Department of Forensic Biology
Charles S. Hirsch Center for Forensic Sciences
421 East 26th Street, New York, NY 10016
Telephone: 212-323-1200 Fax: 212-323-1590
Email: DNALab@ocme.nyc.gov
Official Website: www.nyc.gov/ocme

NEW YORK CITY OCME DEPARTMENT OF FORENSIC BIOLOGY: STRmix™ v2.4.08 MCMC UNCERTAINTY DATA REVIEW

OBJECTIVE

STRmix™ is a probabilistic genotyping software that utilizes a fully continuous approach to DNA sample interpretation. The NYC OCME performed an internal validation of STRmix™ v2.4.05 (1) and began using it in casework in January 2017. Subsequently, a performance check of an upgraded version of the software, STRmix™ v2.4.08 (2), was completed in June 2019 and has been used in casework since. In its evaluation of sample data, STRmix™ uses a variety of biological parameters (i.e. stutter, peak heights, peak height ratios, locus efficiency, etc.) and mathematical techniques to aid in mixture deconvolution, and the software also provides a statistical weight for comparisons of reference samples in the form of a likelihood ratio (LR) (3).

The mathematical technique used within STRmix™ is the widely used and accepted MCMC or Markov chain Monte Carlo method which allows for the assessment of different combinations of genotypes to explain the data most accurately. While this method is suitable, there is inherent variability between MCMC runs due to the random nature of the process. To aid in assigning statistical weight to a comparison, STRmix™ uses the Effective Sample Size (ESS) which is an estimate of the number of independent iterations within an individual MCMC run. This ESS value is ultimately used to generate a gamma distribution of the weights for each of the genotype combinations. This is done by re-sampling independent iterations of the MCMC run to account for the variability that would be found if the MCMC process was repeated on the same set of sample data. By using this distribution of weights, an LR distribution is created.

In combination with re-sampling of the allele frequency databases, the software then uses the highest posterior density (HPD) 1-sided lower-bound method to generate the LR assuming a 99% coverage of the LR distribution (3; 4). Notification was recently made by the developers of STRmix™ that this method may not be giving the desired 99% coverage. While this method is used in casework and is one of the layers of conservatism within the LR calculation, it could affect the assumptions in relation to the statistical weighting made on a sample (4). Within STRmix™, there is an option to not account for the MCMC uncertainty, and therefore exclude it from the LR calculation (3).

The following evaluation outlines a data review that was performed using STRmix™ v2.4.08 where MCMC uncertainty was not included within the LR calculations. The results of this data review were directly compared to the results of the NYC OCME's performance check of v2.4.08 from June 2019 (2).

SAMPLES

A total of 65 samples were used for evaluation. All samples were previously used for the internal validation of STRmix™ v2.4.05 (1) and the performance check of STRmix™ v2.4.08 (2).

- 5 single source samples with associated H_p true reference samples
- 30 apparent two-person (2p) mixtures with H_p true and H_d true reference samples that have LRs supporting inclusion, supporting exclusion, or within the uninformative range
- 30 apparent three-person (3p) mixtures with H_p true and H_d true reference samples that have LRs supporting inclusion, supporting exclusion, or within the uninformative range

EXPERIMENTS

All resulting LRs were generated using a theta (θ) value of 0.03; unified LRs were generated with the Factor of N! turned on and the 99.0% 1-sided lower-bound HPD value (1000 iterations) applied. Additionally, MCMC uncertainty was checked off before all LR calculations were performed for any v2.4.08-MU runs. All other settings and parameters follow the current standard protocols used by the Department of Forensic Biology at the OCME for casework with STRmix™ v2.4.08.

Note: For the purposes of this study, runs completed during the original performance check of STRmix™ v2.4.08 in June 2019 may be referred to as 'v2.4.08' and new runs completed not including MCMC uncertainty may be referred to as 'v.2.4.08-MU'. All differences within the LRs are represented as the order of magnitude (OM) change when using v2.4.08-MU in comparison back to v2.4.08 results with a positive OM resulting in a higher LR in v2.4.08-MU and a negative OM resulting in a lower LR in v2.4.08-MU.

Experiment #1: Likelihood Ratio Comparison

Experimental Design

Likelihood ratios were calculated for 65 single source, apparent 2p, and apparent 3p samples in v2.4.08-MU. These calculations were performed using the original deconvolutions and LR seed of these same samples from Experiment 1 of the performance check of v2.4.08 (2).

