forensic sciences

MDPI

Article

Comparison of Likelihood Ratios from Probabilistic Genotyping for Two-Person Mixtures across Different Assays and Instruments

Dennis McNevin [1,*] and Mark Barash [2]

[1] Centre for Forensic Science, School of Mathematical & Physical Sciences, Faculty of Science, University of Technology Sydney, Sydney, NSW P.O. Box 123, Australia
[2] Department of Justice Studies, San Jose State University, San Jose, CA 95192, USA; mark.barash@sjsu.edu
* Correspondence: dennis.mcnevin@uts.edu.au; Tel.: +61-2-9514-1717

Abstract: Continuous probabilistic genotyping (PG) provides a means to estimate the probative value of DNA mixtures tendered as evidence in court and subject to alternative propositions about the contributors to the mixtures. The weight of that evidence, however, may be valued differently, depending on which forensic laboratory undertook the DNA analysis. There is a need, therefore, to have a means for the comparison of likelihood ratios (LRs) generated by continuous PG amongst different laboratories for the same initial DNA sample. Such a comparison would enable the courts and the public to make judgements about the reliability of this type of evidence. There are particular mixtures and methods for which such a comparison is meaningful, and this study explores them for the short tandem repeat (STR) electropherograms of two-person mixtures obtained from the PROVEDIt Database. We demonstrate a common maximum attainable LR for a given set of common STR loci and a given DNA mixture that is consistent across three different STR profiling assays and two different capillary electrophoresis instruments.

Keywords: DNA mixture; probabilistic genotyping; short tandem repeat; PROVEDIt; comparison

Citation: McNevin, D.; Barash, M. Comparison of Likelihood Ratios from Probabilistic Genotyping for Two-Person Mixtures across Different Assays and Instruments. *Forensic Sci.* **2024**, *4*, 441–452. https://doi.org/10.3390/forensicsci4030028

Academic Editors: Sara C. Zapico and Hiroshi Ikegaya

Received: 6 June 2024
Revised: 30 July 2024
Accepted: 27 August 2024
Published: 2 September 2024

1. Introduction

Continuous probabilistic genotyping (PG) is increasingly used to estimate the weight of evidence for short tandem repeat (STR) profiles with more than one contributor. The output from PG is a ratio of the likelihoods for two alternative propositions concerning the mixture, called the likelihood ratio (LR). The value of the LR for a DNA mixture will depend upon a number of factors including (but not limited to) the following: the propositions, the proportion of DNA from each contributor to the mixture, the quality and quantity of DNA template, the profiling assay, the instrumentation used for DNA profiling, the allele frequencies employed for estimating genotype frequencies and any corrections for population structure. For these reasons, it has been argued that inter-laboratory comparisons of LRs generated from common DNA mixtures are too difficult because of the "human factors, laboratory policy, and elements outside the province of the software" [1], and "conflating too many variables" [2]. This reasoning is problematic as it contradicts the fundamental scientific requirements of reproducibility and consistency in results among multiple laboratories, as highlighted in the President's Council of Advisors on Science and Technology (PCAST) report [3] and the National Institute of Standards and Technology (NIST) Scientific Foundation Review [4]. At the very least, a PG system "that returns reproducible LRs from different laboratory specific parameters might be considered trustworthy" [5]. Bright et al. [6] described a collaborative exercise in which two mixed GlobalFiler™ DNA electropherograms (epgs, from the PROVEDIt Database, https://lftdi.camden.rutgers.edu/provedit/files/ [7]) were submitted to participating laboratories. The LRs returned by the laboratories for one epg clustered into

two distinct groups, attributed to differences in the assigned epg peak heights which were in turn a result of the epg signal processing settings (GeneMapper® ID-X, Thermo Fisher Scientific, Waltham, MA, USA). It is therefore essential that any interlaboratory comparison includes this signal processing as part of the LR pipeline. Boodoosingh, Kelly, Curran, and Kalafut [5] have recently compared the LRs generated by one PG system (STRmix™, ESR, Auckland, NZ, USA) from the raw epgs of different laboratories but it is unclear what signal processing was applied. They maintain that "the true value of an LR cannot be known, making accuracy difficult to measure".

McNevin et al. [8,9] have proposed a framework for interlaboratory comparisons of the LRs generated by continuous PG from the same DNA mixtures that identifies a maximum attainable LR. This might be as close to a "true" value as is possible. Their framework accounts for most of the differences amongst individual DNA profiling pipelines by minimizing their influence. This is possible under specific conditions, namely the following:

- Each laboratory should examine aliquots of a dilution series of the same mixture consisting of equal proportions of high abundance DNA from each contributor to the mixture.
- Each laboratory applies their own DNA profiling pipeline to each aliquot of the dilution series.
- Each laboratory uses their own continuous PG system to generate an LR for each aliquot of the dilution series according to the following:

$$\mathrm{LR} = \frac{\mathrm{P}(G|H_1)}{\mathrm{P}(G|H_2)}$$

where the propositions H_1 and H_2 are as follows:

H_1: The donor of a given reference profile X is a contributor to the mixture, which also consists of N other contributors, which are unrelated to each other, and to the donor of profile X.

H_2: The donor of reference profile X is not a contributor to the mixture, which consists of $N + 1$ other contributors, which are unrelated to each other, and to the donor of profile X.

- The LR should be calculated using:
  - only those loci in common amongst the participating laboratories;
  - the same population allele frequencies;
  - the same population genetic model for calculating genotype frequencies from allele frequencies (e.g., Hardy–Weinberg proportions);
  - the same population sub-structure correction (e.g., $\theta = 0$).

