TrueAllele® Casework Validation of the Verogen ForenSeq DNA Signature Prep Kit Primer Set B and the MiSeq FGx™

Matthew M. Legler[1], Jennifer M. Bracamontes[1], MS, and Mandi Van Buren[2], MS
[1]Cybergenetics, Pittsburgh, PA
[2]Kern Regional Crime Laboratory, Bakersfield, CA

Abstract

Next Generation Sequencing (NGS) is an exciting technology that will allow forensic labs to extract more DNA information from a single sample. Currently, DNA data interpretation is becoming an increasingly complex task as sample types (like touch DNA samples) and newer testing methods (like NGS) become more common. As new DNA technology becomes available, a lab will look to extend their current, established tools for new data interpretation. Toward this end, Cybergenetics, the creators of the TrueAllele® probabilistic genotyping software, updated their system to support NGS STR data produced by the Verogen ForenSeq DNA Signature Prep Kit Primer Set B and the MiSeq FGx™.

The Kern Regional Crime Laboratory along with Cybergenetics validated the updated TrueAllele software to establish the sensitivity, specificity, and reproducibility of the process. The validation includes testing across different axes, including:

For single-source data, three studies were performed:
• Sensitivity & Stochastic - consisting of two dilution series.
• Reproducibility & Repeatability - two users set up the same data.
• Known references - 45 known donors with unique genotypes.

For case-type data, two studies were performed:
• Mock Casework - 21 mock casework samples of known composition involving varying sample types and contributor number.
• Mixture Studies - samples of known composition involving varying contributor number and mixture weight (MW) ratios.

Finally, for concordance, three studies were performed:
• Capillary Electrophoresis (CE) - A set of 45 single-source samples prepared using two sequencing methods (CE and NGS).
• Plate Setup - samples of varying contributor number prepared using two plate setup methods.
• Primer Mix Autosomal STR – Comparison of STR markers of 57 samples of varying contributor number prepared using two NGS primer sets (Primer Mix A & Primer Mix B).

This validation showed that the system produced sensitive, specific, and reproducible results for DNA data containing one to five contributors produced by the Kern Regional Crime Laboratory using the ForenSeq DNA Signature Prep kit Primer Set B with the MiSeq FGx Next Generation Sequencer.

Background

NGS, or massive parallel sequencing, is a newer methodology that can be used for forensic DNA applications. This technology produces identification information based on both DNA size (like traditional STR analysis) and DNA base pair sequence. NGS data comprises of traditional STR markers as well as single nucleotide polymorphisms (SNPs), mitochondrial DNA (mtDNA), and other forensically relevant markers.

Using NGS, DNA laboratories can analyze many DNA samples at one time (in parallel). Utilizing this methodology, a lab can also increase the amount of DNA data produced for a sample, leading to a higher discriminatory power for DNA identification. Additionally, having the DNA sequence information allows labs to potentially distinguish between people having the same allele value at a DNA location (isoalleles) as the DNA sequences between the two people can differ. NGS analysis can also be used for better analysis of degraded DNA samples, genealogy applications, and high-throughput of samples in less time (due to parallel sequencing).

Both NGS and CE workflows start with DNA extraction, quantification, and amplification using polymerase chain reaction (PCR). The both also produce DNA data that an analyst can interpret. For the CE workflow, the STR DNA fragments are size separated and the relative fluorescent intensity for each DNA fragment (i.e., allele) is captured in an electropherogram. In the NGS workflow, there are some additional preparation steps (e.g., library preparation) after amplification. Once prepared, the sample is sequenced and the output includes both STR, SNP, and other forensic marker data. The NGS output also includes DNA sequence and read amount for each allele.

Data

Data Set	Study Name	Description
Single-Source	Sensitivity & Stochastic	two dilution series (4 ng to 8 pg)
	Reproducibility & Repeatability	two users set up the same data
	Known Studies	45 known donors with unique genotypes
Case-type	Mock Casework	21 mock casework samples of known composition involving varying sample types (touch, body fluids) and contributor number (1 to 5)
	Mixture Studies	samples of known composition involving varying contributor number (2 to 4) and mixture weight (MW) ratios
Concordance	Capillary Electrophoresis	45 single source samples preparing using two sequencing methods (CE and NGS)
	Plate Setup	6 samples of varying contributor number prepared using two plate setup methods
	Primer Mix Autosomal STR	57 samples of varying contributor number prepared using two NGS primer mixes (Primer Mix A & Primer Mix B)