Results

1p

All LR calculations for each ethnic group between v2.4.08 and v2.4.08-MU were within one order of magnitude (OM). Any observed changes resulted in LRs being slightly higher when the MCMC uncertainty function was not used. The largest difference was a 0.15 OM difference for a unified LR seen in sample 16-3M_50pg when using v2.4.08-MU. All conclusions (whether exclusion, support for exclusion, support for inclusion, or falling within the uninformative range) were the same between v2.4.08 and v2.4.08-MU (Tables 1-3).

For each ethnic group, both the HPD and unified LR results were highly correlated between v2.4.08 and v2.4.08-MU with *r* values greater than 0.99 (Table 4).

2p

For the majority of the apparent 2p mixtures (25 out of 30 samples), the LRs calculated for each ethnic group between v2.4.08 and v2.4.08-MU were within +/-1 OM. The majority of the changes observed showed higher LRs with the MCMC uncertainty function turned off. The largest difference was 0.82 OM for the HPD and unified LRs seen in the samples listed below when using v2.4.08-MU.

- 35-3p_CST8F_27M30_28M30_75pg_5-1-1_Newton
- 38-M1_C4_100_15-1_6M_22F_Athena when using v2.4.08-MU.

All conclusions (whether exclusion, support for exclusion, support for inclusion, or falling within the uninformative range) for these 25 samples were the same between v2.4.08 and v2.4.08-MU (Tables 5-7).

For the remaining five apparent 2p samples listed below, the point estimate LRs did not change, as expected, when using v2.4.08-MU; however, there were larger OM differences observed for the HPD and unified LRs.

- 22-M3_37.5pg_1-3-5-1_3M-17F-18F-29F_Newton
- 36-3p_CST8F_27M30_28M30_50pg_5-1-1_Newton
- 93-3p_23M30-CST_14M-18F30_37.5pg_5-2-1_Newton
- 12-M1_C2_500_20-1_6M_22F_Athena
- 88-M2_C4_100_4-1_15M_5F_Newton)

The differences ranged between 1.62 to 5.52 OM with use of v2.4.08-MU. Although the OM was greater in these samples, the change in LRs were low with the LRs approaching the point estimate value when the MCMC uncertainty was turned off. All conclusions for these samples were support for exclusion and remained the same between v2.4.08 and v2.4.08-MU (Tables 5-7).
For each ethnic group, both the HPD and unified LR results for all 2p mixtures were highly correlated between v2.4.08 and v2.4.08-MU with *r* values greater than 0.99 (Table 8).

3p

With the exception of one sample, all LRs calculated for each ethnic group between v2.4.08 and v2.4.08-MU were within +/-1 OM. The majority of the LRs were higher with the MCMC uncertainty function turned off. The largest difference observed was 0.47 OM in the HPD and unified LRs seen in sample 14-M3_50pg_1-3-3-1_3M-17F-18F-29F_Newton when v2.4.08-MU was used. For one sample, 19-M1_C3_250_5-2-1_12F_13M_6M_Newton, a difference between 1.85-2.11 OM was observed for the HPD and unified LRs after the use of v2.4.08-MU. Similar to the results observed for the 2p mixtures, the change in LRs associated with this sample were still low with the results higher and approaching the point estimate value when the MCMC uncertainty was not applied (Tables 9-11). For 29 out of 30 samples, all conclusions (whether exclusion, support for exclusion, support for inclusion, or falling within the uninformative range) were the same between v2.4.08 and v2.4.08-MU.

Only one sample, 07-M1_C2_500_5-1-1_12F_13M_6M_Athena, had a conclusion change. The conclusion went from support for exclusion to falling within the uninformative range. The lowest unified LR for v2.4.08 was 7.73×10^{-4} (0.000773) and the lowest unified LR for v2.4.08-MU was 1.05×10^{-3} (0.00105). While the difference between these two LRs is small, 0.13 OM, both values are also close to the lower boundary of the uninformative range, 0.001-1,000. While inherent variability between MCMC runs could account for small changes to LR values (see Experiment 3 below), disabling the MCMC uncertainty feature also caused a small enough difference in the LR result to change the overall conclusion (Tables 9-11). Even though this sample was run as an H_d true hypothesis, the LR calculation results for both v2.4.08 and v2.4.08-MU are as expected since

the comparison profile does share some alleles with the true contributors across the overall mixture.