Under these conditions, the LR should plateau at the same value for higher concentrations of DNA (that is, less diluted aliquots), regardless of the laboratory. This is because the ambiguities associated with allele designation have been minimized such that the likelihoods for true genotype sets are maximized.

The simplest mixture is a two-person mixture for which the simplest propositions are as follows:

H_1: The donor of a given reference profile X is a contributor to the mixture which also consists of one other contributor that is unrelated to the donor of profile X.

H_2: The donor of reference profile X is not a contributor to the mixture which consists of two other contributors, which are unrelated to each other, and to the donor of profile X.

Here, we demonstrate that the approach of McNevin et al. [8,9] for interlaboratory comparison of continuous PG systems is appropriate for two-person mixtures and does indeed lead to reproducible LRs for different combinations of STR genotyping assays and capillary electrophoresis (CE) instruments.

2. Materials and Methods

Electropherograms (epgs) for two-person mixtures were downloaded from the PROVEDIt Database [7] (https://lftdi.camden.rutgers.edu/provedit/files/, accessed on 4 March 2024). This database has been used in a number of similar studies, e.g., [10,11]. Only epgs that satisfied the following criteria were retained:

- Contributors consisted of Sample IDs 44 and 45 in the ratio 1:1.
- The DNA was pristine or "untreated" (i.e., no DNase degradation, Fragmentase® degradation, UV damage, sonication or humic acid inhibition).
- STR profiling assays included AmpFLSTR™ Identifiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific: TFS, Waltham, MA, USA) [12], GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA, USA) [13] or PowerPlex® Fusion 6C System (Promega Corporation, Madison, WI, USA) [14].

The final list of 36 epgs and the conditions under which they were generated is available in Supplementary Table S1. The range of conditions and numbers of epgs for each is summarized in the columns of Table 1.

Table 1. Summary of the conditions under which the PROVEDIt epgs were generated for this study (numbers in brackets refer to the numbers of epgs). Full details are provided in Supplementary Table S1.

STR Profiling Assays	DNA Amounts (ng)	Genetic Analyzers	CE Injection Times (s)	CE Size Standards
Identifiler™ Plus (12)	0.03125 (9)	3130 (12)	5 (12)	GeneScan™ 600 LIZ™ (24)
GlobalFiler™ (12)	0.0625 (9)	3500 (24)	10 (4)	WEN ILS 500 (12)
Fusion 6C (12)	0.125 (9)		15 (8)	
	0.25 (9)		20 (4)	
			25 (8)	

The 36 epgs were imported into the Open Source Independent Review and Interpretation System (OSIRIS, Version 2.16, https://www.ncbi.nlm.nih.gov/osiris/) [15] as part of their plate folders, together with plate-specific ladders. Default analysis settings were applied except for minimum RFU, which were 50 for analysis and detection and 150 for the ladder and the internal lane standards (ILS). Three ILS were employed—ABI-LIZ-600-80 to 400 for Identifiler™ Plus, ABI-LIZ-600-60 to 460 for GlobalFiler™, and Promega-ILS-WEN-500 for Fusion 6C (Supplementary Table S1). Example electropherograms for 0.25 ng DNA template amounts and 5 s CE injection times for Identifiler™ Plus, GlobalFiler™, and Fusion 6C are shown in Supplementary Figures S1, S2 and S3, respectively. Tab-delimited analysis data were exported from OSIRIS and saved as Excel spreadsheets.

EuroForMix (Version 4.0.8, http://www.euroformix.com/, accessed on 4 March 2024) [16] was the continuous PG system used for calculating LRs. Allele frequencies for a population with European ancestry [17] were used to create a population frequency CSV file for the 15 loci shared by Identifiler™ Plus, GlobalFiler™, and Fusion 6C (CSF1PO, FGA, TH01, TPOX, VWA, D13S317, D16S539, D18S51, D19S433, D21S11, D2S1338, D3S1358, D5S818, D7S820, and D8S1179). CSV evidence input files for the same loci were created using the exported OSIRIS data for the 36 epgs in Supplementary Table S1. CSV reference files were created using the PROVEDIt known genotypes for Sample IDs 1 (as a known non-contributor to the mixtures), 44, and 45 (as known contributors). Degradation, backward stutter, and forward stutter were permitted as model options. For each epg, a quantitative LR (maximum likelihood based) was calculated for the three proposition pairs described in Table 2.

Table 2. Proposition pairs applied to each PROVEDIt epg used in this study in order to calculate LRs in EuroForMix (Version 4.0.8).

Proposition Pairs	Contributors under H_1	Contributors under H_2
1 (false)	Sample ID 1 + 1 unknown	2 unknowns unrelated to Sample ID 1
2 (true)	Sample ID 44 + 1 unknown	2 unknowns unrelated to Sample ID 44
3 (true)	Sample ID 45 + 1 unknown	2 unknowns unrelated to Sample ID 45

3. Results

The LRs calculated for the three proposition pairs are plotted in Figure 1 as functions of the DNA template input amounts. The $\log_{10}$LR generally decreased below zero with increasing template amount for proposition pair 1 (false propositions) and generally increased above zero with template amount for proposition pairs 2 and 3 (true propositions). There is evidence of a plateau for proposition pair 3 at $\log_{10}\text{LR} \approx 14$. For proposition pair 1 at a given template amount, $\log_{10}$LR is generally lower for longer CE injection times. This is because a longer injection time results in more DNA being delivered to the capillary, resulting in a stronger signal. We define a quantity we call CE mass (ng·s), which is the product of the DNA template amount (ng) and the CE injection time (s) and is indicative of the amount of DNA delivered to the capillary which, in turn, should be indicative of the amount of information available for determining likelihoods. The LRs calculated for the three proposition pairs are plotted in Figure 2 as functions of the CE mass.

