Detection and analysis of DNA mixtures with the MiSeq FGx™

Damani Johnson, B.S.*, Ryan Gutierrez, Ph.D., Tim Kalafut, Ph.D., Rachel Houston, Ph.D.
Department of Forensic Science, Sam Houston State University, Huntsville, Texas, USA

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

WASHINGTON, DC ISFG 2022 INTERNATIONAL SOCIETY FOR FORENSIC GENETICS

INTRODUCTION

The detection and deconvolution of mixtures is an ongoing focus of research and training in forensic laboratories. Mixtures represent a crucial portion of casework with sexual assault kit processing, touch DNA samples, and property crime all providing potential mixed casework samples. Recent advancements in mixture interpretation involve the implementation of probabilistic genotyping software and likelihood ratios. Probabilistic genotyping software allows for modeling of stochastic effects, including allele drop-out and drop-in not possible with binary approaches to mixture interpretation.

There have also been advancements in next-generation sequencing (NGS) assay design that have resulted in forensically relevant and commercially available products with potential benefits over capillary electrophoresis (CE). NGS provides many potential advantages over CE, including enhanced sensitivity, the detection of sequenced-based isoalleles that can differentiate amplicons of the same length, and the ability to multiplex more and multiple types of genetic markers. These benefits may be most noticeable when dealing with highly degraded or low-template DNA samples, where smaller amplicon sizes and larger combined multiplexes including SNP and STR markers may result in much more informative analyses.

While CE mixture analysis is currently aided by probabilistic genotyping software for more complex mixtures with high numbers of contributors, one potential disadvantage of NGS is that there currently is no probabilistic genotyping software available. However, binary deconvolution tools similar to those used for CE are available for NGS as well. ArmedXpert™ (NicheVision) is a widely used CE analysis software with binary mixture deconvolution functions (**Figure 1**), and MixtureAce™ (NicheVision) is the comparable tool for NGS. MixtureAce™ arranges NGS sequence data in a familiar format resembling an electropherogram.

This study aimed to assess the potential benefits and limits of NGS when used for mixture detection. It analyzed both mock forensic and mixture samples using CE and NGS technologies, and compared them in metrics including concordance, limits of detection, and mixture analysis.

MATERIALS & METHODS

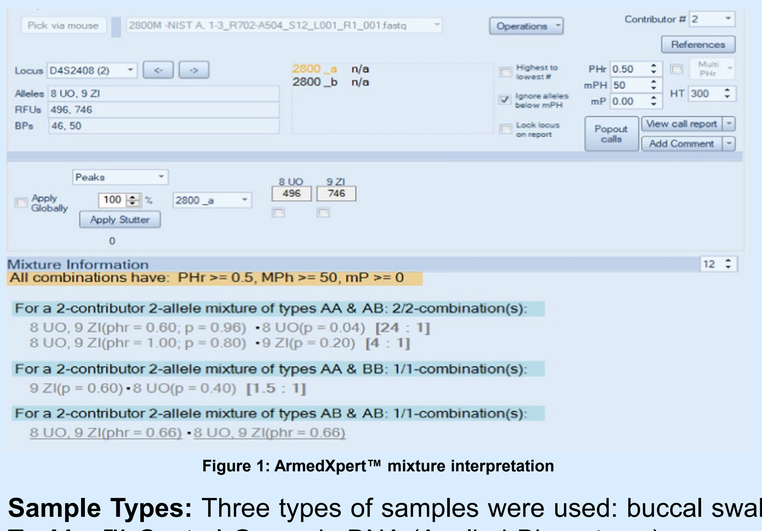

Figure 1: ArmedXpert™ mixture interpretation

Table 1: Mixture Samples

Mixture	Major	Medium	Medium 2	Minor
2 Person	Female 1	Male 1		
3 Person 1	Female 2	Male 1	Female 1	
3 Person 2	Male 2	Male 1	Female 1	
3 Person 3	Male 2	Female 3	Female 1	
3 Person 4	Male 2	Female 3	Female 4	
3 Person 5	Female 5	Female 3	Female 4	
4 Person 1	Male 2	Female 2	Male 1	Female 1
4 Person 2	Female 3	Female 2	Male 1	Female 1
4 Person 3	Female 3	Female 4	Male 1	Female 1
4 Person 4	Female 3	Female 4	Female 5	Female 1
4 Person 5	Female 3	Female 4	Female 5	Male 3

