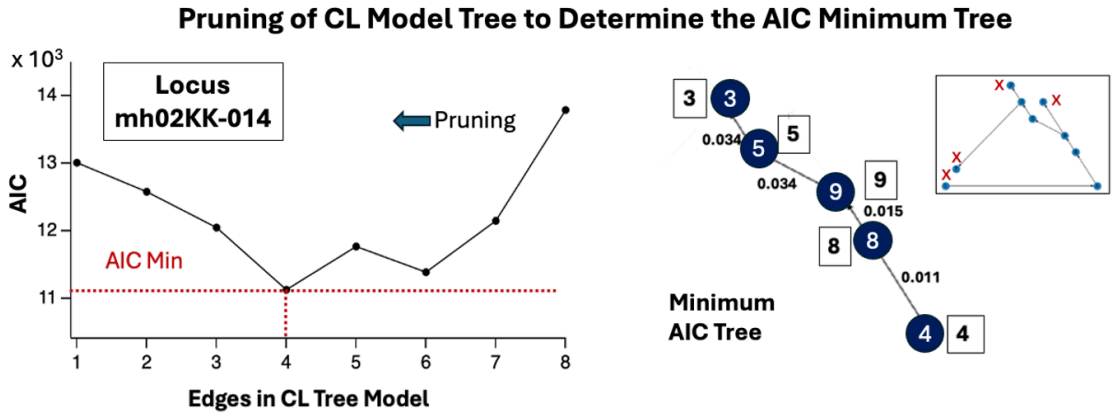

Figure 9: Variation of Akaike Information Criterion (AIC) versus edges in the Chow-Liu graphical model (inset) when eliminated in succession based on edge mutual information content. The CL model corresponding to the AIC minimum is shown on the right. The node labels are the consecutive polymorphic sites across the locus sequence.

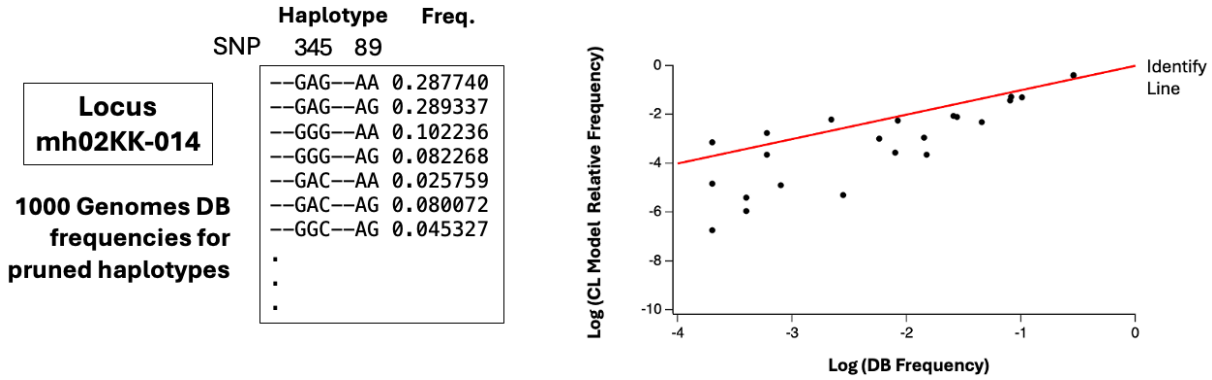

Figure 10: Depiction (left) of truncated haplotypes based on the AIC minimum CL model in Figure 3. Polymorphic sites with the dash are treated as untyped and only sites 3,4,5,8, and 9 are used to determine the allele frequency. The figure on the right compares the CL prediction based on the tree in Figure 3 to the counting based method using the informative sites only**.**

Comparing LR Values Across Specimen Samples and Mixture Deconvolution Solvers

When comparing LR values obtained for different mixture deconvolution methods for the same sample, a key consideration is how the respective approach weights the relative certainties of allele pairing for the POI, since this weight appears in the numerator of the LR value for each locus: the strength of evidence favoring the prosecution hypothesis. The quantification of the relative strength of evidence per respective allele-pairing is greatly dependent on the internal model used to represent the expected variability of the allelic signals at their observed total signal fractions. In MCMC methods, relative weights reflect the relative number of iterations in which the allele pair is visited during execution of the Markov Chain used to deconvolve the sample into its constituent multi-locus genotypes. In NexGenID, we derive the weightings analytically using the Akaike Information Theory framework to rank models and compute normalized relative weights. In the AIC framework the likelihood function used to compute each AIC value incorporates the explicit *de novo* model of the expected variability of allele signal based on the input quantity of amplifiable DNA.