For proposition pair 1 (false propositions), the LRs were always less than one ($\log_{10}$LRs always less than zero). For proposition pairs 2 and 3 (true propositions), LRs were always greater than one ($\log_{10}$LRs always greater than zero) with a sole exception. For one epg (Fusion 6C, 0.0625 ng DNA, 5 s CE injection), proposition pairs 2 and 3 returned $\log_{10}$LRs of -9.301 and -8.265, respectively. This epg was derived from the same PCR and the same 3500 Genetic Analyzer as two other epgs (Fusion 6C, 0.0625 ng DNA, 15 s CE injection; Fusion 6C, 0.0625 ng DNA, 25 s CE injection), and the genotypes for all three epgs are compared in Table 3. In general, for these epgs, the number of extraneous alleles (those that are not bold in Table 3) increased as the CE injection time was reduced, especially for D8S1179, shown in Figure 3, which looks to be the result of an elevated baseline or "waterfall", perhaps caused by poor spectral calibration or poor-quality formamide. Extraneous alleles increase the likelihood for H_2 (unknown contributors). When D8S1179 was removed from the analyses for the outlier epg (Fusion 6C, 0.0625 ng DNA, 5 s CE injection), the $\log_{10}$LRs for proposition pairs 2 and 3 changed to 6.95 and 4.636, respectively.

The plateaus for $\log_{10}$LR occur for CE mass > 3.2 for proposition pair 2 and for CE mass > 1.6 for proposition pair 3 (Figure 2). At lower CE masses, the $\log_{10}$LR increases with CE mass towards the plateau, but there are a number of outliers and always for GlobalFiler™ and PowerPlex® Fusion 6C. Examination of the epgs for these outliers demonstrates that the lower-than-expected LRs are attributable to unexpected allele dropout. In particular, the dropout of allele 12 at D5S818 (from Sample ID 44) occurred in all three epgs derived from GlobalFiler™ and PowerPlex® Fusion 6C amplification of 0.125 ng DNA but not for Identifiler™ Plus (Figure 4).

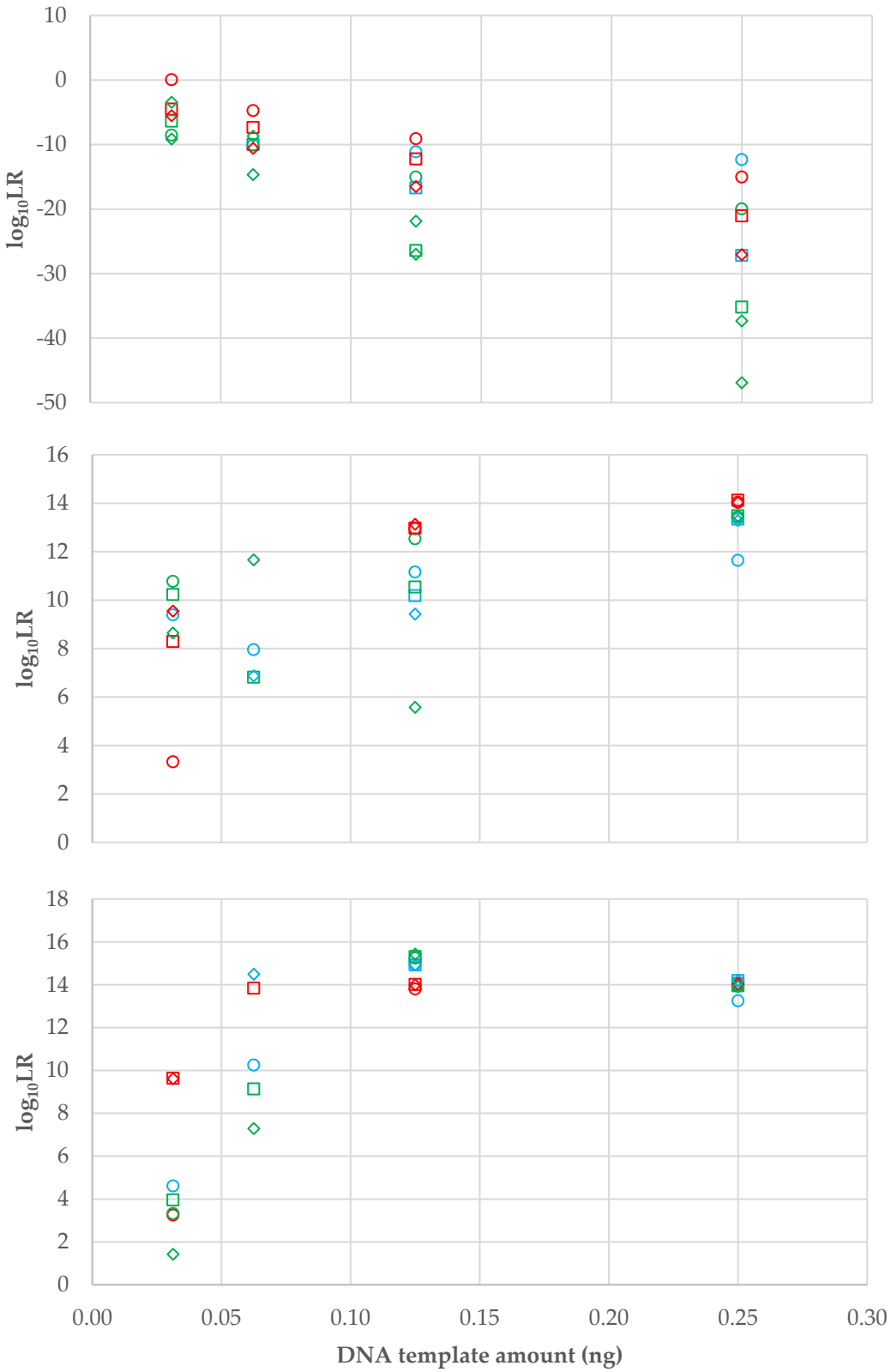

Figure 1. LRs calculated for proposition pairs 1 (**top**), 2 (**middle**), and 3 (**bottom**). STR profiling assays are Identifiler™ Plus, GlobalFiler™ and Fusion 6C. CE injection times are 5 s (○), 10–15 s (□) and 20–25 s (◊).