Sample Types: Three types of samples were used: buccal swabs for the sensitivity and mixture studies and mock forensic samples. TaqMan™ Control Genomic DNA (Applied Biosystems) was used for the sensitivity dilution series. Buccal swabs were collected via self-swabbing. The mock forensic samples included 22 previously extracted buccal, bone, muscle, hair roots, and blood spots from the Southeast Texas Applied Forensic Science Facility. 2-4 person mixtures were created in varying proportions from 1:1 to 40:1 according to Table 1.

DNA Extraction: Buccal samples were extracted using the EZ1® DNA Investigator® Kit (QIAGEN) with purification on the EZ1® Advanced XL using the Trace protocol.

DNA Quantification: Samples were quantified using the Investigator® Quantiplex® Pro DNA kit (QIAGEN) on a 7500 Real-Time PCR System (Applied Biosystems).

Capillary Electrophoresis: A target of 1 ng was amplified using the GlobalFiler™ PCR Amplification kit (Applied Biosystems) with 29 cycles, then analyzed using a 3500 Genetic Analyzer (Applied Biosystems).

Sequencing: A target of 1 ng was amplified using the ForenSeq™ DNA Signature Prep Primer Mix A. Sample libraries were purified and normalized using a bead-based protocol and pooled for analysis on the MiSeq FGx™ (Verogen).

Software Analysis: CE data was analyzed and deconvolutions were performed using ArmedXpert™. NGS data was analyzed and deconvolutions were performed using MixtureAce™. Non-mixture samples were analyzed using GeneMapper™ IDX v1.4 (Thermo Fisher) and Universal Analysis Software v1.3 (Verogen).

RESULTS & DISCUSSION

Concordance

- All samples were concordant for length-based genotypes between the ForenSeq™ and GlobalFiler™ kits
- In several bone and muscle samples, additional alleles were recovered using one amplification chemistry, most often with NGS over CE.
- No exclusionary discordant alleles were recorded with any samples.

Sensitivity

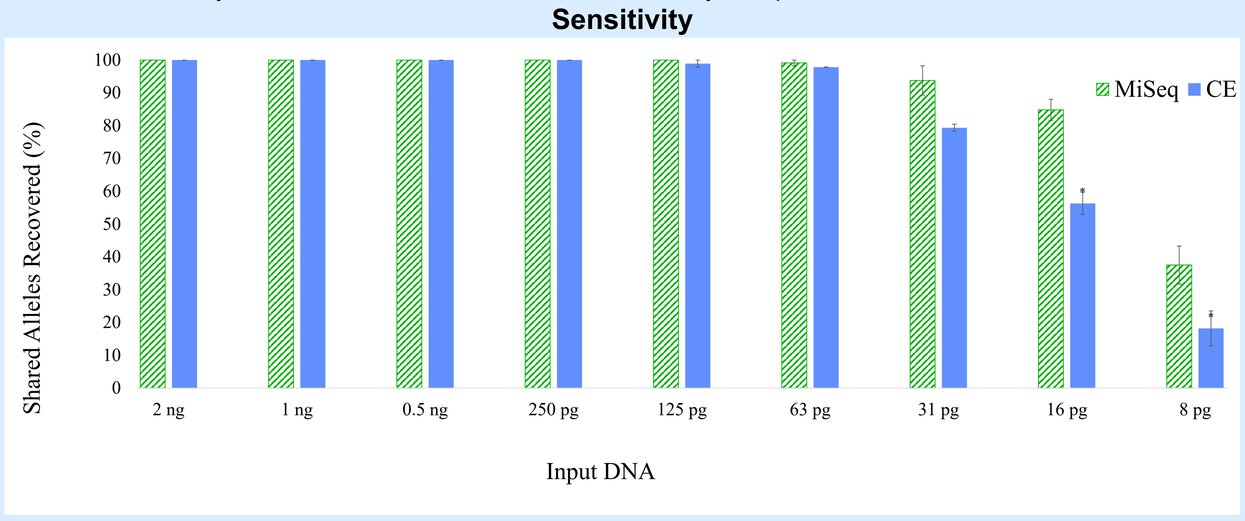

Figure 2: Sensitivity comparison of NGS and CE

- 16 pg of input DNA was sufficient to recover the majority of CODIS eligible loci included in ForenSeq™ Signature Prep Kit Primer Mix A
- 16 pg and 8 pg inputs resulted in significantly higher recovery using NGS compared to GlobalFiler™ (**Figure 2**).

