B40 The Generation of a Universal Protocol Data Set to Validate Probabilistic Genotyping Software for Uniformity Between Laboratories

Josephine Buckley, BA, Elkins Park, PA 19027; Fabio Oldoni, PhD, Arcadia University, Glenside, PA 19038; Megan M. Foley, MSFS, Center for Forensic Science Research & Education, Willow Grove, PA 19090*

Learning Overview: After attending this presentation, attendees will understand: (1) if a shared data set can be used by laboratories for the validation and verification of probabilistic genotyping calculations on mixed DNA profiles using MaSTR™ software, and (2) how to implement suggested parameters and guidelines for MaSTR™ software into their workflow for mixture interpretation.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing guidelines to laboratories with limited resources and by introducing a sustainable avenue for switching from limited binary approaches to the more statistically powerful approach of probabilistic genotyping using the software MaSTR™ software.

Forensic scientists have been utilizing DNA analysis as an investigative tool for over 30 years. Samples collected from crime scenes often contain multiple contributors, degraded DNA, and/or low template DNA, which confound the interpretation process.[1] These complex mixtures are also often characterized by an increased presence of stochastic factors that further complicate interpretation.[2] Crime labs in the United States have historically used a binary method to interpret and perform statistical analysis on mixtures. As the mixtures produced from crime scene samples have become more complex, however, many labs have started to switch over to the non-binary approach of probabilistic genotyping, which eliminates the need for binary thresholds and increases the statistical power of interpretations.[3] A fully continuous probabilistic model uses peak heights and a variety of parameters to identify which contributor genotypes are best described by the data and then calculates a Likelihood Ratio (LR) to provide a statistical weight for the hypotheses chosen for court reports.[4]

Despite the obvious advantages to the probabilistic approach, the switch to the probabilistic approach has been slow in the United States.[2] Traditionally, laboratories have conducted an end-user validation to ensure that the software works as intended in their lab, using their protocols and instrumentation to produce data sets for this evaluation.[5,6] This process can take a substantial amount of time and resources away from casework. Many laboratories, therefore, continue to analyze complex mixtures using less informative, traditional approaches or report these mixtures as uninterpretable or inconclusive. This work examined the feasibility of generating a universal data set that could be utilized by laboratories interested in validating MaSTR™ probabilistic genotyping software in their laboratories, alongside a smaller data set created in-house for verification. In this study, software parameters (e.g., variance factor for peak height ratios) were validated using a large data sets of 44, 30, 20, and 10 single-source GlobalFiler™ profiles run through the MaSTR™ software in various replicates, depending on data set. From these large data sets, subsets of Short Tandem Repeat (STR) profiles were used to create multiple smaller data sets of varying sizes from which software parameters were determined. Variation within parameters between the large and small data sets were then analyzed to evaluate the extent of the variability between running data sets of varying sizes. This process was then repeated to determine the variation between data sets generated using different amplifications kits. Data sets will be generated from the same amplification kits under different conditions, (e.g., different labs, PCR cycles, Capillary Electrophoresis [CE] injection times, etc.). The software parameters between these data sets will then be compared to determine if the variability between the parameters is significant. Finally, MaSTR™ software will be used to calculate LRs for the same DNA mixtures, using all of the data set parameters, and the variability between the calculated LRs will be assessed.

Preliminary results showed that the variability in generated parameters, between data sets of varying sizes was minimal. These preliminary results indicate that the variation within parameters can be effectively modeled for the MaSTR™ software, using all of the data set parameters generated from data sets of at least ten STR profiles. Some of the data set parameters generated from data sets of four STR profiles exhibited less variation than the larger data sets. These results indicate that for data sets with four or less profiles, the characteristics of the individual profiles impact the extent of variation within parameters. This study aims to provide a more sustainable alternative to a full in-house validation, which requires the use of unique internally generated data sets for the validation of MaSTR™ probabilistic genotyping software.

Reference(s):

1. Kelly, Hannah; Bright, Jo-Anne; Buckleton John S; Curran, James M. A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science & Justice* 54, no.1 (January 2014): 66-70. https://doi.org/10.1016/j.scijus.2013.07.003.
2. Moretti, Tamyra; Gill, Peter; Garcia, Lynn. The Elements of DNA Profile Interpretation and Probabilistic Genotyping. *National Institute of Justice*. (May 2019) https://learning.forensicac.org/mod/page/view.php?id=1250.
3. Coble, Michael D.; Bright, Jo-Anne. Probabilistic genotyping software: An overview. *Forensic Science International* 38, (January 2019): 219-224. https://doi.org/10.1016/j.fsigen.2018.11.009.
4. Butler, John M. Coping with Potential Missing Alleles. In *Advanced Topics in Forensic DNA Typing: Interpretation*. 333-348. Walthan: Academic Press, 2014.
5. Haned, Hinda; Gill, Peter; Lohmueller, Kirk; Inman, Keith; Rudin, Norah. Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. *Science & Justice* 56, no. 2 (March 2016) 104-108. https://doi.org/10.1016/j.scijus.2015.11.007.
6. Scientific Working Group on DNA Analysis Methods. *Guidelines for the Validation of Probabilistic Genotyping Systems.* Accessed June 24, 2020). https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf.

Probabilistic Genotyping, STR Analysis, Software Validation

*Presenting Author