Practical Implementation of ForenSeq™ DNA Signature Sequence-Based Mixture Interpretation, Sample Comparison and Populations Statistics Methods for Criminal Casework

Stephanie Sarnese, MS[1]; Erica Black, BS[1], Cassidy Torgrimson, MS[1]; Susan Belote, BS[1]; Meghan Didier, MS[1]; Kyla Hackman, MFS[1]; Cydne Holt, PhD[1]

[1]Tetracore, Inc., Rockville, MD 20850

INTRODUCTION

In recent years, massively parallel sequencing (MPS) has been internally validated by forensic laboratories and offers several advantages over length-based STR genotyping. Benefits include a single assay for multiple categories of STRs and SNPs to generate more data from less DNA and additional discrimination power and mixture detection/resolution from nucleotide-level allele calling [1-4]. Although MPS forensic testing has been available for many years, and the chemistry has proven superior to CE, the primary hurdle to practical implementation has been interpretation, comparison & statistical limitations within available software programs.

Tetracore has validated and implemented the following in-house workflow to effectively realize the full data potential of the ForenSeq DNA Signature assay:

UAS-generated allele detection & calling/typing → Import sequence-based genotype data to Seqogram tool → Tetracore's Seqogram data analysis tool for single source or mixed profile interpretation + sequence-based sample comparisons → Import sequence-based alleles to Pop Stats tool (for inclusions) → Tetracore's sequence-based population statistics calculator (RMP or LR) for auSTRs, X STRs & iiSNPs + YHRD for Y STR Stats

MATERIALS AND METHODS

- The following QAS/SWGDAM-based validation studies were conducted to validate ForenSeq DNA Signature (Primer Mix B), the MiSeq FGx instrument and Universal Analysis Software (UAS) v1.3:
  - Repeatability, reproducibility, accuracy, precision, sensitivity and stochastic, STR stutter, DNA mixtures, mock evidence samples, known reference standards, contamination and artifacts, controls, run quality metrics and sample multiplex.
- Internal validation mixture studies and evaluations consisted of:
  - 48 unique mixtures (including NISTD),104 total replicates, 2-, 3- and 4-person mixtures, total inputs ranging from 50 ng to 25 pg and contributor ratios ranging from 1:1 to 1:100, with GlobalFiler™ comparisons for some mixtures.
- Tetracore's v1.0 STR & iiSNP Seqogram Tool, v1.0 Population Statistics calculator & YHRD Pop Stats validation:
  - A series of software requirements and test cases were developed prior to testing, with expected outcomes documented (33 test cases for the STR Seqogram, 60 test cases for the iiSNP Seqogram, single source & mixed samples)
  - Two users performed testing. Results between users were compared for reproducibility and to the truth data.
  - For statistics, "hand calculation" results (Excel) were compared for accuracy. Select GlobalFiler™ mixtures were analyzed in STRmix v2.5.11 to generate inclusion likelihood ratios for the minor profile.

Comparison of MPS Interpretation Software Capabilities

MPS Software	Validation Status	Mixture Interpretation & Documentation				Automated sequence-based allele comparison & population statistics			
	For Forensic Casework Use	Assignment of single source major/minor contributors or multiple possible genotypes	Ability to input known contributor genotype	Automated SNP mixture decon	Highly visible & reviewable interpretation methods, conclusions & calculations	au STRs	X STRs	iSNPs	Y STRs aiSNPs piSNPs
Tetracore's Custom Tools		Binary methods							NA[1,2]
UAS v2.0									NA[1,2]
STRmix NGS R&V		PG Methods							NA[1,2]

Figure 1. Comparison of Tetracore's in-house MPS interpretation tools relative to UAS v2.0 and STRmix NGS R&V programs. Comparison focused on mixture interpretation capabilities and ability to use sequence-based allele data for population statistics in an automated fashion.

[1] Y STRs rely on YHRD for population statistics, which does not currently accept sequence-based data. Length-based data were thus used for Y STR statistics.

[2] ai/piSNPs provide investigative intelligence data in the form of biogeographical ancestry and phenotype (hair/eye color) estimation and are not intended for identity comparisons or population statistics.

