Article:

Bright, J.A., Stevenson, K.E., Curran, J.M., & Buckleton, J.S. (2015). **The variability in likelihood ratios due to different mechanisms**. *Forensic Science International: Genetics*, *14*, 187–190.

INSTITUTE OF ENVIRONMENTAL SCIENCE AND RESEARCH LIMITED
Kenepuru Science Centre: 34 Kenepuru Drive, Kenepuru, Porirua 5022 | PO Box 50348, Porirua 5240, New Zealand
T: +64 4 914 0700 **F:** +64 4 914 0770
www.esr.cri.nz

The variability in likelihood ratios due to different mechanisms

Jo-Anne Bright[1,2,*], Kate E. Stevenson[1], James M. Curran[2], John S. Buckleton[1]

[1]*ESR Ltd, Private Bag 92021 Auckland, New Zealand 1142*

[2] *University of Auckland Department of Statistics, Private Bag 92019 Auckland, New Zealand 1142*

* Corresponding author at: Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142, New Zealand. Email address: Jo.bright@esr.cri.nz .

Recently there has been a drive towards standardisation of forensic DNA interpretation methods resulting in the uptake of probabilistic interpretation software. Some of these software solutions utilise Markov chain Monte Carlo techniques (MCMC). They will not produce an identical answer after repeat interpretations of the same evidence profile because of the Monte Carlo aspect. This is a new source of variability within the forensic DNA analysis process. In this paper we explore the size of the MCMC variability within the interpretation software STRmix™ compared to other sources of variability in forensic DNA profiling including PCR, capillary electrophoresis load and injection, and the makeup of allele frequency databases. The MCMC variability within STRmix™ was shown to be the smallest source of variability in this process.

Keywords: Forensic DNA; Interpretation; Likelihood ratio; MCMC; STRmix™

Introduction

In forensic DNA analysis, a profile is produced from a biological sample and compared with the DNA profile of one or more persons of interest (POI). Stages in a typical DNA analysis include sampling, extraction of the DNA, amplification using PCR, separation of the amplified components via capillary electrophoresis (CE) and production of an electropherogram (epg) [1]. An epg appears as a set of peaks in a plot of fluorescence versus time on the capillary. The height of the peaks is relative to the amount of amplified DNA. The time scale is converted to molecular weight using internal size standards.

The amplification process will amplify the alleles present but also produces by-products, the predominant one being a peak one repeat shorter than the allele, termed back stutter [2]. After production of an epg the next step is most often a human interpretation of the profile. At this stage artefactual peaks may be removed. Certain artefacts are readily recognised by a human and may be discounted as being of allelic origin. However, at least two artefacts, back and forward stutter, are very difficult to differentiate from true allelic peaks. This is particularly difficult in profiles containing DNA from more than one contributor (termed a mixture) where the peaks of a minor contributor are of a similar height to stutter peaks from a major contributor.

In 2011, Dror and Hampikian reported an exercise undertaken studying the variability in DNA profile interpretation [3]. In this article a mixed DNA profile from the sperm fraction of a vaginal sample, where the person of interest (POI) had previously been reported as not excluded, was distributed to 17 individuals from within the same laboratory. They were asked to interpret the profile and offer one of three conclusions: cannot be excluded, excluded or inconclusive. Only one analyst agreed with the original conclusion. Four analysts said the profile was inconclusive and 12 concluded the POI was excluded. The results of this study were reported in New Scientist under the title "Fallible DNA evidence can mean prison or freedom" [4] and have since been used to highlight the subjectivity and bias inherent to certain methods of DNA interpretation.

A 2005 study [5] showed that it can also be difficult to determine how many people have contributed to a mixture. The authors showed that more than 70% of four person mixtures could be incorrectly designated as two or three person mixtures. This work has more recently been repeated for the European Standard Set of loci [6] where the authors demonstrated that an increase in the number of loci within a multiplex and the resulting increases in discriminatory power reduces this rate but it will never remove it.

The response to this, and other highly justifiable criticisms, is to promote best practise and encourage standardisation [7-9]. Within Australasia, a standardisation process resulted in the implementation of the continuous DNA profile interpretation software STRmix™ [10-12]. An intended outcome of standardisation processes is to reduce the variability within each region. The implementation of the same software across different laboratories within Europe has been shown to reduce but not eliminate the variability, however [13]. Standardisation is not a panacea but it is at least a step in the right direction. Standardisation efforts represent a focus on repeatability or, in quality parlance, precision.

