B96 Exploring the Impact of DNA Template Mass on the Ability to Infer the Number of Contributors Using Three Interpretation Methods

Lauren E. Alfonse, BS, Boston Univesity School of Medicine, 72 E Concord Street, R806, Boston, MA 02118; Harish Swaminathan, Rutgers University, Center for Computational & Integrative Biology, Camden, NJ 08102; Desmond S. Lun, PhD, Rutgers University, Center for Computational & Integrative Biology, Camden, NJ 08102; Muriel Medard, ScD, Massachusetts Institute of Technology, Dept of Elec Engineering & Computer Science, 77 Massachusetts Avenue, Cambridge, MA 02139; and Catherine M. Grgicak, PhD, Boston University, School of Medicine, Biomedical Forensic Sciences, 72 E Concord Street, Rm R806D, Boston, MA 02118*

After attending this presentation, attendees will understand the impact DNA template mass has on the ability to infer the true number of contributors. This presentation will assess the accuracy of three methods: (1) Maximum Allele Count (MAC); (2) Maximum Likelihood Estimator (MLE), which is available online; and, (3) via the online tool *NOC*It.[1,2]

This presentation will impact the forensic science community by demonstrating that a computational tool that utilizes a continuous probabilistic approach is the preferred method by which to assess the Number Of Contributors (NOC) as it returns higher accuracy rates for low-template samples. Probabilistic approaches also provide the probability distribution over n contributors. This provides the user with information regarding not only the most likely number of contributors but the uncertainty associated with the measurement.

The performance of the methods was tested on single-source samples as well as two-, three-, four-, and five-person mixtures, amplified using 29 cycles, and injected for 10 seconds. Samples were amplified using 0.25, 0.125, 0.063, 0.016, and 0.008ng. MAC and MLE rely on setting an Analytical Threshold (AT) to calculate the NOC. In this study, a constant threshold of 50 Relative Fluorescence Units (RFU) was utilized. Application of MAC and MLE also uses a stutter threshold to filter out the peaks in the stutter position of allelic peaks. The stutter filter specified by the manufacturer's manual was used to filter the stutter peaks at each locus. Allele frequencies from the Caucasian population specified in the AppliedBiosystems® AmpFlSTR® Identifiler® Plus PCR Amplification Kit User's Manual were used to test the *NOC*It and MLE methods.[3] Unlike the MLE and MAC methods, *NOC*It does not utilize any thresholds. Rather, it relies on a set of calibration standards to train the software and then utilizes this training set to model the baseline noise, stutter ratios, and the non-detection rates of stutter and allele peaks. Thus, 92 single-source samples were amplified utilizing the same amplification, run, and analysis protocols described above. Artifacts such as pull-up, -A, etc., were manually removed. The exported allele table was the calibration file used to train *NOC*It.

When 75 mock-evidence samples were interpreted, preliminary results suggest that, regardless of the method, DNA template mass had a significant effect on accurately inferring the NOC to a complex stain. Both the MLE and MAC methods resulted in similar accuracy rates, which ranged from 60% to 13% for the 0.25ng to 0.008ng samples, respectively. In contrast, the accuracy rates of *NOC*It were 87% to 27% for the 0.25ng to 0.008ng samples. MLE and MAC resulted in both overestimates and underestimates. Both methods overestimated 12% of the samples tested. These overestimations were the result of stutter peaks surpassing the stutter ratio threshold. One sample was overestimated when *NOC*It was used to infer the NOC. Underestimations were typically due to high levels of allele drop-out and/or allele sharing between large numbers of individuals. The percentage of samples resulting in underestimations for MLE, MAC, and *NOC*It were 43%, 52%, and 43%, respectively.

Unlike MAC, the MLE and *NOC*It methods provide a probability distribution on the NOC. In all cases, the distribution was unimodal. Further, the uncertainty associated with the result did not change with target but instead increased with the true NOC. For example, a 1:1:2:1:1 mixture amplified using 0.125ng resulted in *NOC*It returning two significant results: Pr(NOC=4) of 0.872 and Pr(NOC=5) of 0.128, suggesting this sample could have originated from four or five contributors. In summary, when utilizing MLE, 24% of the three-, four-, and five-person samples resulted in at least two NOCs exhibiting significant non-zero probabilities (i.e., Probability≥5%.). When *NOC*It was utilized, 9% of the three-, four-, and five-person samples resulted in at least two probable NOCs.

These preliminary results suggest that all methods are limited in their ability to accurately infer the NOC for samples containing low-template quantities. Though *NOC*It outperformed the other two methods at all templates, these results suggest samples which contain at least one contributor with fewer than ten cells are prone to underestimation. Accuracy rate data from the full study that includes an additional 360 samples will be presented. Data will also be provided regarding the minimum number of calibration samples needed to train *NOC*It.

References:

1. http://forensim.r-forge.r-project.org/. Accessed July 24, 2014.
2. http://www.bu.edu/dnamixtures/ Accessed July 24, 2014.
3. AppliedBiosystems. *AmpFlSTR® Identifiler® Plus PCR Amplification Kit User's Manual*. 1st ed: Applied Biosystems; 2006.

Number of Contributors, Complex DNA Interpretation, Low-Template DNA