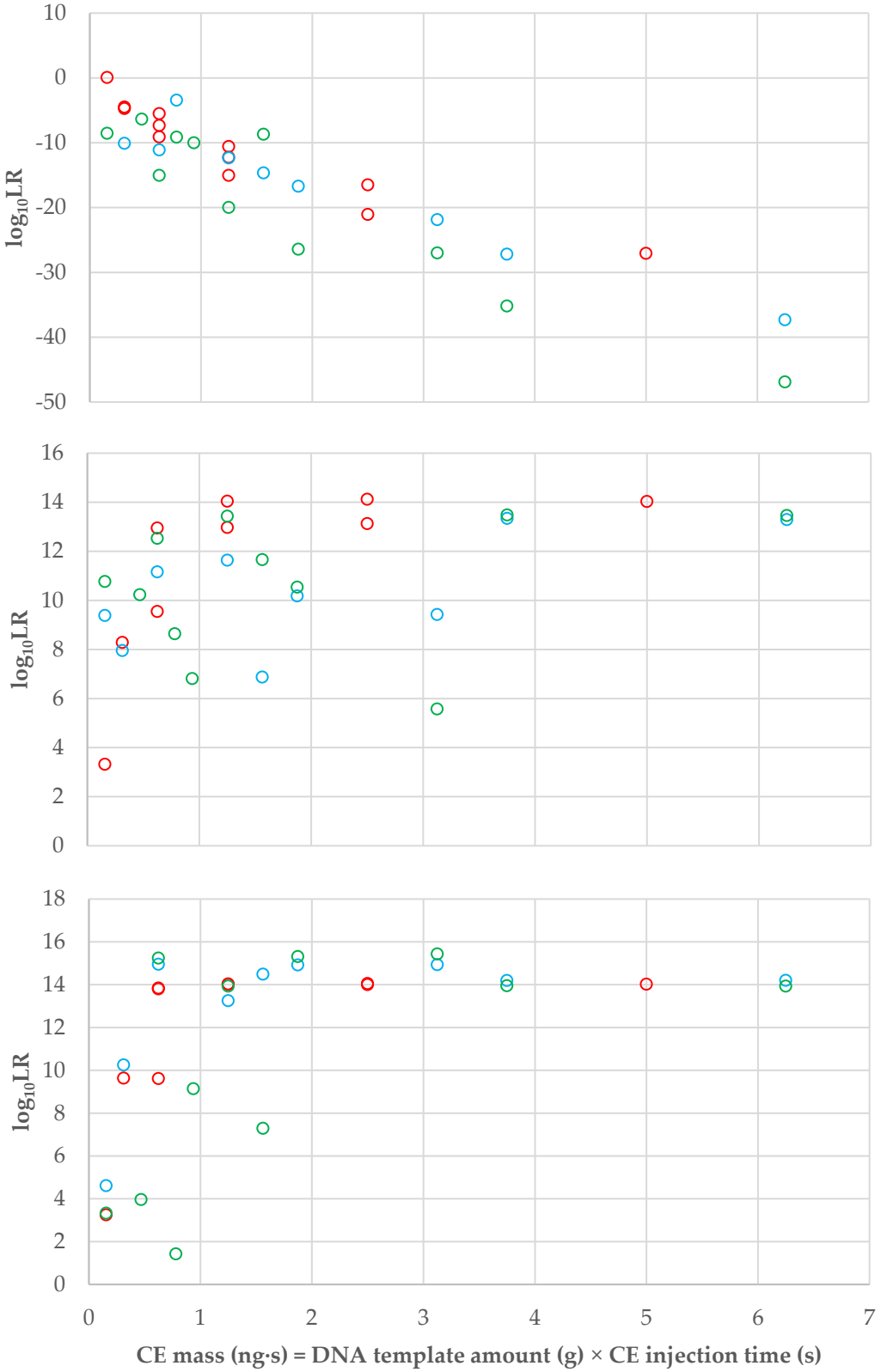

Figure 2. LRs calculated for proposition pairs 1 (**top**), 2 (**middle**), and 3 (**bottom**). STR profiling assays are Identifiler™ Plus, GlobalFiler™ and Fusion 6C.

Table 3. STR genotypes for the three CE injections of 0.0625 ng DNA template amounts amplified by PowerPlex® Fusion 6C as well as reference profiles for Sample IDs 44 and 45. Alleles in **bold** are shared with the reference profiles.