Efforts to remove (or decrease) between or within laboratory variability tend to focus on the symptoms rather than the causes. It is certainly conceivable that the result could be highly repeatable but patently false. For example, taken to the extreme, we could imagine laboratories issuing a report for every positive correspondence with LR = 1 billion. This result would be highly repeatable. The focus on repeatability as opposed to scientific logic and rigor has detrimental consequences.

Each of the steps in the analytical process could be repeated and if repeated then the resulting LR would be different. The sample itself is just one of potentially many samples that could have been collected from a scene or from a larger exhibit. The very concept that there is one and only one answer is flawed and in fact highly detrimental.

STRmix™ utilises Markov chain Monte Carlo (MCMC) methods. STRmix™ will not produce an identical answer after repeat interpretations of the same epg because of the Monte Carlo aspect. This is a new source of variability within the forensic DNA analysis process and may be perceived as a disadvantage.

In this paper we explore the variability that would occur if aspects of this process from the PCR stage were repeated.

Method

DNA from saliva collected from two individuals was extracted using Promega DNA IQ™ extraction chemistry (Madison, WI). The extracted DNA was quantitated using the Applied Biosystems Quantifiler™ real time PCR kit (CA, Life Technologies). An equivalent amount of DNA from both individuals was aliquoted into one sample. Ten replicate amplifications of 0.1 ng total DNA from the aliquoted sample were performed using the Applied Biosystems Identifiler™ multiplex. The replicate amplifications were set up by hand using calibrated micro pipettors. All amplifications were undertaken concurrently in an Applied Biosystems 9700 thermal cycler with a silver block. Low level mixed DNA profiles where the contributors are in roughly equal proportions are likely to have the greatest MCMC variability and were selected for this reason. They will also have discernible PCR to PCR variability. As such, they represent a situation where the pressure on reproducibility is high.

Aliquots of 1 μL of amplified DNA from each of the ten replicates were separated concurrently on an Applied Biosytems 3130xl capillary electrophoresis (CE) instrument. Subsequently, one of the replicates was loaded onto one plate in ten different wells and analysed. Both plates were set up manually, again using calibrated micro pipettors. In a final experiment, one replicate was repeat injected onto the CE instrument ten times from the same set up plate.

Each mixed DNA profile was compared to the reference profile from one of the known contributors (referred to as Male A). A likelihood ratio (*LR*) was calculated using STRmix™ which employs a continuous method of DNA profile interpretation [10-12] considering the following two hypotheses:

H_p: Male A and one unknown contributor

H_d: Two unknown contributors

All statistical calculations were performed using the same Caucasian population allele frequencies (not previously published) and a θ value of 0.01 was applied using the sub population correction model of Balding and Nichols [14]. No correction for allele frequency sampling effects was undertaken. Following laboratory standard operating procedure, STRmix™ was run for a total of 400,000 accepts following a period of 100,000 burn-in accepts.

In order to investigate the variability of the MCMC process within STRmix™ the *LR* was calculated in triplicate for each of the ten replicate amplifications, the ten repeat CE loads and the ten multiple injections.

