STRmix™ Collaborative Exercise on DNA Mixture Interpretation

Jo-Anne Bright [a], Kevin Cheng [a,b], Zane Kerr [a], Catherine McGovern [a], Hannah Kelly [a], John Buckleton [a,b]

[a] Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142 New Zealand
[b] University of Auckland, Department of Statistics, Auckland, New Zealand

Preface

The work and results for this manuscript has been published in Forensic Science International: Genetics, Volume 40, May 2019, Pages 1-8, https://doi.org/10.1016/j.fsigen.2019.01.006 [1]. The authors refer the reader to the paper for more detail and other insights not covered within this manuscript. Figures 3, 4, 5, and 6, and Tables 1 and 2 are reprinted from Forensic Science International: Genetics, 40, Bright et al., STRmix™ collaborative exercise on DNA mixture interpretation, Pages 1-8, Copyright 2019, with permission from Elsevier.

Introduction

Accuracy and precision are two of the main principles of the scientific method. Accuracy is how estimates compare to the true answer or true standard; whereas precision is the repeatability of estimates. Perfect accuracy and perfect precision are extremely unlikely. Instead, scientists in all disciplines aim to achieve high accuracy and high precision; this is where estimates are closely clustered around the true answer.

Forensic biologists have the same goal when assigning a match statistic. The likelihood ratio (*LR*) is acknowledged to be a powerful and relevant statistic used to assign the weight of evidence and is recommended by the DNA commission of the International Society of Forensic Genetics (ISFG) for mixed DNA profiles [2]. The *LR* is a ratio of two conditional probabilities, probability densities, or numbers proportional to these, where the probability of the evidence (the DNA profile) is evaluated under two competing and mutually exclusive propositions, for example:

H_1: The suspect contributed to the DNA

H_2: The suspect did not contribute to the DNA

Utilising sound experimental design, an understanding of biological modelling assumptions, and vigorous validation studies, including ground truth known trials, forensic practitioners can provide the judicial system with reliable assignments of the *LR*. In addition, deliberate decisions may be made that lead to a conservative estimate of the *LR*, thus conceding any doubt to the POI. Precision is important in a method to ensure that the same conclusion is drawn from repeat interpretation of the same evidence.

NIST have undertaken a number of inter-laboratory studies to “assess effectiveness and accuracy of analytical protocols used across participating laboratories”, most recently with MIX13 [3]. This study highlighted differences in the reported match statistics between laboratories due to the use of different statistics (RMP, CPI, or *LR*), different parameters (e.g. analytical thresholds), different

allele frequencies, and different subjective decisions being made by different analysts. These differences results in decreased precision of the estimates.

The development and transition to probabilistic genotyping (PG) solutions using semi-continuous and continuous models was partly in a response to intra and inter-laboratory trials in order to improve reproducibility of the results. However, there are still some aspects of DNA profile interpretation that are not controlled by PG software, and therefore some variability is still to be expected.

More recently a number of inter-laboratory studies using PG solutions have been reported. For example, the use of LRmix and LRmix Studio in the Euroforgen-NoE and GHEP-ISFG collaborative exercises, and the use of STRmix™ [4-6].

These studies showed that there were still differences between results with different parameters used, including analytical and stochastic thresholds, and allele frequencies. In the current work, we investigated the effect on the *LR* of truly subjective decisions such as the assignment of the number of contributors and propositions.

Method

This collaborative exercise included 174 participants, across 42 laboratories. It aimed to identify the sources of variation remaining in DNA interpretation when other sources of variation were removed. Each participant was asked to analyse two complex GlobalFiler™ mixtures from the PROVEDIt dataset [7] , assign the number of contributors and form suitable propositions based on the provided case circumstances (Table 1). In addition, profiles already analysed in GeneMapper ID-X were provided as STRmix™ input files for those participants unable to analyse the mixture themselves.

Table 1: Samples, case circumstances, and references provided to analysts (replicated from Table 1 in [1] with permission).