Locus	CE Injection Time			References	
	5	**15**	**25**	**Sample ID 44**	**Sample ID 45**
Amelogenin	**X**, **Y**	**X**, **Y**	**X**, **Y**	**X**, **Y**	**X**, **Y**
CSF1PO	**11**, **12**	**11**, **12**	**11**, **12**	**11**, **12**	**11**, **12**
FGA	20, **21**, **23**, **24**, **25**	**21**, **23**, **24**	**21**, **23**, **24**	**23**, **24**	**21**, **25**
TH01	5.2, 6.3, **9**, **9.3**	3, 4, 5, **6**, 6.3, **7**, **9**, **9.3**, 10	**6**, **7**, **9**, **9.3**	**6**, **9.3**	**7**, **9**
TPOX	**8**, **11**	**8**, **11**	**8**, **11**	**11**, **11**	**8**, **11**
vWA	**13**, **20**	**13**, **20**	**13**, **20**	**20**, **20**	**13**, **16**
D2S1338	**16**, **19**, **25**	**19**, **25**	**19**	**16**, **19**	**19**, **25**
D3S1358	14, **15**, **16**, **17**	**15**, **16**, **17**	**15**, **16**, **17**	**15**, **16**	**15**, **17**
D5S818	**11**, **12**	**11**, **12**	**11**, **12**	**11**, **12**	**11**, **11**
D7S820	**8**, **9**, **10**	**8**, **9**, **10**	**8**, **9**, **10**	**8**, **10**	**9**, **10**
D8S1179	6, 6.2, 6.3, 7, 7.1, 7.2, 7.3, 8, 8.2, 8.3, 9.1, 9.3, **10**, 10.1, 10.3, 11.1, 11.2, 12, 12.2, 13, 13.2, **14**, **15**	6.1, 6.3, 7.3, 8.3, 9, **10**, 11.3, 12.2, 13, **14**, **15**	**10**, **14**, **15**	**14**, **15**	**10**, **14**
D13S317	**8**, **13**	**8**, **12**, **13**	**8**, **12**	**8**, **8**	**12**, **13**
D16S539	4.1, 9, **10**, **11**	4.1, **10**, **11**	**10**, **11**	**10**, **11**	**10**, **11**
D18S51	**12**, **14**, 15, **16**	**12**, **14**, **16**	**12**, **14**, **16**	**12**, **14**	**16**, **17**
D19S433	**11**, **13**, **14**, 14.1, 14.2, **15**	**11**, **13**, **14**, **15**	**11**, **13**, **14**, **15**	**13**, **15**	**11**, **14**
D21S11	**28**, **31**, **32**	**28**, **31**, **32**	**28**, **31**, **32**	**28**, **32**	**28**, **31**

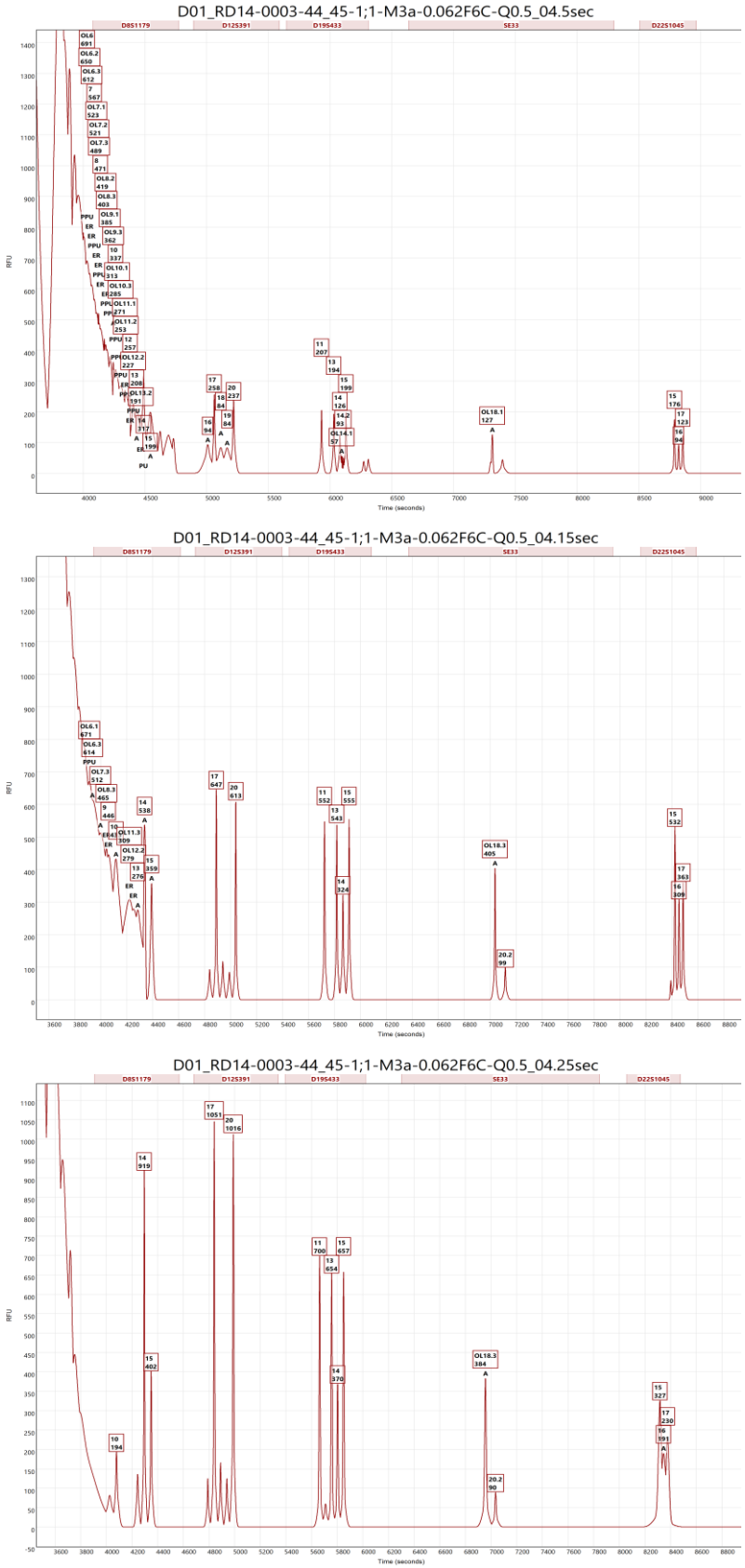

Figure 3. Electropherograms (CXR dye) for the three CE injections of 0.0625 ng DNA template amounts amplified by PowerPlex® Fusion 6C System. CE injection times are 5 s (**top**), 15 s (**middle**), and 25 s (**bottom**). D8S1179 is the first locus at the left in all three epgs.