Isoalleles

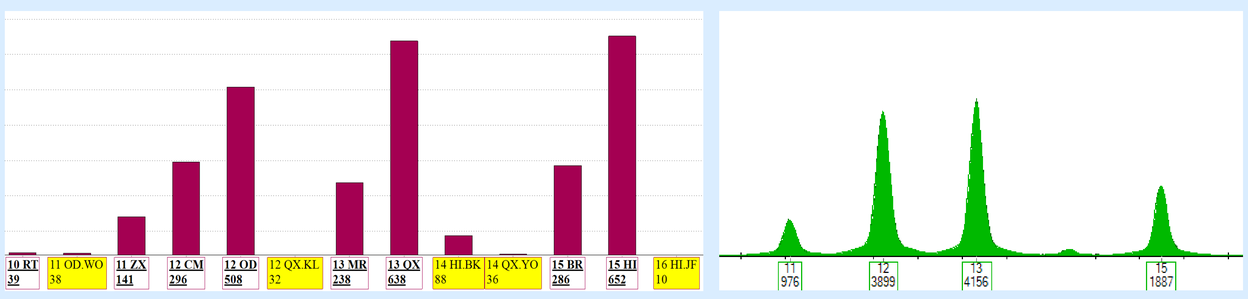

Figure 3: Software displays of the same sample's D8S1179 locus in MixtureAce™ (left) and ArmedXpert™ (right)

- Able to detect unique alleles within complex mixtures that are indistinguishable or hidden in a CE electropherogram (**Figure 3**).
- MixtureAce™ displays alleles separated by both length and sequence, with each sequence assigned a two-letter code.

Heterozygote Balance and Stutter

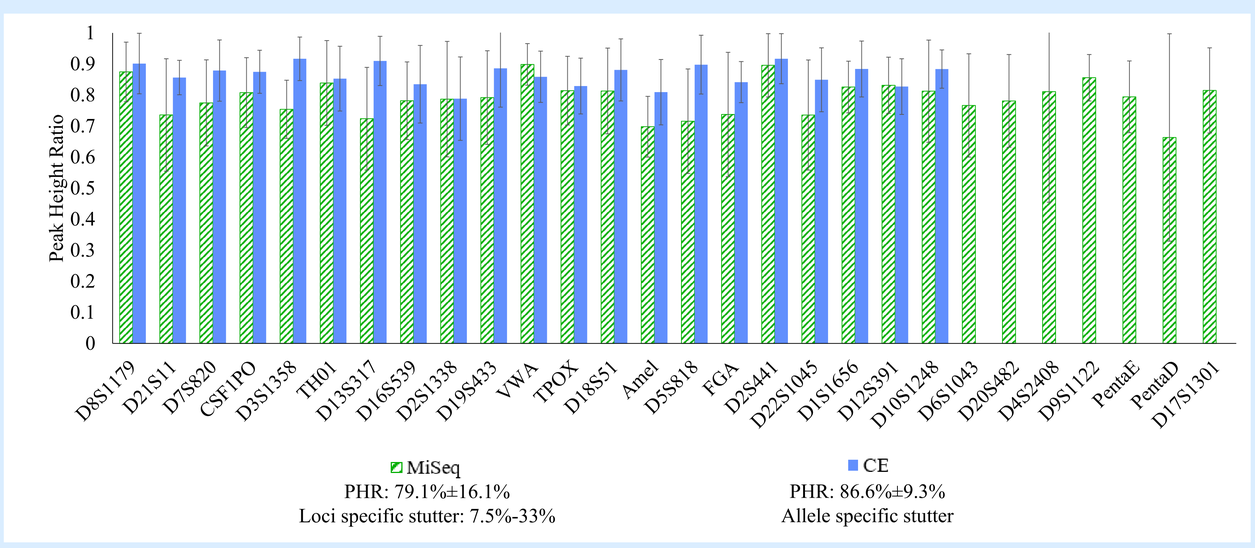

Figure 4: Heterozygote balance and stutter comparison of NGS and CE

- Average peak height ratio for autosomal STRs in NGS was 79.1% and in CE was 86.6% (**Figure 4**).
- While not statistically different, these differences may play a role in mixture interpretation.
- Stutter ratios for NGS used locus-specific thresholds, while CE analysis used allele-specific stutter filtering.