Sample	Case Circumstance	References
RD14-0003-44_45_46_47-1;1;4;1-M3a-0.105GF-Q0.8	The DNA profile was obtained from a semen stained sample from underwear collected from the complainant after an alleged sexual assault. The male complainant alleges he has been sexually assaulted by two male individuals. DNA swabs from the complainant and a person of interest have been taken for analysis.	Two reference profiles have been submitted, one described as having come from the complainant and one from the suspect.
RD14-0003-30_31_32-1;4;4-M2a-0.75GF-Q0.6	The DNA profile was obtained from a semen stained anal swab after an alleged sexual assault. The male complainant alleges he has been sexually assaulted by two male	Two reference profiles have been submitted, one described as having come from the complainant (“Sample

	individuals. DNA swabs from the complainant and a person of interest have been taken for analysis.	2 complainant") and one from the suspect ("Sample 2 suspect").

Participants were also provided with the STRmix™ run parameter files and settings so that allele frequency, FST, and other PG parameters were fixed [8].

Results and Discussion

Sample Two - RD14-0003-30_31_32-1;4;4-M2a-0.75GF-Q0.6

Sample two is an experimentally designed three-person mixture with a target contributor ratio of 1:4:4. The target total template amount was 750 pg.

From the one-hundred and seventy-four (174) participants, there was a total of one-hundred and seventy-six (176) submissions, as some participants submitted multiple results. The results from Figure 6 in [1] have been replicated in Figure 1 below. One-hundred and fifty-one (151) of these submissions were interpreted under the assumption of three-contributors – shown as red triangles. The remaining submissions were interpreted assuming four-contributors – shown as blue diamonds.