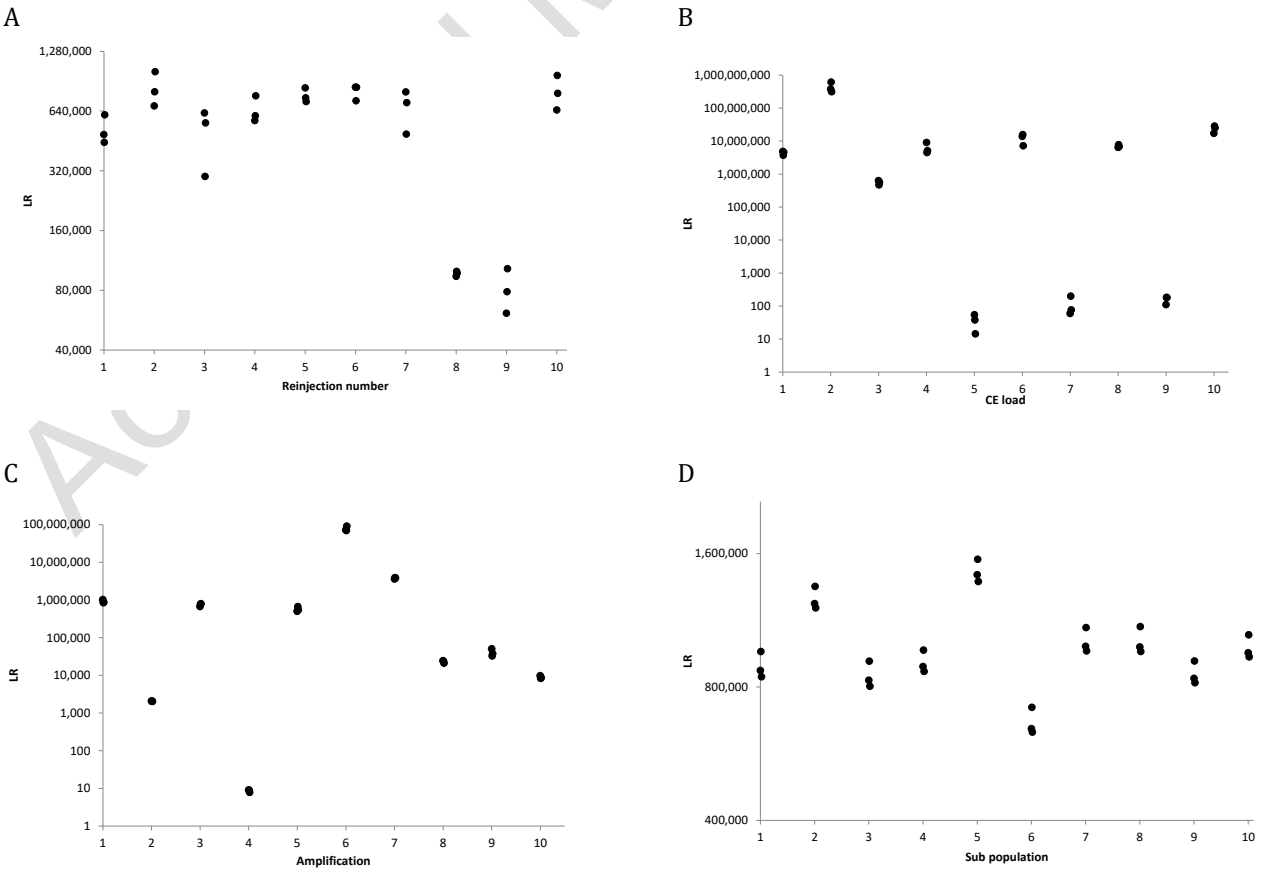
The allele frequencies are compiled from 8248 self-declared Caucasians from the New Zealand population. In order to investigate variability between allele frequency databases of the same ethnic group, the population database was subsequently randomly divided into ten separate populations each of 824 to 832 individuals. Allele frequencies for each of these subpopulations were determined and the *LR* for one sample calculated in STRmix™ for each of the subpopulations as described above.

Results

Figure 1A is the plot of *LR*s for one amplified sample after repeat injection on the CE. The variability of the MCMC process within STRmix™ can be observed by comparing the *LR within* each reinjection number. The variability of the combined CE injection and MCMC processes can be observed by comparing the *LR between* each reinjection. Figure 1B is the plot of the *LR*s for a different amplified sample after repeat manual loads to a CE plate. As in Figure 1A the variability of the MCMC process can be observed by comparing the *LR within* each CE load number and the combined CE load, injection and MCMC process variability can be observed by comparing the *LR between* each CE load number. Inspection of the peaks heights of the internal size standard for each of the reinjection and CE load profiles indicated no significant difference (data not shown). The difference in the observed *LR* between the ten reinjections of the same sample could be accounted for by stochastic variation in eletrokinetic injection from the low template sample DNA.

Figure 1C is the plot of the *LR* for each of the ten replicate amplifications. The variability across amplifications is an accumulation of amplification, CE load and injection variabilities. Finally, Figure 1D is the plot of *LR*s for one amplified sample after recalculation using ten different allele frequency files prepared by randomly splitting the one larger population used previously for Figures 1A through C.

Figure 1: Plot of *LR*s calculated in triplicate after replicate processes: repeat CE injection of the same amplified DNA (panel A), manual CE load of the same amplified DNA (panel B), repeat amplification of the same DNA extract (panel C) and calculation of the *LR* using a different 1/10th portion of a large allele frequency database.