For each ethnic group, both the HPD and unified LR results for all 3p mixtures were highly correlated between v2.4.08 and v2.4.08-MU with *r* values greater than 0.99 (Table 12).

Discussion

Overall, only minor differences were observed when comparing LR calculations between v2.4.08 and v2.4.08-MU. The average change across all samples was small. For the HPD LRs, the average OM change and standard deviation (SD) are as follows: 1p (0.02,0.03), 2p (0.74,1.48), and 3p (0.17,0.36). For the unified LRs, the average OM change and standard deviation (SD) are as follows: 1p (0.02,0.05), 2p (0.74,1.48), and 3p (0.17,0.36) (Tables 4, 8, and 12). The largest LR changes were observed for those in support for exclusion where v2.4.08-MU generally raised the LR. A direct comparison of all samples is represented in Figure 1, showing that the majority of the LR results with the MCMC uncertainty setting on are highly correlated with those LR results with the MCMC uncertainty setting off.

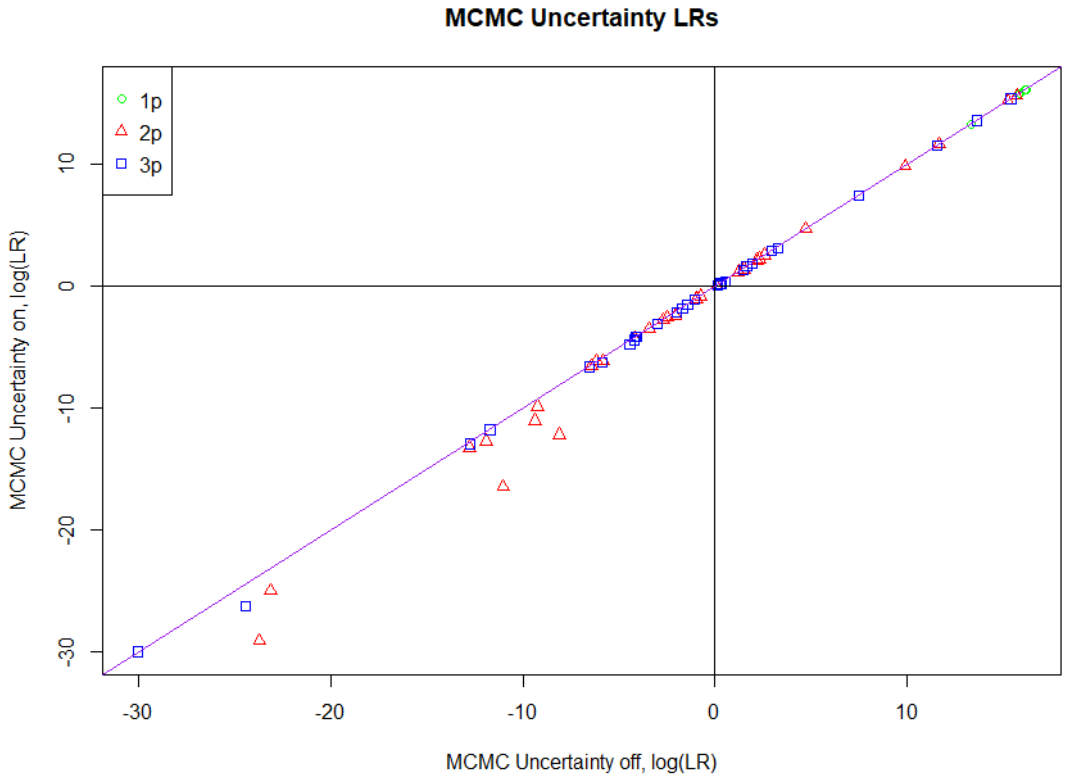

Figure 1: Comparison of log(LRs) in STRmix™ v2.4.08 with the MCMC uncertainty setting turned on versus the MCMC uncertainty setting turned off for 1p, apparent 2p,

and apparent 3p samples. The solid line represents the 1:1 result between the log(LRs).

Experiment #2: Non-Contributor Evaluation

Experimental Design

Within the performance check of v2.4.08, a database of non-contributor LRs were examined with a focus on those which fell outside of the laboratory's uninformative range of 0.001-1,000 (2). For any instances of a non-contributor supporting inclusion (i.e. showing an LR greater than 1,000), additional LR calculations were performed with the seed set to apply θ, Factor of N!, and the 99% 1-sided lower-bound HPD value. These samples were run in v2.4.08-MU to apply the same factors of conservatism. Because the database function only allows for use of one population, the database LRs and all comparisons were completed using the Caucasian population frequencies.