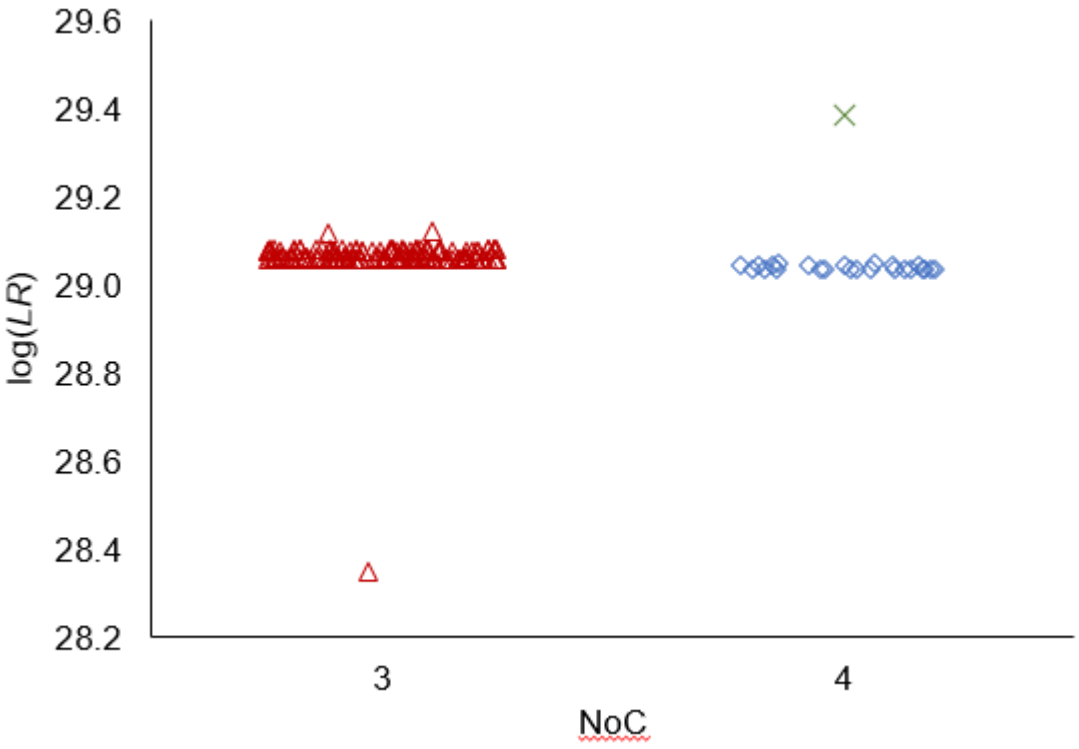


Figure 1: Log(*LR*) versus number of contributors (NoC) – 3 (red triangles) and 4 (blue diamonds) – for sample two. A green cross indicates a submission using a bespoke method to manage artefacts (replicated from Figure 6 in [1] with permission).

All submissions interpreted the mixture as either:

H_1: The DNA originated from the complainant, the suspect, and one unknown-unrelated individual.

H_2: The DNA originated from the complainant, two unknown-unrelated individuals.

Or,

H_1: The DNA originated from the complainant, the suspect, and two unknown-unrelated individuals.

H_2: The DNA originated from the complainant, three unknown-unrelated individuals.

There are two outliers, one of which is the submission giving the lowest log(*LR*) of 28.3. In this interpretation, the analyst ignored the locus CSF1PO. In this particular case, omitting a locus, resulted in a lower *LR* compared with the other observations interpreted assuming three contributors. The reason why this locus was ignored was not provided (we had not requested information of this type). This is an example demonstrating that subjective decisions can impact the precision of the *LR*.

The second outlier was an observation from a submission where the analyst interpreted the mixture using a bespoke method to manage artefacts, which resulted in a log(*LR*) of 29.4.

Ignoring the two outliers, the log(*LR*), ranged from 29.0 to 29.1, regardless of the assigned number of contributors. The intra-laboratory range, a measure of the variability between analysts within the same laboratory, was 0.05.

The small amount of variability between *LR*s is because STRmix™ utilizes a Markov chain Monte Carlo (MCMC) interpretation method. This is a method that involves random sampling. A typical *LR* range of about one-order of magnitude is expected due to MCMC variation using STRmix™. Inspection of the results for sample two showed minimal variation and consequently high precision.

Sample 2, with the exception of the bespoke artefact management method, demonstrates two factors that can impact the precision of the *LR*. These include:

- subjective decisions around ignoring a locus; and
- MCMC variation.

Sample One - RD14-0003-44_45_46_47-1;1;4;1-M3a-0.105GF-Q0.8

Sample one is an experimentally designed four-person mixture with the target contributor ratios of 4:1:1:1. This sample contained a target total DNA amount of 105 pg. This sample was more complex than sample two, and five analysts stated that the mixture was too complicated electing not to interpret it.

There was a total of one-hundred and seventy-three (173) submissions, and eleven interpreted the profile assuming three contributors. The remaining submissions assigned four-contributors. The results from Figure 3 in [1] have been replicated in Figure 2.