Methods

Data generation
NGS kit: Verogen ForenSeq DNA Signature Prep Kit Primer Mix B (27 autosomal STR markers, X- and Y-chromosome markers, and SNPs)
NGS sequencer: MiSeq FGx™
CE kit: GlobalFiler™
CE instrument: 3130

Genotype separation
Software: TrueAllele Casework (autosomal STRs only)

Match information
Likelihood ratio (LR) statistics
Software: TrueAllele VUIer™
Co-ancestry coefficient: 1%

Exact genotype LR distributions
Calculated in VUIer directly from genotypes to evaluate validation metrics and error rates [3]

Validation metrics
Sensitivity: inclusionary match strength of true contributors
Specificity: exclusionary LRs for non-contributors
Reproducibility: similarity of results (within-group standard deviation)

NGS workflow

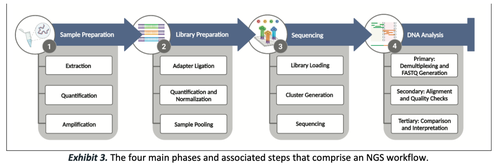
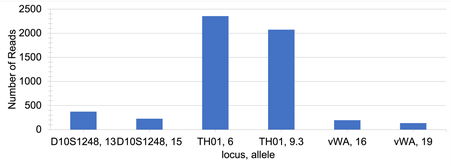


Exhibit 3. The four main phases and associated steps that comprise an NGS workflow.

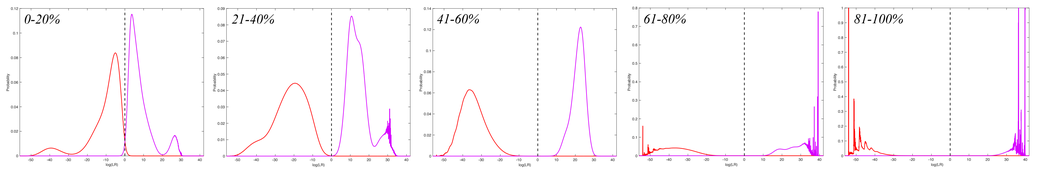
Image from [1]

Samples are prepared in the same way as traditional DNA analysis, but using specific NGS testing kits to amplify the DNA samples (top left). The MiSeq FGx Sequencing System (top right, image from [2]) is used to generate the NGS DNA data. NGS DNA data output includes both autosomal STR data (select loci, bottom) and SNPs.

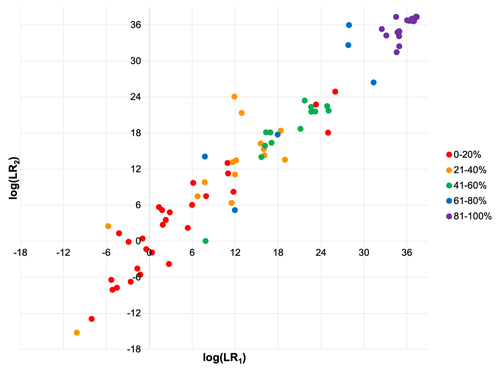


Results

Mixture samples: sensitivity, specificity, and reproducibility

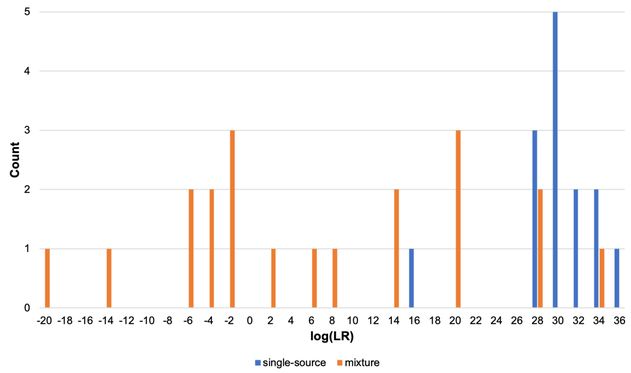


Line graphs show exact non-contributor (red) and contributor (pink) distributions for the two-person NGS mixture data for different MW ranges.