The statistical variation of allele templates when sub-sampled from the tube that contains the entire extract DNA into the PCR reaction at each locus establishes a lower limit on expected variability of the signal (due to Poisson fluctuations). Other factors increasing variability above this lower limit include amplicon production variability due to <100% amplification efficiency, PCR base substitution errors impacting clustering of identical sequences into allele candidates, read depth, which determines sampling of amplicons from the amplified pool onto the sequencer chip, and base calling errors during sequencing. In NexGenID, these factors are not fit to the data but are computed *ab initio* for the end-to-end process based on the process parameter values, such as polymerase error rate, amplification efficiency, etc. As this process-based approach is unique to NexGenID, we thus must expect different weighting and LR values when compared to other solvers due to the different (and less NGS-explicit) representation of expected data variability. This variation is observed when comparing NexGenID to EuroForMix LR values on identical input data, EuroForMix attempts to estimate the noise per each sample. Our perspective is that this is prone to overfitting and that noise parameters should not be fit along with the deconvolution

solution per each sample analyzed but set *ab initio* from the end-to-end sample processing process whose parameter values should remain fixed during the deconvolution fitting.

One additional consideration in LR value interpretation related to allelic signal variation is its impact on the sample-to-sample LR values when technical replicate subsamples are drawn from the same DNA extract for end-to-end analysis. This variability brings to the fore the debate of the meaning and utility of reporting a confidence interval for an LR. The LR value is a quantity derived from a fit of a model to the specific instantiation of the data. While this fit may be optimal for the instance, it is prudent to quantify to what extent the LR of this optimal solution for the instance is reflective of that which could be obtained by averaging multiple technical replicates from the whole extracted DNA specimen. When data is obtained from a subsample of a specimen it has a bias, as this bias will be reflected in the model fit parameters obtained for this subsample. To interpret this one result as reflecting the composition of the whole extract, we should try to account for the sampling variation in our reporting; thus, the fit parameters themselves should ideally be reported with uncertainties as should quantities derived from them, such as contributor LR values. Our perspective is that we should acknowledge that the same extract of DNA subsampled and processed again at the same or a different lab will not give the same LR value. This is readily seen in the results we obtained over 3 technical replicate analyses for a 3-person mixture with DNA inputs of 0.5, 1.0, and 5 ng that is shown in Figure 11. The error bar represents 2 standard deviations. Note for lower input DNA quantity both within a sample, and between samples with different input DNA quantity, the LR variability is much higher for content with lower input DNA quantity in sampled specimens. Using the AIC framework, we were able to show analytically and directly how LR values vary with expected variation of the respective allele signal data for a sample contributor. The variability of LR values across technical replicates is a certainty, so how best report this in courts of law? We investigated one approach worthy of further consideration.

Internally, NexGenID implements a data bootstrapping method based on the expected variability of the allelic signals to report a central LR value over a statistical ensemble of solutions centered on the single instance result. The software reports the median values over this ensemble. We took this approach once step further to address the question posed above. Consider, as an analogy, a Gaussian distributed variable $G(\mu)$ with mean value μ_0 and standard deviation s. Further, assume that for a sample μ^* drawn randomly from $G(\mu)$ we define a new Gaussian variable $G'(\mu)$ with mean μ^* that has the same standard deviation as the parent $G(\mu)$. If we now draw a sample from $G'(\mu)$, the following holds: $\mu^* - 2*s < \mu_0 < \mu^* + 2*s$. In other words, the 95% probability mass interval around the sampled value μ^* will contain the parent sampling distribution population mean μ_0 with 95% confidence. While we did not perform the mathematical proof, we have verified this assertion numerically. Thus, if we estimate the 95% confidence interval of an LR around a biased single sample, this can provide a window containing the population mean value. Under the simplifying assumption of that our LR values over replicates will be Gaussian distributed (see

Figure 12 below), we can report that the population LR value for our total extracted sample is within the 95% confidence interval we estimate from the analysis of our single biased sample.