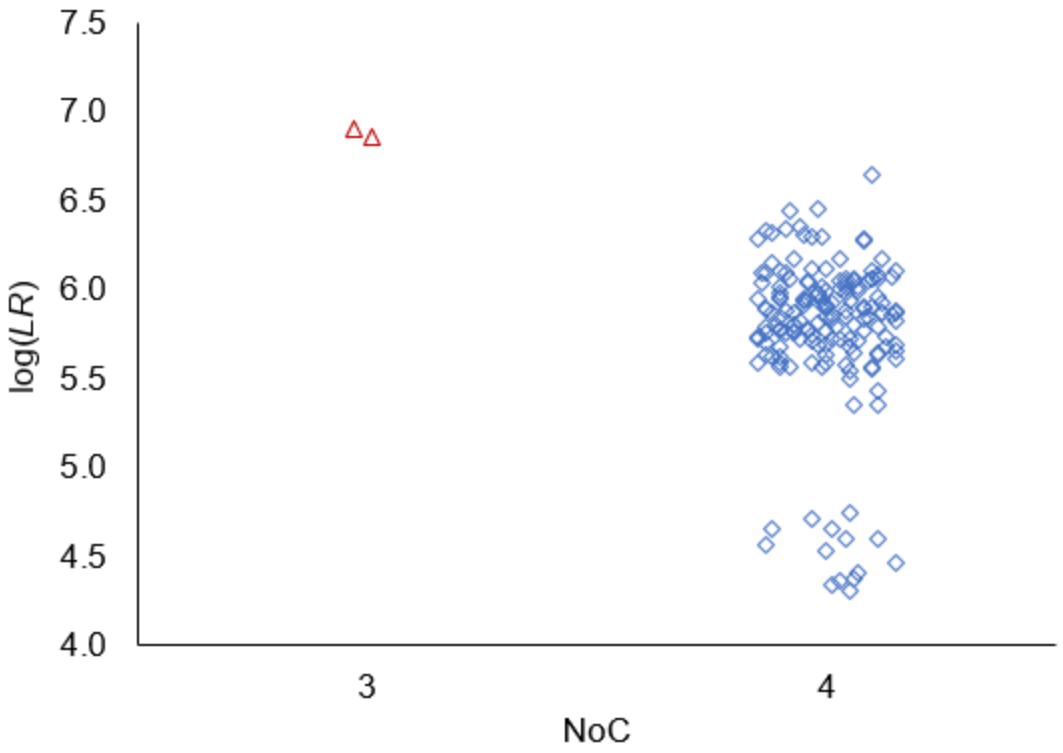


Figure 2: Log(*LR*) versus interpretations under different number of contributors (NoC) – 3 (red triangles) and 4 (blue diamonds) – for sample one. Nine (9) of the eleven (11) interpretations under the assumption of three contributors resulted in an exclusion (*LR* = 0) and are not plotted in the figure (replicated from Figure 3 in [1] with permission).

Whilst eleven submissions interpreted the profile assuming three contributors, in Figure 2, only two observations are shown. This is because nine of these submissions resulted in an *LR* of 0 because of an exclusion (*LR* = 0) at the locus D18S51. These exclusionary *LR*s are not shown in this plot. Two submissions reinterpreted the mixture ignoring D18S51 and obtained the *LR*s in Figure 2 above.

Figure 3 reproduces the peaks observed at D18S11. Given the case circumstances it is reasonable to expect the complainant's genotype to be present in the profile. This aligns with the peaks 13 and 18, labelled with the letter 'C' in blue. Assuming three contributors, the remaining five peaks (12, 14, 15, 16, and 17) need to be explained by the two other contributors. This means at least one of these peaks has to be explained as drop-in. This could be either the 16 or the 17 peak as these fall below the empirical drop-in cap of 150 rfu.

The suspect's genotype is [13, 15] labelled by the letter 'S' in red. For this genotype combination to be possible, the four remaining peaks (12, 14, 16, and 17) need to be explained by the remaining contributor. This can only be possible if both the 16 and 17 peaks are accepted as drop-in. This is statistically very unlikely and STRmix™ did not accept this genotype combination with two drop-in peaks in any of these interpretations. As a result, an *LR* of zero was obtained when comparing the suspect to the interpretation assuming three contributors.