Mixture Interpretation

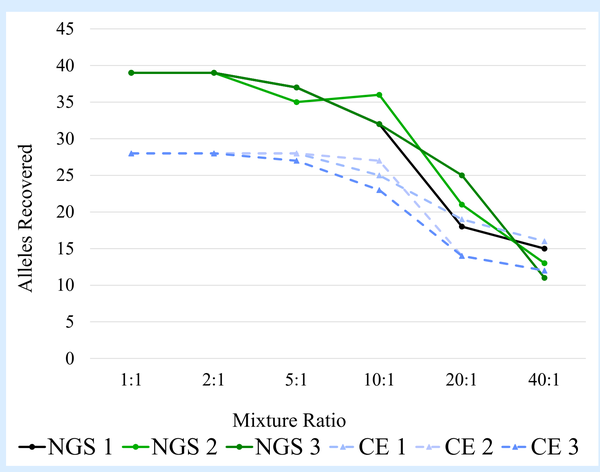

Figure 5: Allele recovery at various 2-person mixture ratios in NGS and CE

Mixture	NGS	NGS CODIS	CE
20:1	3.52E+12	1.82E+09	6.9125E+6
20:1	2.83E+18	1.25E+10	1.4632E+11
20:1	3.32E+18	8.94E+12	8.8174E+7
40:1	1.33E+10	1.40E+07	7.206E+5
40:1	1.11E+10	6.31E+07	2.8934E+9
40:1	3.27E+08	3.95E+05	7.206E+5

Table 2: Random Match Probability calculations for 2-person mixtures

- In 2-person mixtures, all mixture proportions down to 40:1 were detected in both ArmedXpert™ and MixtureAce™ (**Figure 5**).
- There were 38 possible unique minor alleles with NGS and 27 with CE, which provided noticeable advantages at mixture ratios closer than 20:1.
- At 20:1 and 40:1 ratios, NGS and CE produced similar RMPs in CODIS-eligible loci, and NGS consistently generated higher RMPs using all available loci (**Table 2**).

CONCLUSIONS

Advantages

- More autosomal STRs
- Includes Y-STRs and X-STRs
- Isoalleles provide better match statistics
- Less allele masking

Disadvantages

- Elevated stutter (thresholds)
- Peak height ratio variability
- Lack of probabilistic genotyping systems

Recommendations

- Implement probabilistic genotyping
- Simplified assay design (ForenSeq™ MainstAY)

ACKNOWLEDGEMENTS

The authors would like to acknowledge Verogen for the provision of reagents to this project. Additionally, we would like to acknowledge NicheVision Forensics for providing the MixtureAce™ software. Finally, we would like to acknowledge the Sam Houston State University Department of Forensic Science as well as the Southeast Texas Applied Forensic Science Facility and those deceased donors who helped in advancing forensic research.