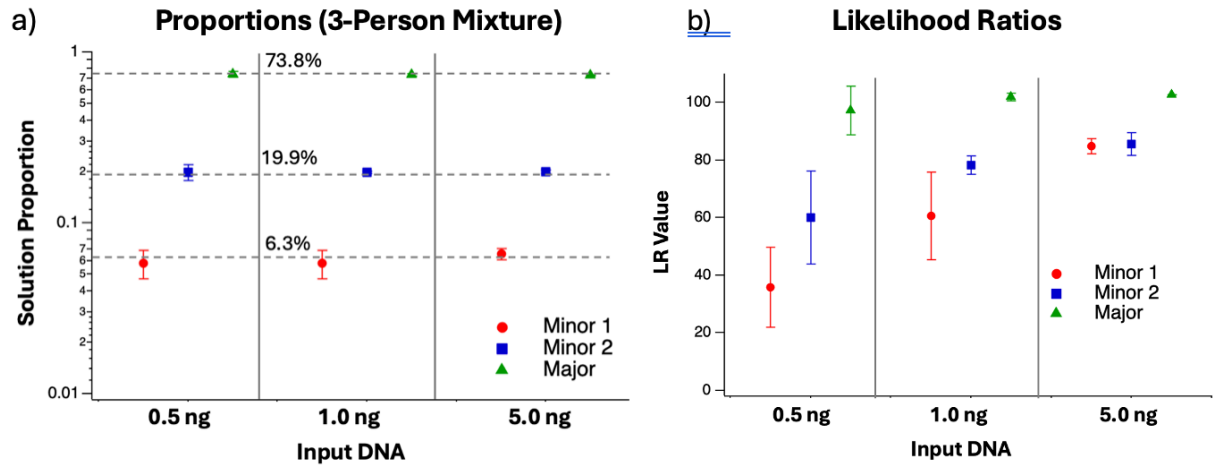

Figure 11: Solution proportion values (mean and 95% confidence interval) for end-to-end technical replicate analyses of a 3-person mixture (left) and associated Likelihood Ratio values with 95% confidence intervals (right).

Results from 100 NexGenID solutions of a single sample are shown in Figure 12 below along with the confidence band obtained over technical replicate analyses from the whole sample extract. The confidence interval over 3 technical replicates is shown in red and the NexGenID solution ensemble from the single replicate are the data points with 95% confidence ellipses. These results are encouraging and support further investigation of this concept of LR reporting. The ***lower bound of this confidence window*** could then be used as the most conservative value to report with respect to the inclusion of a Person of Interest in the total extracted DNA.