Results

The re-calculations of all 19 samples with the use of additional factors of conservatism were within one OM of each other and highly correlated between v2.4.08 and v2.4.08-MU with the largest difference being 0.22 OM. All *r* values were greater than 0.99 (Tables 13-14). For both the unified and HPD LRs, two samples fell above 1,000 (supporting inclusion) - 38-M1_C4_100_15-1_6M_22F_Newton.csv in comparison to profile 3544 and 53-M2_C4_100_5-5-1_17M_27F_9M_Newton.csv in comparison to 9515. These two samples were the same two samples that remained above 1,000 during the v2.4.08 performance check (2). Similar to that performance check, these results are due to the nature of the DNA profiles of the two samples where the minor profile shares common alleles at lower amounts of input DNA and is not due to a failure of STRmix™. Overall, applying these factors of conservatism to the LR calculation with the MCMC uncertainty turned off is concordant with results with the MCMC uncertainty turned on. No additional H_d false comparisons were observed.

Experiment #3: Precision Testing

Experimental Design

As the calculation of the 99% 1-sided HPD likelihood ratio within the software relies on a random sampling process, replicate LR calculations result in slightly different LR values. For three samples representing a range of LRs (high, low, and within the uninformative range), an additional 9 replicate LR calculations were performed applying the MCMC uncertainty feature and 9 replicate LR calculations were performed disabling the MCMC uncertainty feature. A comparison was

made between the LRs in order to observe the variability between the LRs for replicate runs of the same sample.

Results

The replicate calculations for all three samples were highly consistent within each set of 10 runs (Tables 15-17; Figure 2). All replicate LRs were within +/-1 OM of the mean for the replicate set as is shown in Figure 2.