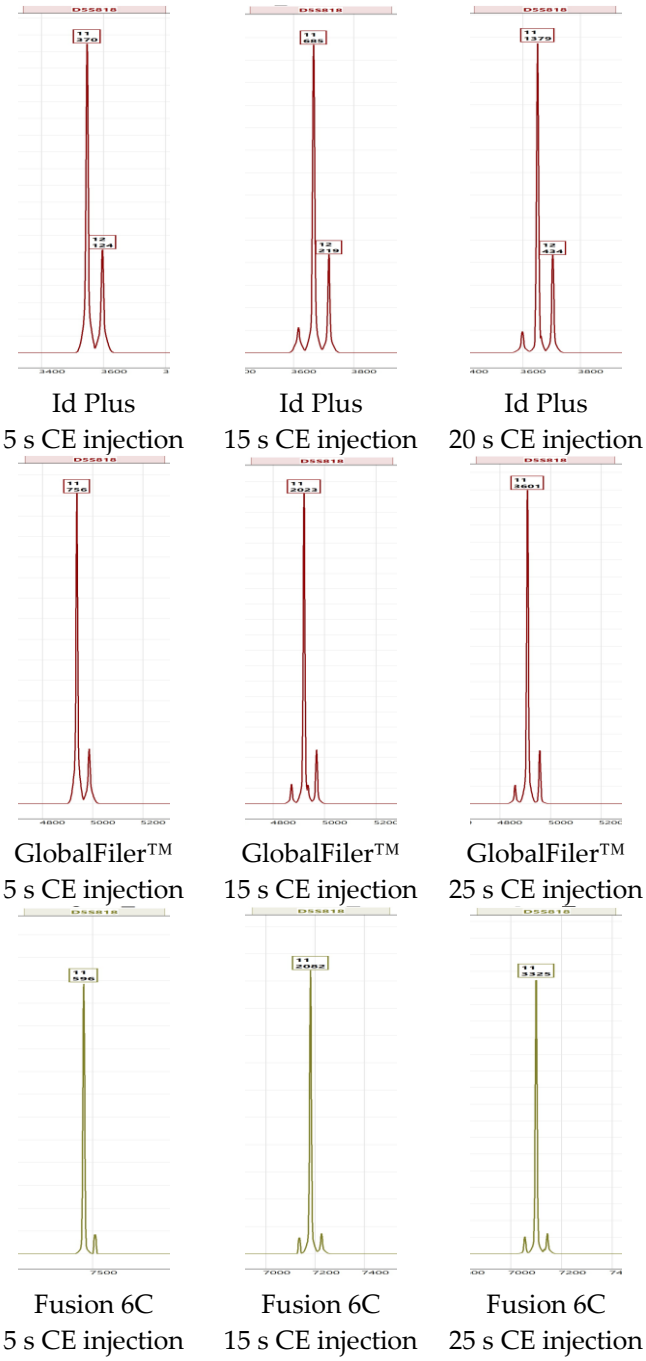

Figure 4. Detail from epgs at D5S818 derived from amplification of 0.125 ng DNA with various CE injection times showing dropout of allele 12 for GlobalFiler™ and PowerPlex® Fusion 6C but not for Identifiler™ Plus.

4. Discussion

LRs have been generated using the EuroForMix continuous PG system on epgs derived from three different STR profiling assays applied to four different DNA template amounts which, in turn, were subjected to three different CE injection times (Supplementary Table S1). LRs could not be generated for all the proposition pairs (Table 2) and all 36 epgs. For some combinations of the proposition pairs and epgs, "the specified model could not explain the data". These are labelled "unexplained" in Supplementary Table S1 (five epgs: one for proposition pair 1, four for proposition pair 2, and three for proposition pair 3). For three epgs, both proposition pairs 2 and 3 were unexplained. For these five epgs, the quantitative LR (maximum likelihood based) model may have tried to fit degradation and/or stutter

when one or both are not present and, while "the solution is to turn off the corresponding model option which caused the problem" [18], degradation, backward stutter and forward stutter were successfully modeled in all the other epgs and so these options were retained for consistency.

Elevated baselines at the shortest CE elution times were a consistent feature of the epgs in this dataset that were generated on a 3500 Genetic Analyzer (Thermo Fisher Scientific, Waltham MA, USA) (but not on a 3130 Genetic Analyzer), including negative controls, perhaps because of the higher sensitivity of the instrument and/or the signal processing applied by OSIRIS. Many potential artefacts, like the ones at D8S1179 for Fusion 6C amplification of 0.0625 ng DNA (Table 3 and Figure 3), would be eliminated by using a higher minimum RFU threshold, but misleading LRs were only returned for one epg. A lower minimum RFU (50) was employed to allow as many artefacts as possible which would properly test the continuous PG algorithm (EuroForMix) and our proposed strategy for interlaboratory comparisons, as well as improve the sensitivity for the detection of mixtures.