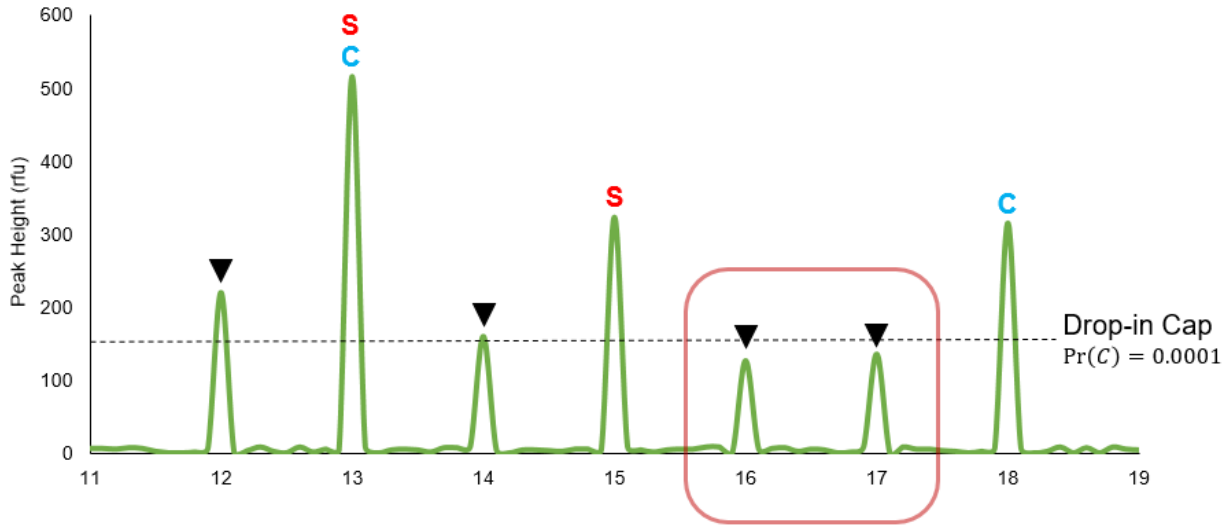


Figure 3: The complainant's and the suspect's known genotypes at this locus are respectively labelled as the letters 'C' (blue) and 'S' (red). Under the assumption of three contributors, the four peaks (12, 14, 16, and 17) labelled by the black triangles need to be explained by the one contributor. This means that both the 16 and 17 peaks, with heights below the drop-in cap of 150 rfu (dashed black line), need to be explained as drop-in, which is statistically unlikely (reproduced from Figure 4 in [1] with permission).

As stated earlier, an *LR* range within one-order of magnitude can be explained by MCMC variation. However, in Figure 2 the log(*LR*) range for those that interpreted the mixture ranged from 4.3 to 6.6. The intra-laboratory or within-laboratory range was 2.09.

The disparity observed in the *LR* was due to differences in GeneMapper ID-X analysis methods, resulting in differences in the peaks labelled and the peak heights. This is because participants either analysed raw electropherograms (epgs) using their analysis methods in GeneMapper ID-X or used the pre-analysed inputs provided.

We identified four different input files corresponding to four different analysis methods or analysis settings that can be configured within GeneMapper ID-X. Table 2 shows a summary of the different analysis methods and Figure 4 shows how the different analysis methods impact the *LR*.

Table 2: Differences in analysis methods (reproduced from Table 2 in [1] with permission).

Method	**A**	**B**	**C**	**D**
Smoothing	Light	Light	Light	No
Normalization	No	Yes	No	No
Baseline Window	51	51	33	51