REFERENCES

1. NIST. DNA Mixture Interpretation: A NIST Scientific Foundation Review. 2021.
2. Buckleton, J.S., et al., The Probabilistic Genotyping Software STRmix: Utility and Evidence for its Validity. J Forensic Sci, 2019. 64(2): p. 393-405.
3. Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int, 2006. 160(2-3): p. 90-101.
4. Bright, J.A., et al., Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. Forensic Sci Int Genet, 2016. 23: p. 226-239.
5. Jager, A.C., et al., Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Sci Int Genet, 2017. 28: p. 52-70.
6. Holt, C.L., et al., Human Mitochondrial Control Region and mtGenome: Design and Forensic Validation of NGS Multiplexes, Sequencing and Analytical Software. Genes (Basel), 2021. 12(4).
7. Verogen, ForenSeq™ DNA Signature Prep Reference Guide. 2018.
8. FBI National DNA Index 534 System (NDIS) Operational Procedures Manual.
9. Hollard, C., et al., Automation and developmental validation of the ForenSeq DNA Signature Preparation kit for high-throughput analysis in forensic laboratories. Forensic Sci Int Genet, 2019. 40: p. 37-45.
10. Thanakiatkrai, P., et al., Performance comparison of MiSeq forensic genomics system and STR-CE using control and mock IED samples. Forensic Science International: Genetics Supplement Series, 2017. 6: p. e320-e321.
11. Guo, F., et al., Massively parallel sequencing of forensic STRs and SNPs using the Illumina ForenSeq DNA Signature Prep Kit on the MiSeq FGx Forensic Genomics System. Forensic Sci Int Genet, 2017. 31: p. 135-148.
12. Applied Biosystems. GlobalFiler and GlobalFiler IQC PCR Amplification Kits.
13. Novroski, N.M.M., et al., Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forensic Sci Int Genet, 2016. 25: p. 214-226.
14. Devesse, L., et al., Concordance of the 554 ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups. Forensic Sci Int Genet, 2018. 34: p. 57-61.
15. Gettings, K.B., et al., Sequence-based U.S. population data for 27 autosomal STR loci. Forensic Science International: Genetics, 2018. 37: p. 106-115.
16. Barrio, P.A., et al., Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. Forensic Science International: Genetics, 2019. 42: p. 49-55.
17. Gettings, K.B., et al., STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. Forensic Sci Int Genet, 2017. 31: p. 111-117.
18. Phillips, C., et al., Global patterns of STR sequence variation: Sequencing the CEPH human genome diversity panel for 58 forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit. ELECTROPHORESIS, 2018. 39(21): p. 2708-2724.
19. Young, B., et al., Match statistics for sequence-based alleles in profiles from forensic PCR-mps kits. ELECTROPHORESIS, 2021. 42(6): p. 756-765.
20. Young, B.A., 578 et al., Estimating number of contributors in massively parallel sequencing data of STR loci. Forensic Science International: Genetics, 2019. 38: p. 15-22.
21. Verogen, ForenSeq Universal Analysis Software v1.3 Reference Guide. 2018.
22. Sharma, V., et al., Evaluation of ArmedXpert software tools, MixtureAce and Mixture Interpretation, to analyze MPS-STR data. Forensic Science International: Genetics, 2022. 56.
23. Young, B., T. Faris, and L. Armogida, A nomenclature for sequence-based forensic DNA analysis. Forensic Science International: Genetics, 2019. 42: p. 14-20.
24. Goor, R.M., et al., A mathematical approach to the analysis of multiplex DNA profiles. Bull Math Biol, 2011. 73(8): p. 1909-31.
25. Kalafut, T., et al., Implementation and validation of an improved allele specific stutter filtering method for electropherogram interpretation. Forensic Sci Int Genet, 2018. 35: p. 50-56.
26. Hill, C 599 .R., et al., U.S. population data for 29 autosomal STR loci. Forensic Sci Int Genet, 2013. 7(3): p. e82-3.
27. Steffen, C.R., et al., Corrigendum to 'U.S. Population Data for 29 Autosomal STR Loci' [Forensic Sci. Int. Genet. 7 (2013) e82–e83]. Forensic Science International: Genetics, 2017. 31: p. e36-e40.
28. Xavier, C. and W. Parson, Evaluation of the Illumina ForenSeq DNA Signature Prep Kit - MPS forensic application for the MiSeq FGx benchtop sequencer. Forensic Sci Int Genet, 2017. 28: p. 188-194.
29. Silvia, A.L., N. Shugarts, and J. Smith, A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex. 2017, Springer Science + Business Media: Germany. p. 73.
30. Hussing, C., et al., Sequencing of 231 forensic genetic markers using the MiSeq FGx forensic genomics system - an evaluation of the assay and software. Forensic Sci Res, 2018. 3(2): p. 111-123.
31. Sharma, V., et al., Qualitative and quantitative assessment of Illumina's forensic STR and SNP kits on MiSeq FGx. PLoS One, 2017. 12(11): p. e0187932.