ACKNOWLEDGEMENTS

The authors would like to thank all analysts and reviewers of this poster and validation reports for their time, energy and participation in this effort.

REFERENCES

1. K. Cheng, J.A. Bright, H. Kelly, Y.Y. Liu, M.H. Lin, M. Kruijer, D. Taylor, J. Buckleton, Developmental validation of STRmix™ NGS, a probabilistic genotyping tool for the interpretation of autosomal STRs from forensic profiles generated using NGS, Forensic Sci. Int. Genet. 62 (2023) 102804. https://doi.org/10.1016/j.fsigen.2022.102804.
2. R.S. Just, L.I. Moreno, J.B. Smerick, J.A. Irwin, Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens, Forensic Sci. Int. Genet. 28 (2017) 1–9. http://dx.doi.org/10.1016/j.fsigen.2017.01.004.
3. Verogen, ForenSeq D.N.A. Signature Prep Kit Reference Guide VD2018005 Rev. D, 2022.
4. A.C. Jäger, M.L. Alvarez, C.P. Davis, E. Guzmán, Y. Han, L. Way, P. Walichiewicz, D. Silva, N. Pham, G. Caves, J. Bruand, F. Schlesinger, S.J.K. Pond, J. Varlaro, K.M. Stephens, C. L. Holt, 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. 28 (2017) 52-70.
5. K.B. Gettings, L.A. Bourke, C.R. Steffen, K.M. Kiesler, P.M Vallone, Sequence-based U.S. population data for 27 autosomal STR loci, Forensic Sci. Int. Genet. 37 (2018) 106-115.

RESULTS & DISCUSSION

- Tetracore's primary capabilities as related to sequence-based allele designations and population statistics are summarized here (& shown in comparison to OTS software programs in Figure 1):
  - Binary mixture interpretation (with or without a known contributor genotype); automated for SNP 2-person mixtures
  - Automated sequence-based allele comparisons
  - Automated RMP/mRMP and LR pop stats calculations (using NIST 1036 sequence-based allele frequency data) [5] & NRCII formula recommendations
  - Documentation of methods, conclusions & calculations to enable case file review (even if reviewer does not have UAS access)
- STR & iiSNP data interpretation tools enabled successful single source and/or mixed profile interpretation for all marker categories.
- Deduced contributor genotypes resulted in expected inclusions and exclusions (see Figures 2 & 3).

Mixture M1:M3 1:10, 25pg total input	GlobalFiler	DNA Signature
Total # of deduced minor contributor alleles	18	23
auSTR LR Stat	1.11E12 (Binary LB) 3.60E12 (PG LB)	5.47E18 (Binary SB)

Figure 4. Example of the increased resolution & statistical power of sequence-based (SB) DNA Signature using binary statistics vs length-based (LB) analysis of GlobalFiler using binary statistics or advanced probabilistic genotyping (STRmix v2.5.11).

- Comparison of mixed sample results between DNA Signature and GlobalFiler indicated the following:
  - Observed mixture ratios were consistent between Sig Prep & GF profiles, supporting use of calculated ratios for mixture deconvolution.
  - Number of fully resolved loci increased with Sig Prep due to additional loci (*e.g.*, 26 vs 21 auSTRs) and sequence-based allele calling to improve accuracy in NOC estimation.
  - Sequence-based allele frequencies increased the power of discrimination for major and minor contributor comparisons (Figure 4).
  - RMPs of partial, minor contributor data were on the order of 1 in a trillion or less (down to 25 pg input DNA) and major contributor profiles were ~1 in a decillion at 250 pg input, demonstrating increased power of discrimination provided sequenced-based allele frequencies.