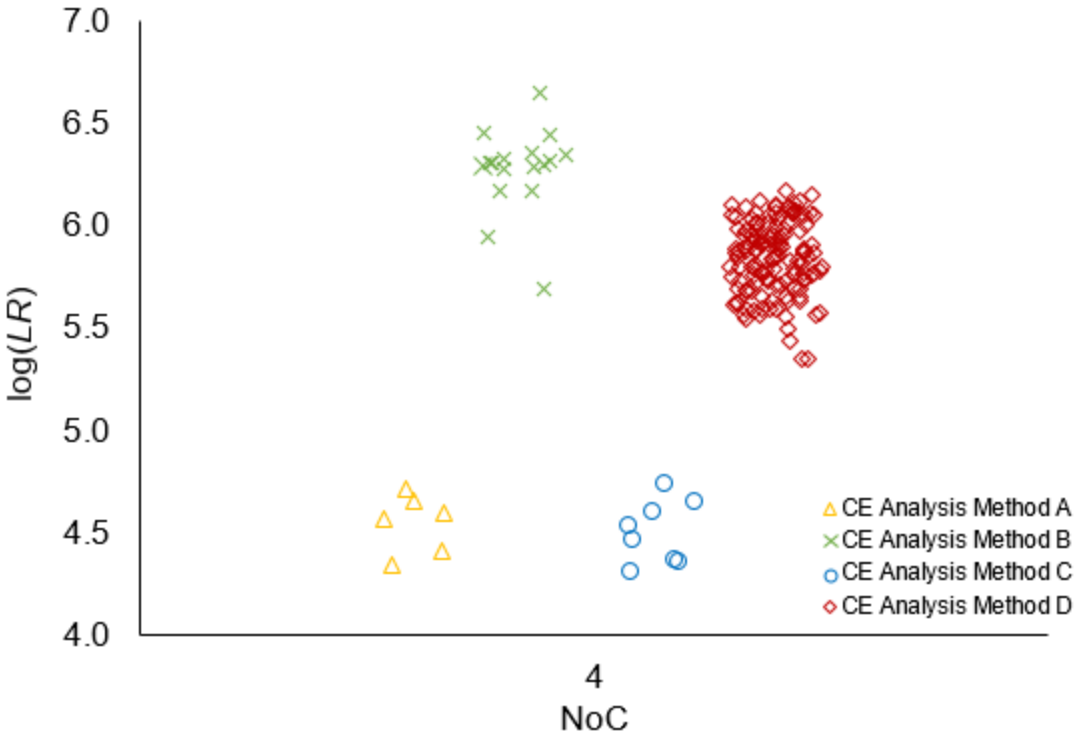
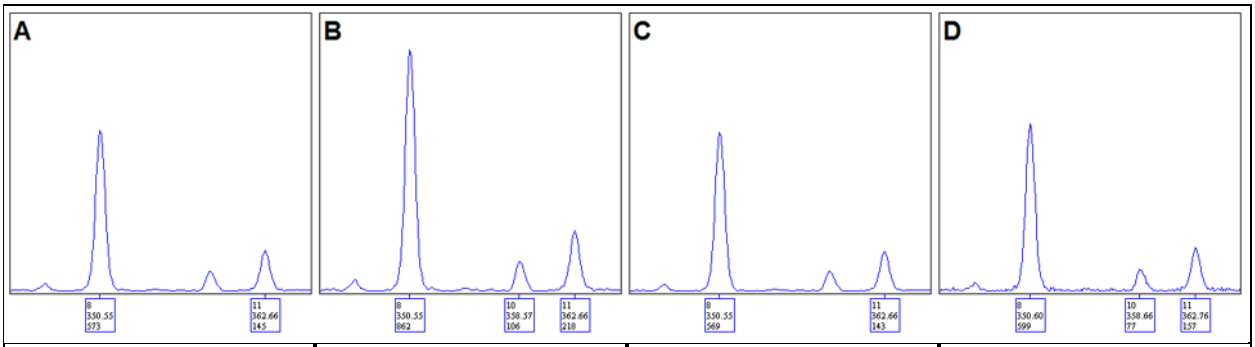


Figure 4: Log(*LR*) under the assumption of four contributors for sample one. The different symbols and colours represent different GeneMapper ID-X analysis methods (replicated from Figure 5 in [1] with permission).

To understand why each analysis method results in different inputs, Figure 5 shows the epg for the TPOX locus given each of the analysis methods. Peaks observed in analysis method B are taller than the other analysis methods, and analysis methods B and D both have an extra peak – the 10 peak above the analytical threshold of 75 rfu.



Analysis Method A		Analysis Method B		Analysis Method C		Analysis Method D	
Allele	Height (rfu)	Allele	Height (rfu)	Allele	Height (rfu)	Allele	Height (rfu)
8	573	8	862	8	569	8	599
		10	106			10	77
11	145	11	218	11	143	11	157

Figure 5: Replicated illustration of the TPOX locus for sample one with a table of observed peak heights. Panels A-D correspond to the different analysis methods outlined in Table 2.