Dropout occurred for allele 12 at D5S818 (from Sample ID 44) for all three epgs derived from GlobalFiler™ and PowerPlex® Fusion 6C amplification of 0.125 ng DNA (Figure 4). This had the effect of yielding LRs for these STR kits below those obtained for Identifiler™ Plus. The genotypes for the two contributors (Sample IDs 44 and 45) at this locus are 11, 12, and 11, 11; thus, we would expect the peak height for allele 12 to be approximately one-third of the combined peak height for allele 11, but it seems to disappear below the analytical threshold for GlobalFiler™ and PowerPlex® Fusion 6C. This could be a PCR 'stochastic effect' artefact where a small imbalance in the initial PCR amplifications in favor of allele 11 (the major peak) is enough to render allele 12 (the minor peak) undetectable. It may also be possible that the smoothing algorithm for OSIRIS obscures low peaks with small signal-to-noise ratios. In this way, we can quickly identify potential shortcomings for particular combinations of probabilistic genotyping and STR profiling pipelines in particular laboratories. This is another valuable aspect of the inter-laboratory comparison framework proposed by McNevin et al. [8,9].

Figure 2 shows that for proposition pair 1, the $\log_{10}$LR decreases steadily with CE mass (the product of DNA template amount and CE injection time). The weight of evidence for this false proposition is therefore increased as the amount of fluorescently labeled DNA fragments delivered to the capillaries is increased. We can therefore increase support for exclusion of true non-contributors by either increasing the DNA template amount for PCR and/or the CE injection time. This result is in agreement with the theoretical treatment by McNevin et al. [8], and the empirical evidence offered by Taylor [19], namely that LR $\longrightarrow 0$ ($\log_{10}\text{LR} \longrightarrow -\infty$) for false propositions as the probability of non-contributor alleles decreases with the increasing true contributor allele peak height.

It should be noted that there are differences between PG algorithms in this region (for false propositions of the type represented by proposition pair 1). Buckleton, et al. [20] found that EuroForMix reports higher LRs than STRmix™ for true non-contributors, primarily because EuroForMix uses maximum likelihood estimation to calculate the LR whereas STRmix™ uses Markov chain Monte Carlo (MCMC) simulation. For STRmix™, then, we might expect a steeper decline in $\log_{10}$LR with increasing CE mass for proposition pair 1 in Figure 2. In fact, some of the LRs for STRmix™ in this region might be zero ($\log_{10}\text{LR} = -\infty$).

Figure 2 also shows that for proposition pairs 2 and 3, the $\log_{10}$LR increases with CE mass until it plateaus at a constant value. For proposition pair 2, the plateau occurs at $\log_{10}$LR in the range of 13 to 14 and for proposition pair 3, at $\log_{10}\text{LR} \approx 14$. Once again, this is in agreement with McNevin et al. [8], namely that the LR plateaus when no further information is gained from the increased peak heights because the non-contributor alleles have vanishingly low probabilities and, for mixtures with contributors in equal proportion, the probability that the allele from any contributor is confused with epg artefacts is also small.

This study was limited to two-person mixtures (Sample IDs 44 and 45 as contributors to equal proportion mixtures in the PROVEDIt Database). The same methodology could be applied to three, four, or more contributors in equal proportion. It has demonstrated the existence of a common maximum attainable LR for a given set of common STR loci, regardless of STR profiling assay and capillary electrophoresis instrument. At least for the case of two-person mixtures, it appears that there is, in fact, a reproducible LR under the conditions specified here (high CE masses). Figure 1 in Riman, Iyer, and Vallone [10] and Figure 2 in Riman, Iyer, and Vallone [11] also suggest a maximum attainable LR that is the same for both STRmix™ and EuroForMix for two-, three- and four-person mixtures.

Boodoosingh, Kelly, Curran, and Kalafut [5] suggest that the LR may be reproducible (if not accurate) over a much broader range of conditions than the ones specified by us. They found that "the largest differences in the reported LR . . . comes from changes in the propositions that were used". Choosing the right propositions could therefore also be part of any inter-laboratory comparison, as suggested by McNevin et al. [8,9].

5. Conclusions

We have provided experimental support for the framework for the comparison of continuous PG systems amongst different laboratories proposed by McNevin et al. [8,9] for two-person mixtures. The only amendment we propose is the use of CE mass as an indicator of DNA concentration. This is because CE mass is a better indicator of DNA mass delivered to capillaries, and this is reflected in the better resolution of the LR plateaus for proposition pairs 2 and 3 in Figure 2 compared with Figure 1. In fact, this suggests a strategy for improving the probative value of low template DNA evidence in general— increase the CE injection time [21]. Any changes to injection time, however, should be cautiously implemented within the confines of established laboratory standard operating procedures and after thorough validation of the methods to ensure reliability and reproducibility of results. This is because increased injection times may also result in more artefacts like stutter, allele drop in, and heterozygote imbalance. Moving forward, it is recommended that further research explores the applicability of this framework across a broader range of DNA mixture complexities and different PG systems. Such studies would help to standardize methodologies across laboratories, contributing significantly to the reliability of DNA evidence used in judicial settings.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/forensicsci4030028/s1: Supplementary Figure S1: Example electropherogram for Identifiler™ Plus with 0.25 ng DNA template amount and 5 s CE injection time. Supplementary Figure S2: Example electropherogram for GlobalFiler™ with 0.25 ng DNA template amount and 5 s CE injection time. Supplementary Figure S3: Example electropherogram for Fusion 6C with 0.25 ng DNA template amount and 5 s CE injection time. Supplementary Table S1: Electropherogram information and LRs.