Conclusion

Inspection of Figures 1A through 1D confirms that a repeat of any part of the process from PCR to *LR* calculation can lead to a different result. This result may differ in a small way or a larger way and overall, the MCMC process displayed the least variability. The division of the population into ten smaller subpopulations had the next smallest effect followed by CE injection. Aside from the two outliers (reinjection 8 and 9) the variability due to CE injection was similar to that observed to the MCMC variability. CE load and PCR had the greatest variability. The variability accumulates for each additional step in the process of generating a DNA profile and calculating an *LR*. The relative amount of variability for each process can be determined by comparing the variability between different plots in Figure 1.

For the situations examined the MCMC variance was a small fraction of the total potential variability. The magnitude of the MCMC variation depends mainly on two factors; the size of the weights for genotype sets in H_p and the length the MCMC analysis is run [15, 16]. In validation trials and casework we have observed an increased MCMC variability when the weights for genotype sets under H_p are small. Weights can be small for low level contributors to profiles and where many possible genotype combinations are being considered for the POI, for example in mixed DNA profiles. Generally, for low level contributors the weights decrease as the number of contributors to a profile increase. Increasing the number of MCMC accepts during STRmix™ interpretation of a profile improves the run to run precision. Our laboratory default is 400,000 post burn-in accepts. This is likely to be excessive for many 'simple' single source and two person mixed DNA profiles as in this trial.

We would expect that the variability in the sampling and extraction process would be even greater. This is most easily glimpsed at sampling where different stains may be examined and collected at the scene. These samples would undoubtedly be of varying quality and quantity of DNA, may be mixed and contain DNA from different individuals in varying proportions.

The *LR* itself is conditional on modelling assumptions. In the case of STRmix™ this includes the behaviour of the peak and stutter heights [10, 17] and a population genetic model [14]. However well-justified these models may be it would be incorrect to say that any one of them was the 'true' model. Taylor et al. previously investigated the effect of different sources of uncertainty on likelihood ratio calculations including allele frequencies, θ and population makeup [16]. They concluded that all sources of uncertainty will interact to produce an *LR* distribution that is wider than if the individual sources of variation were considered.

Acknowledgements

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. We would like to thank Johanna Veth, Heidi Baker and two anonymous referees who helpful comments greatly improved this paper.

References

1. Butler, J.M., Advanced topics in forensic DNA typing: methodology. 2012, San Diego: Academic Press.

2. Walsh, P.S., N.J. Fildes, and R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res., (1996). 24: 2807-2812.
3. Dror, I.E. and G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation. Science & Justice, (2011). 51(4): 204-208.
4. Geddes, L., Fallible DNA evidence can mean prison or freedom. New Scientist, (2010). 207(2773): 8-11.
5. Paoletti, D.R., et al., Empirical analysis of the STR profiles resulting from conceptual mixtures. Journal of Forensic Sciences . (2005). 50: 1361-1366.
6. Curran, J.M. and J. Buckleton, Uncertainty in the number of contributors for the European Standard Set of loci. Forensic Science International: Genetics, (2014). 11(1): 205-206.
7. Scientific Working Group on DNA Analysis Methods (SWGDAM). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2010 [cited; Available from: http://www.fbi.gov/hq/lab/html/codis_swgdam.pdf.
8. Morling, N., et al., Interpretation of DNA mixtures--European consensus on principles. Forensic Science International: Genetics, (2007). 1(3-4): 291-292.
9. Haelser, A., Issues in gathering, interpreting and delivering DNA evidence, in Expert Evidence Conference. 2011: Canberra.
10. Bright, J.-A., et al., Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics, (2013). 7(2): 296-304.
11. Bright, J.-A., et al., Degradation of forensic DNA profiles. Australian Journal of Forensic Sciences, (2013). 45(4): 445-449.
12. Taylor, D., J.-A. Bright, and J. Buckleton, The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics, (2013). 7(5): 516-528.
13. Prieto, L., et al., Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. Forensic Science International: Genetics, (2014). 9: 47-54.
14. Balding, D.J. and R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International, (1994). 64: 125-140.
15. Taylor, D., Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. Forensic Science International: Genetics, (2014). 11: 144-153.
16. Taylor, D., et al., An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics, (2014). 11: 56-63.
17. Brookes, C., et al., Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics, (2012). 6(1): 58-63.