Methods B and D take more information into the STRmix™ interpretation and this results in more weight being assigned to the genotype [8,10]. This genotype aligns with the POI's genotype and therefore *LR*s assigned using input files generated using methods B and D are slightly higher than the *LR*s assigned using the input files from analysis methods A and C. This just one example that we observed in our results where the analysis method affected the precision of the match statistic.

In Figure 4, by identifying and categorizing the individual methods used, we can see that if we did not take into consideration these different analysis methods there will be a decrease in precision. However, looking at the group of *LR*s for each analysis method independently, there is a decrease in the amount of variability and consequently an increase in the precision of the estimates.

All, but one, submission interpreted the mixture as either:

> H_1: The DNA originated from the complainant, the suspect, and one unknown-unrelated individual.
>
> H_2: The DNA originated from the complainant, two unknown-unrelated individuals.

Or,

> H_1: The DNA originated from the complainant, the suspect, and two unknown-unrelated individuals.
>
> H_2: The DNA originated from the complainant, three unknown-unrelated individuals.

One submission, using analysis method B, interpreted the mixture without assuming the presence of the complainant's DNA in the evidence profile. The proposition set was,

> H_1: The DNA originated from the suspect, and three unknown-unrelated individuals.
>
> H_2: The DNA originated from four unknown-unrelated individuals.

The resulting log(*LR*) is 5.683 and is within one-order of magnitude to the other submissions from analysis method B (the largest log(*LR*) = 6.644). This is because the complainant's profile is the major component within this mixture. Therefore, conditioning on the complainant does not greatly assist with the deconvolution of this mixture.

Excluding the differences generated due to the GeneMapper ID-X analysis methods, the remaining variability between the *LR*s are all within one-order of magnitude. This variability can be attributed to the random sampling process of MCMC.

Conclusion

We show the precision from the submitted results, with the likelihood ratio ranging from 2.02×10^4 to 7.92×10^6 (sample one) and 2.21×10^{28} to 2.43×10^{29} (sample two). Differences in the *LR*s calculated for each submission can be attributed to:

- Varying the number of contributors assumed when interpreting a profile.
- Exclusion of some loci when interpreting a profile.
- Differences in CE data analysis methods, leading to variation in peak heights.
- Run-to-run variation due to random sampling inherent to the MCMC method.

Despite these sources of variation, this study demonstrates very good precision among participants. This is a desirable and pleasing outcome as it reinforces some of the key findings in existing collaborative exercises, such the convergence of results between laboratories and within laboratories as a result of using of PG software, and the importance of increased training to forensic biologists around the use and understanding of PG software. Furthermore, the reproducibility of *LR* reported in this exercise can be attributed to the use of PG software and supports the ongoing transition of forensic laboratories to probabilistic genotyping methods.

Acknowledgements

The work and results for this manuscript has been published in Forensic Science International: Genetics, Volume 40, May 2019, Pages 1-8, https://doi.org/10.1016/j.fsigen.2019.01.006 [1]. The authors of this manuscript would like to acknowledge Tamyra R. Moretti, Michael A. Smith, Frederick R. Bieber, Bruce Budowle, Michael D. Coble, Rashed Alghafri, Paul Stafford Allen, Amy Barber, Vickie Beamer, Christina Buettner, Melanie Russell, Christian Gehrig, Tacha Hicks, Jessica Charak, Kate Cheong-Wing, Anne Ciecko, Christie T. Davis, Michael Donley, Natalie Pedersen, Bill Gartside, Dominic Granger, MaryMargaret Greer-Ritzheimer, Erick Reisinger, Jarrah Kennedy, Erin Grammer, Marla Kaplan, David Hansen, Hans J. Larsen, Alanna Laureano, Christina Li, Eugene Lien, Emilia Lindberg, Ciara Kelly, Ben Mallinder, Simon Malsom, Alyse Yacovone-Margetts, Andrew McWhorter, Sapana M. Prajapati, Tamar Powell, Gary Shutler, Kate Stevenson, April R. Stonehouse, Lindsey Smith, Julie Murakami, Eric Halsing, Darren Wright, Leigh Clark, and Duncan A. Taylor for their participation in this collaborative study.