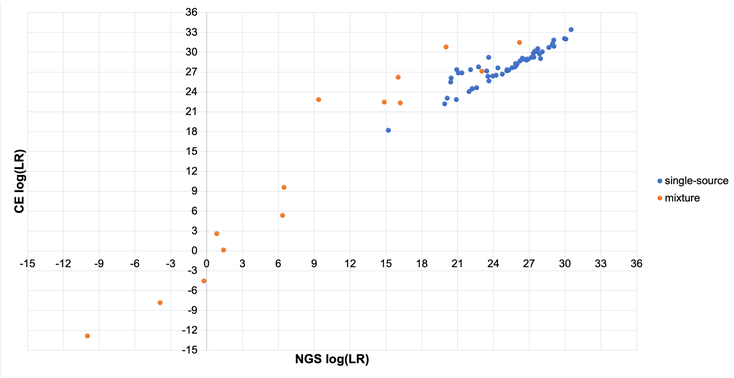
Scatterplot of log(LR) comparison between independent replicate computer runs on the same NGS mixture data. The results are separated into MW ranges, with each range shown as a different series color.

Single-source and mixture samples



Histogram of log(LR) values for evidence to reference genotype comparisons for the NGS case-type mock data samples, for single-source (blue) and mixture (orange) samples.

Scatterplot of log(LR) comparison between NGS and CE methods on the same sample, for single-source (blue) and mixture (orange) samples. NGS match statistics were calculated from GlobalFiler loci.



Conclusions

- NGS is a new, exciting DNA technology that empowers laboratories to get more identification information from their DNA evidence. NGS data can combine autosomal STR, sex chromosome and SNP data, along with DNA sequence information.
- TrueAllele Casework probabilistic genotyping software can reliably interpret single-source, mixture, and case-type NGS autosomal STR data.
- This study established TrueAllele's sensitivity, specificity, reproducibility, and concordance with CE methodology for NGS data generated by the Verogen ForenSeq DNA Signature Prep Kit on a MiSeq FGx™ sequencing instrument.
- On mixture data, sensitivity and specificity exact contributor and non-contributor distributions show that as mixture weight increases, so does average inclusionary and exclusionary match information. So too does distribution separation.
- Additionally, TrueAllele gave similar match statistics between two independent contributor runs for mixture data as shown in the mixture reproducibility scatterplot. The log(LR) results roughly follow the equal information line, showing reproducibility.
- When compared to traditional CE method for producing DNA data, the TrueAllele-inferred genotype results from autosomal STR NGS data gave similar DNA information and match statistics for both single-source and case-type samples (including DNA mixtures).
- In summary, a DNA laboratory can get useful information from their DNA evidence using next generation sequencing technology. This study established that the autosomal STR NGS data can be reliably interpreted, with similar mixture weights and match statistics to the STR data generated by CE methods. SNP interpretation using TrueAllele is currently under development.

References

[1] Martin, M., Estus, E., Krauss, S., Sorrell, C., Shute, R., Ferrara, L., Johnson, T., and Melton, P. Landscape Study of Next Generation Sequencing Technologies for Forensic Applications. *Forensic Technology Center of Excellence, U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Office of Investigative and Forensic Sciences*, April 2023.

[2] Verogen. MiSeq FGx Sequencing System. https://verogen.com/products/miseq-fgx-sequencing-system/. Accessed 12 September 2023.

[3] Perlin, M.W. Efficient construction of match strength distributions for uncertain multi-locus genotypes. *Heliyon*, 4(10):e00824, 2018.

Validation

• Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. PLoS ONE. 2009;4(12):e8327.

• Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA mixture interpretation. J Forensic Sci. 2011;56(6):1430-1447.

• Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. Sci Justice. 2013;53(2):103-114.

• Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. J Forensic Sci. 2013;58(6):1458-1466.

• Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. PLOS ONE. 2014;9(3):e92837.

• Perlin MW, Hornyak J, Sugimoto G, Miller K. TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. J Forensic Sci. 2015;60(4):857-868.

• Greenspoon SA, Schiermeier-Wood L, Jenkins BA. Establishing the limits of TrueAllele® Casework: a validation study. J Forensic Sci. 2015;60(5):1263-1276.

• Bauer DW, Butt N, Hornyak JM, Perlin MW. Validating TrueAllele® interpretation of DNA mixtures containing up to ten unknown contributors. J Forensic Sci. 2020;65(2):380-398.

Cybergenetics

Cybergenetics © 2023