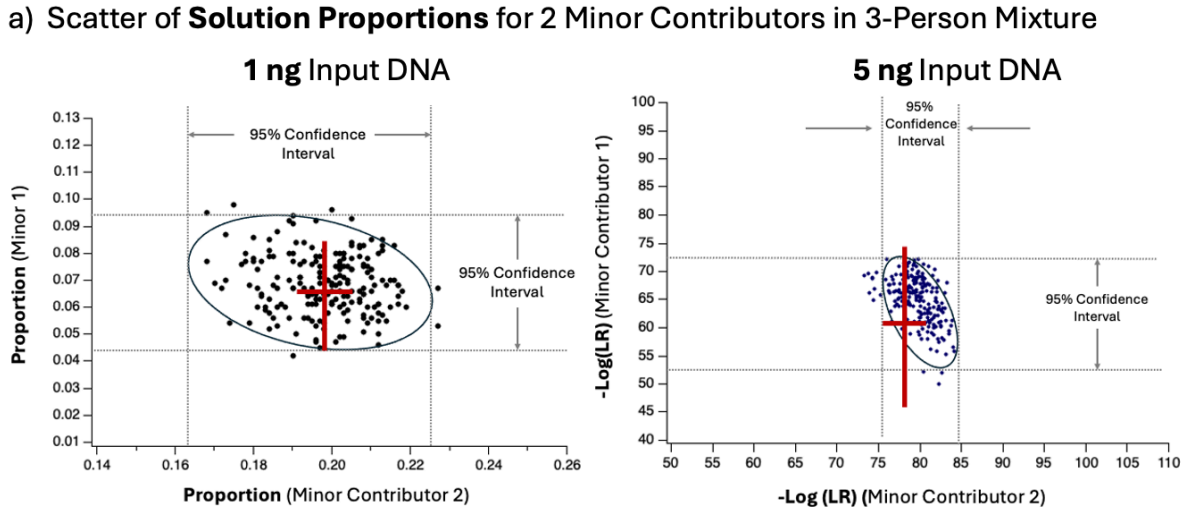

Figure 12: Distribution of proportion values (left) and Likelihood Ratios (right) for the 2 minor contributors in a 3-person mixture following 200 bootstrap replicates of a single sample using our measurement noise model. The ellipses are used to define a 95% confidence interval. The red bars indicate 95% confidence intervals for the mean values of 3 end-to-end technical replicate analyses. In all cases, the confidence intervals overlap indicating the bootstrap method intervals contain the mean value of technical replicates resampled from the sample tube and analyzed end-to-end.

Applicability to Criminal Justice

The NIST Foundational Review raised fundamental reliability questions about the use of current DNA methods for the analysis of DNA mixtures. STRs replaced RFLPs more than 20 years ago, cementing STR-CE as the gold standard. However, even with modern probabilistic genotyping solutions, significant limitations with mixtures remain. Ultimately, a new DNA typing system is needed to complement STRs that 1) has equivalent discriminatory power and sensitivity; 2) is not subject to STR's inherent amplification artifacts and 3) is compatible with modern sequencing technology utilized in forensic laboratories. A microhap-based sequencing assay and analysis methodology offers multiple advantages for complex and imbalanced mixtures that address reliability questions of current DNA methods and thus must be further developed to complement current methods. For example, the increased power of discrimination, elimination of stutter artifacts, and reduction of allelic or size-specific amplification bias will enhance mixture deconvolution by improving detection of minor contributors and improving the certainty of genotype deconvolution. The data presented demonstrate a pathway for implementation of an optimized microhap sequencing assay with an applicable end-to-end bioinformatic software platform, NexGenID, adaptable to the use with open-source algorithms, that could significantly impact the analysis of low-level DNA samples. The number of samples considered suitable for comparison will increase and enable retesting of cold cases where a minor contributor was assumed present but was not suitable for comparison. This effort also identified novel research areas that may inform future solutions to improve analyses of sequenced microhaplotypes. Together, these findings underscore that as massively parallel sequencing is adopted in forensic DNA laboratories, microhaplotypes are well positioned to integrate seamlessly alongside STRs rather than requiring an entirely new interpretive ecosystem.

Another important outcome of this effort was its contribution to forensic workforce development. Through collaboration between Bode and GW, master's-level students were engaged as research interns and received hands-on training in laboratory methods routinely used in forensic DNA laboratories, including sample preparation, experimental design, next-generation sequencing workflows, and data interpretation. Direct involvement in this research fostered critical thinking and problem-solving skills by exposing students to method development and emerging analytical challenges beyond standard coursework. This training is particularly valuable to the forensic community, as these students represent the near-term workforce and will enter operational laboratories with experience in next-generation sequencing technologies that are expected to be adopted by forensic laboratories in the near future. Results from this project will be presented as an oral presentation by one of the student researchers at the American Academy of Forensic Sciences annual meeting in February 2026, providing national and international visibility for the project while enhancing the student's professional development and employment prospects.

Products

Scholarly Products

- Manuscript in progress
- Fenske R, Cavanaugh S, Douglas H, Standage D, Lagace R, Feldman A, Kidd K, Podini D. The Analysis of a Microhaplotype Panel for Complex Mixture Deconvolution Using

Next-Generation Sequencing [master's thesis]. Washington, DC: The George Washington University, 2024.

- Cavanaugh S, Feldman A, Lin J, Borchert L, Fenske R, Kim A, Sriram T, Bever R, Kidd K, Podini D, Davoren J. Optimization of Microhaplotypes for Advanced DNA Mixture Deconvolution. 2025. Dryad, Dataset. https://doi.org/10.5061/dryad.k98sf7mmt
- Study-level information and a link to the Dryad dataset for NIJ award 15PNIJ-22-GG-04393-DNAX have been submitted to the National Archive of Criminal Justice Data (NACJD).