This work was supported in part by grant 2017-DN-BX-K541 from the US National Institute of Justice. The opinions or assertions contained in this document are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Department of Justice.

References

[1] J.A. Bright, K. Cheng, Z. Kerr, C. McGovern, H. Kelly, T.R. Moretti, M.A. Smith, F.R. Bieber, B. Budowle, M.D. Coble, R. Alghafri, P.S. Allen, A. Barber, V. Beamer, C. Buettner, M. Russell, C. Gehrig, T. Hicks, J. Charak, K. Cheong-Wing, A. Ciecko, C.T. Davis, M. Donley, N. Pedersen, B. Gartside, D. Granger, M. Greer-Ritzheimer, E. Reisinger, J. Kennedy, E. Grammer, M. Kaplan, D. Hansen, H.J. Larsen, A. Laureano, C. Li, E. Lien, E. Lindberg, C. Kelly, B. Mallinder, S. Malsom, A. Yacovone-Margetts, A. McWhorter, S.M. Prajapati, T. Powell, G. Shutler, K. Stevenson, A.R. Stonehouse, L. Smith, J. Murakami, E. Halsing, D. Wright, L. Clark, D.A. Taylor, J. Buckleton, STRmix collaborative exercise on DNA mixture interpretation, Forensic Sci Int Genet 40 (2019) 1-8.

[2] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures, Forensic Science International 160(2) (2006) 90-101.

[3] J.M. Butler, M.C. Kline, M.D. Coble, NIST Interlaboratory Studies Involving DNA Mixtures (MIX05 and MIX13): Variation Observed and Lessons Learned, Forensic Science International: Genetics 37 (2018) 81-94.

[4] L. Prieto, H. Haned, A. Mosquera, M. Crespillo, M. Alemañ, M. Aler, F. Álvarez, C. Baeza-Richer, A. Dominguez, C. Doutremepuich, M.J. Farfán, M. Fenger-Grøn, J.M. García-Ganivet, E. González-Moya, L. Hombreiro, M.V. Lareu, B. Martínez-Jarreta, S. Merigioli, P. Milans del Bosch, N. Morling, M. Muñoz-Nieto, E. Ortega-González, S. Pedrosa, R. Pérez, C. Solís, I. Yurrebaso, P. Gill, Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles, Forensic Science International: Genetics 9 (2014) 47-54.

[5] P.A. Barrio, M. Crespillo, J.A. Luque, M. Aler, C. Baeza-Richer, L. Baldassarri, E. Carnevali, P. Coufalova, I. Flores, O. García, M.A. García, R. González, A. Hernández, V. Inglés, G.M. Luque, A. Mosquera-Miguel, S. Pedrosa, M.L. Pontes, M.J. Porto, Y. Posada, M.I. Ramella, T. Ribeiro, E. Riego, A. Sala, V.G. Saragoni, A. Serrano, S. Vannelli, GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: Results and evaluation, Forensic Science International: Genetics 35 (2018) 156-163.

[6] S. Cooper, C. McGovern, J.A. Bright, D. Taylor, J. Buckleton, Investigating a common approach to DNA profile interpretation using probabilistic software, Forensic Science International: Genetics 16 (2015) 121-131.

[7] L.E. Alfonse, A.D. Garrett, D.S. Lun, K.R. Duffy, C.M. Grgicak, A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Science International: Genetics 32 (2018) 62-70.

[8] H. Kelly, J.-A. Bright, M. Kruijver, S. Cooper, D. Taylor, K. Duke, M. Strong, V. Beamer, C. Buettner, J. Buckleton, A sensitivity analysis to determine the robustness of STRmix™ with respect to laboratory calibration, Forensic Science International: Genetics 35 (2018) 113-122.