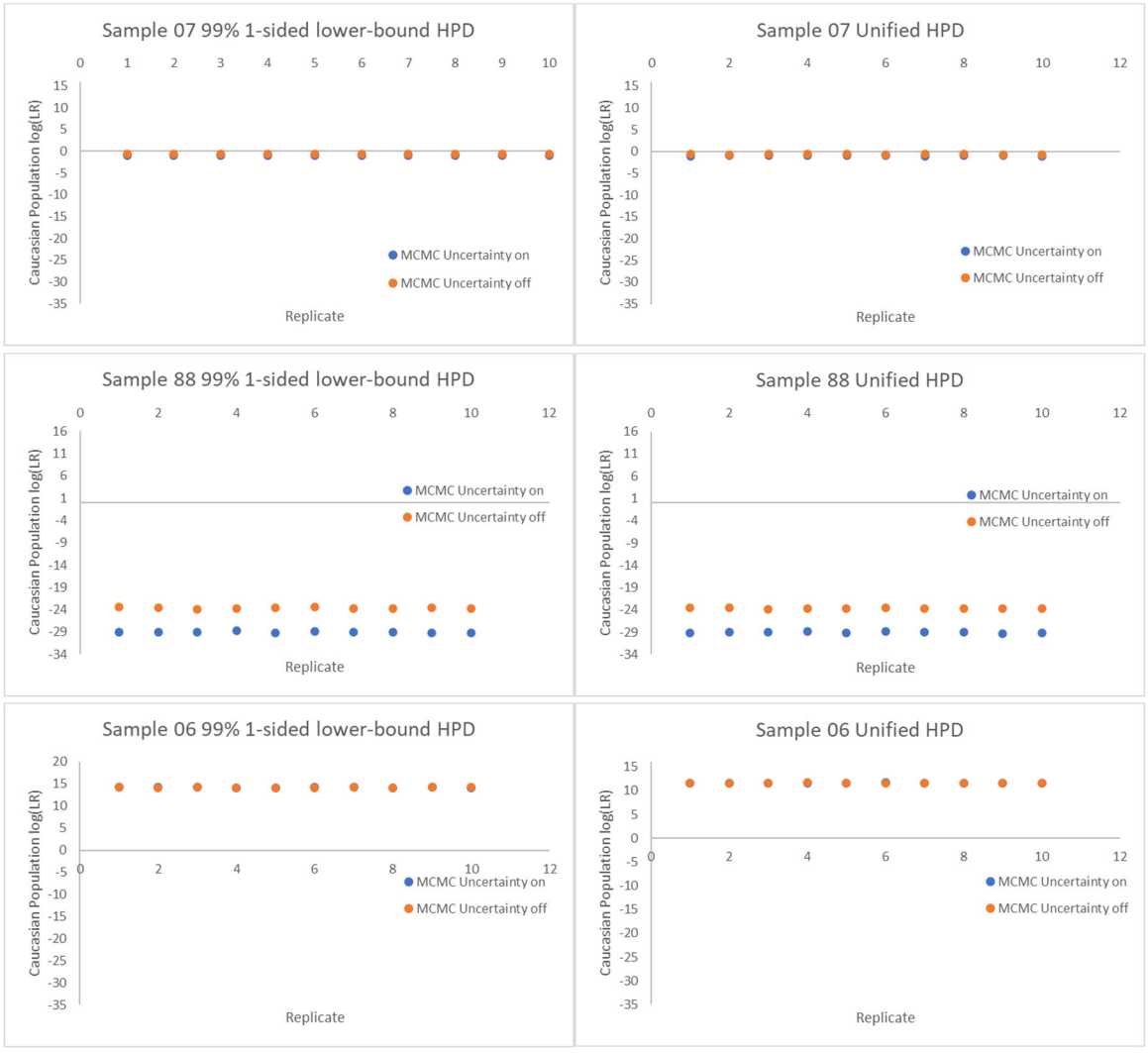

Figure 2: Comparison of log(LRs) in STRmix™ v2.4.08 with the MCMC uncertainty setting turned on versus the MCMC uncertainty setting turned off for replicates of apparent 2p (Samples 06 and 88) and apparent 3p (Sample 07) samples.

For samples 06 and 88, the conclusions (whether exclusion, support for exclusion, support for inclusion, or falling within the uninformative range) were all the same within replicate sets for v2.4.08 and v2.4.08-MU and between the replicate sets for each sample. As both samples 06 and 88 had LRs far from the uninformative range, the variability induced by the replicate LR calculations was not enough to change the conclusions (Tables 16-17).

For sample 07, all LRs were close to the lower boundary of the uninformative range, 0.001. The inherent variability from the re-sampling process in calculating the LR, for both the MCMC uncertainty function on and off, affected the conclusion with some of the replicates supporting exclusion while others fell within the uninformative range. As the samples were already within +/- 1 OM of the uninformative range lower boundary, even a small amount of variability could move a sample from one conclusion to the other. While most replicates run in v2.4.08 supported exclusion, replicates run with v2.4.08-MU mostly fell within the uninformative range (Table 15). Similar to the results seen in Experiments 1 and 2, the means of the replicate runs for all racial groups were higher with v2.4.08-MU except for the 99% 1-sided lower-bound HPD in Caucasians for Sample 06 (Table 18). The variances showed the same trend with all variances for v.2.4.08-MU being higher than v2.4.08 with the same exception, the 99% 1-sided lower-bound HPD in Caucasians for Sample 06 (Table 19). This highlights the fact that generally, v2.4.08-MU does not account for the uncertainty of the MCMC process and therefore, can exhibit more variability from run to run. Based on this data, having the MCMC function on tends to generally both lower the LR and lessen the variability seen from run to run.

Conclusion

This data review was completed to assess the differences between LR calculations within STRmix™ v2.4.08 with the MCMC uncertainty function turned on and with the MCMC uncertainty function turned off. Generally, the data review showed that the LRs produced using both methods are consistent and highly correlated to each other. Additionally, for those LR values where variability was seen between the two methods, the LR results were generally higher when the MCMC uncertainty function was not enabled. Overall, these results demonstrate that applying the MCMC uncertainty function applies a layer of conservatism that will most often lead to the same or lower LRs for a given comparison. Therefore, it is recommended to leave the function turned on even though it may not be giving the desired 99% coverage.

References

1. **Office of Chief Medical Examiner of the City of New York.** *Internal Validation of STRmix™ V2.4 for Fusion NYC OCME.* New York City : NYC OCME, December 20, 2019.
2. —. *STRmix™ V2.4.08 Performance Check.* New York City : NYC OCME, July 24, 2018.
3. **Institute of Environmental Science and Research Limited.** *STRmix™ V2.4 User's manual.* July 6, 2016.
4. —. The highest posterior density for the Monte Carlo effect in STRmix™. June 30, 2020. Message to Eugene Lien.