Dissemination Activities

1) Cavanaugh S, Podini D, Feldman A, Lin J, Becker J, Kidd K, Plaza D, Davoren J. Constructing a new laboratory and analytic workflow for mixture deconvolution using microhaplotypes. Oral Presentation at Green Mountain DNA Conference. Burlington, VT. July 22, 2024.

2) Fenske R, Cavanaugh S, Douglas H, Standage D, Lagace R, Feldman A, Lin J, Kidd K, Podini D. The Construction of a Microhaplotype Panel for Complex Mixture Deconvolution Using Next-Gen Sequencing. Oral Presentation at AAFS 77th Annual Conference. Baltimore, MD. Feb 21, 2025.

3) Borchert L, Cavanaugh S, Felman A, Kim EK, Lin J, Sriram T, Podini D. Evaluation of the NexGenID software for complex mixture deconvolution using microhaplotypes. Oral Presentation at AAFS 78th Annual Conference. New Orleans, LA Feb 2026.

References

1. National Institutes of Standards and Testing. DNA Mixture Interpretation: A NIST Scientific Foundation Review. NISTIR 83151-DRAFT. 2021.
2. Bleka Ø., Storvik G., Gill P. EuroForMix: An open-source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. Forensic Science International: Genetics 21 (2016) 35–44. https://doi.org/10.1016/j.fsigen.2015.11.008
3. Kidd KK, Speed WC. Criteria for selecting microhaplotypes: mixture detection and deconvolution. Investigative Genetics 6 (2015) Article 1. DOI 10.1186/s13323-014-0018-3
4. Standage DS, Mitchell RN. MicroHapDB: A portable and extensible database of all published microhaplotype marker and frequency data. Frontiers in Genetics 11 (2020) 781. DOI:10.3389/fgene.2020.00781
5. Kidd KK, Pakstis AJ, Gandotra N, Scharfe C, Podini D. A multipurpose panel of microhaplotypes for use with STR markers in casework. Forensic Science International: Genetics 60 (2022) 102729. DOI: 10.1016/j.fsigen.2022.102729
6. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526 (2015) 68-74.
7. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Research 12(6) (2002) 996-1006. DOI: 10.1101/gr.229102

8. Bennett L, Oldoni F, Long K, Cisana S, Madella K, Wootton S, Chang J, Hasegawa R, Lagacé R, Kidd KK, Podini D. Mixture deconvolution by massively parallel sequencing of microhaplotypes. Int J Legal Med 133 (2019) 719-729. DOI: 10.1007/s00414-019-02010-7
9. President's Council of Advisors on Science and Technology. Report to the President - Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. September 2016.
10. Bille TW, Fahrig G, Weitz SM, Peiffer GA. An improved process for the collection and DNA analysis of fired cartridge cases. Forensic Sci Int Genetics 46 (2020) 102238. https://doi.org/10.1016/j.fsigen.2020.102238
11. Ristow PG, D'Amato ME. Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics. Forensic Sci Int Genetics Suppl Series 6 (2017) e52-e54.
12. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. American J Human Genet 73 (2003) 1402-1422. DOI: 10.1086/380416
13. Rousset F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8 (2008) 103-106.
14. Hammer Ø, Harper DAT, Ryan PD. PAST 2.17c Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1) (2001) 1-9. Available: https://palaeo-electronica.org/2001_1/past/pastprog/index.html
15. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 155(2) (2000) 945–59.
16. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information Theoretic Approach, Second Edition, Springer, New York, 2002.
17. Akaike H. Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov and F. Csaki (eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest (1973).
18. Akaike H. A new look at the statistical model identification. IEEE Trans Aut Control 19 (1974):716-723.
19. Akaike H. Information measures and model selection. International Statistical Institute 44 (1983) 277-291.
20. Budowle B. "Probabilistic genotyping in Forensic DNA analysis." *American City and County.* 2020 https://www.americancityandcounty.com/2020/11/30/probabilistic-genotyping-in-forensic-dna-analysis/
21. European Network of Forensic Science Institutes (ENFSI). ENFSI Guidelines for Evaluative Reporting in Forensic Science (Approved version 3.0). European Network of Forensic Science Institutes (ENFSI): Wiesbaden, Germany (2015) https://enfsi.eu/wpcontent/uploads/2016/09/m1_guideline.pdf
22. Chow C, Liu C. Approximating discrete probability distributions with dependence trees. IEEE Trans Inf Theory 14(3) (1968) 462-467.