Author Contributions: Conceptualization, D.M. and M.B.; methodology, D.M.; formal analysis, D.M.; investigation, D.M. and M.B.; data curation, D.M.; writing—original draft preparation, D.M.; writing—review and editing, M.B.; visualization, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The electropherograms used in this study were downloaded from the PROVEDIt Database, approved by the Institutional Review Board of the Boston University School of Medicine (Protocol Number H-31941).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gill, P.; Benschop, C.; Buckleton, J.; Bleka, Ø.; Taylor, D. A Review of Probabilistic Genotyping Systems: EuroForMix, DNAStatistX and STRmix™. *Genes* **2021**, *12*, 1559. [CrossRef] [PubMed]
2. Buckleton, J.S.; Bright, J.-A.; Ciecko, A.; Kruijver, M.; Mallinder, B.; Magee, A.; Malsom, S.; Moretti, T.; Weitz, S.; Bille, T.; et al. Response to: Commentary on: Bright et al. (2018) Internal validation of STRmix™—A multi laboratory response to PCAST, Forensic Science International: Genetics, 34: 11–24. *Forensic Sci. Int. Genet.* **2020**, *44*, 102198. [CrossRef] [PubMed]
3. Executive Office of the President President's Council of Advisors on Science and Technology. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*; President's Council of Advisors on Science and Technology: Washington, DC, USA, 2016.
4. Butler, J.M.; Iyer, H.; Press, R.; Taylor, M.K.; Vallone, P.M.; Willis, S. *DNA Mixture Interpretation: A NIST Scientific Foundation Review*; National Institute of Standards and Technology: Gaithersburg, ML, USA, 2021.
5. Boodoosingh, S.; Kelly, H.; Curran, J.M.; Kalafut, T. An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories. *Forensic Sci. Int. Genet.* **2024**, *71*, 103046. [CrossRef] [PubMed]
6. Bright, J.-A.; Cheng, K.; Kerr, Z.; McGovern, C.; Kelly, H.; Moretti, T.R.; Smith, M.A.; Bieber, F.R.; Budowle, B.; Coble, M.D.; et al. STRmix™ collaborative exercise on DNA mixture interpretation. *Forensic Sci. Int. Genet.* **2019**, *40*, 1–8. [CrossRef] [PubMed]
7. Alfonse, L.E.; Garrett, A.D.; Lun, D.S.; Duffy, K.R.; Grgicak, C.M. A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Sci. Int. Genet.* **2018**, *32*, 62–70. [CrossRef] [PubMed]
8. McNevin, D.; Wright, K.; Barash, M.; Gomes, S.; Jamieson, A.; Chaseling, J. Proposed Framework for Comparison of Continuous Probabilistic Genotyping Systems amongst Different Laboratories. *Forensic Sci.* **2021**, *1*, 33–45. [CrossRef]
9. McNevin, D.; Wright, K.; Chaseling, J.; Barash, M. Commentary on: Bright et al. (2018) Internal validation of STRmix™—A multi laboratory response to PCAST, Forensic Science International: Genetics, 34: 11–24. *Forensic Sci. Int. Genet.* **2019**, *41*, e14–e17. [CrossRef] [PubMed]
10. Riman, S.; Iyer, H.; Vallone, P.M. Exploring DNA interpretation software using the PROVEDIt dataset. *Forensic Sci. Int. Genet. Suppl. Ser.* **2019**, *7*, 724–726. [CrossRef]
11. Riman, S.; Iyer, H.; Vallone, P.M. Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset. *PLoS ONE* **2021**, *16*, e0256714. [CrossRef] [PubMed]
12. Applied Biosystems™. AmpFLSTR™ Identifiler™ Plus PCR Amplification Kit. Available online: https://www.thermofisher.com/order/catalog/product/4427368 (accessed on 26 March 2024).
13. Applied Biosystems™. GlobalFiler™ PCR Amplification Kit. Available online: https://www.thermofisher.com/order/catalog/product/4476135 (accessed on 26 March 2024).
14. Promega Corporation. PowerPlex® Fusion 6C System. Available online: https://www.promega.com.au/products/forensic-dna-analysis-ce/str-amplification/powerplex-fusion-6c-system/?catNum=DC2705 (accessed on 26 March 2024).
15. Goor, R.M.; Forman Neall, L.; Hoffman, D.; Sherry, S.T. A Mathematical Approach to the Analysis of Multiplex DNA Profiles. *Bull. Math. Biol.* **2011**, *73*, 1909–1931. [CrossRef] [PubMed]
16. Bleka, Ø.; Storvik, G.; Gill, P. EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Sci. Int. Genet.* **2016**, *21*, 35–44. [CrossRef] [PubMed]
17. Taylor, D.; Bright, J.-A.; McGovern, C.; Neville, S.; Grover, D. Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations. *Forensic Sci. Int. Genet.* **2017**, *28*, e38–e40. [CrossRef] [PubMed]
18. Bleka, Ø.; Gill, P. *Manual for EuroForMix V3.0*; EuroForMix: Oslo, Norway, 2020.
19. Taylor, D. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Sci. Int. Genet.* **2014**, *11*, 144–153. [CrossRef] [PubMed]
20. Buckleton, J.; Susik, M.; Curran, J.M.; Cheng, K.; Taylor, D.; Bright, J.-A.; Kelly, H.; Wivell, R. A diagnosis of the primary difference between EuroForMix and STRmix™. *J. Forensic Sci.* **2024**, *69*, 40–51. [CrossRef] [PubMed]
21. van Oorschot, R.A.H.; Ballantyne, K.N.; Mitchell, R.J. Forensic trace DNA: A review. *Investig. Genet.* **2010**, *1*, 14. [CrossRef] [PubMed]