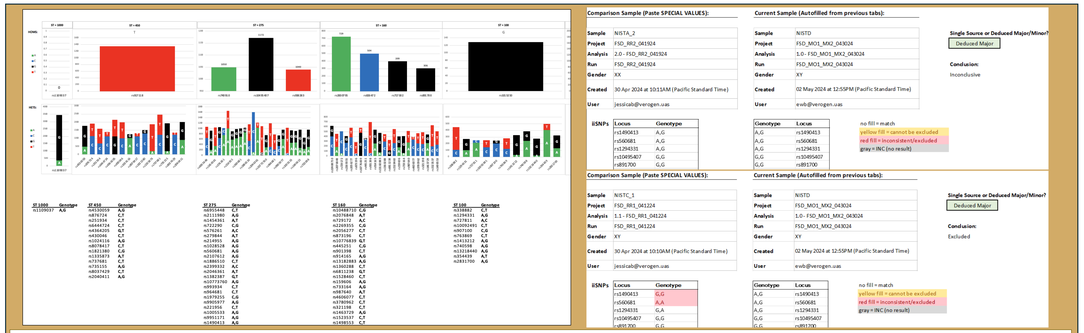

Figure 2. Example of Tetracore's iiSNP data visualization (left) & automated iiSNP sample comparisons. The top comparison is of component NISTA (known major contributor in mixture NISTD) to the iiSNP deduced major contributor of mixed sample NISTD, and results in the expected "inclusion". The bottom comparison is to reference sample NISTC (the known minor component of NISTD mixture), and results in the expected "exclusion".

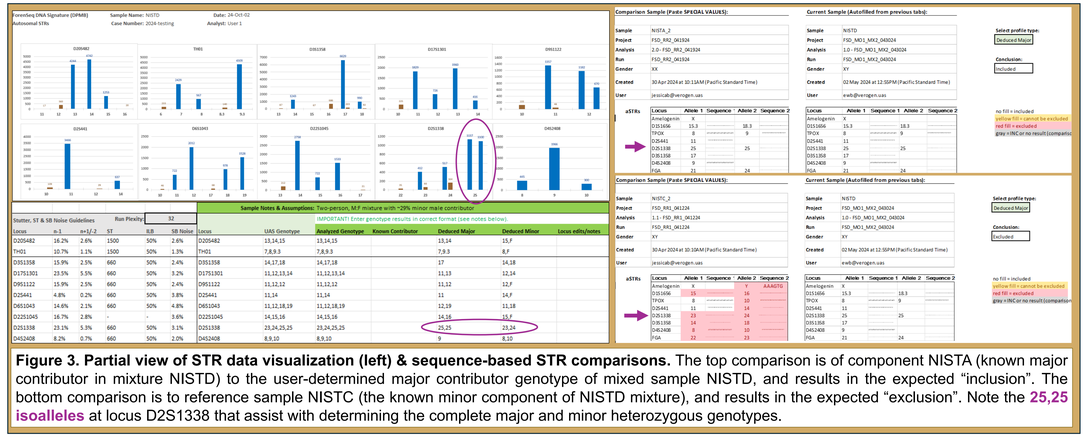

Figure 3. Partial view of STR data visualization (left) & sequence-based STR comparisons. The top comparison is of component NISTA (known major contributor in mixture NISTD) to the user-determined major contributor genotype of mixed sample NISTD, and results in the expected "inclusion". The bottom comparison is to reference sample NISTC (the known minor component of NISTD mixture), and results in the expected "exclusion". Note the **25,25 isoalleles** at locus D2S1338 that assist with determining the complete major and minor heterozygous genotypes.

CONCLUSIONS

Tetracore successfully validated and implemented the ForenSeq DNA Signature MPS System for casework, to include DNA Primer Mix B assay, MiSeq FGx Instrument (Standard Flow Cell), UAS v1.3, Tetracore's v1.0 STR & iiSNP Seqogram Data Interpretation & Comparison Tool, Tetracore's Population Statistics Calculator for auSTRs, X STRs, Y STRs & iiSNP loci, & YHRD for Y STR loci. The UAS is used for allele detection and genotyping (calling), while Tetracore's tools further enable binary mixture interpretation and sequence-based allele comparisons & population statistics.

Tetracore recognizes the benefits of probabilistic genotyping software for mixtures. These validation data can be used for STRmix™ NGS R&V software evaluation for use within the laboratory's workflow. Prior to availability of a casework-ready PG solution, Tetracore's validated interpretation tools enable application of the full benefits of MPS data to casework samples, improving case resolutions over CE workflows and other partially-implemented MPS workflows (*e.g.*, those validated for only some marker category(ies), limited or no mixture procedures, length-based sample comparison and statistics).