



Archived at the Flinders Academic Commons:

<http://dspace.flinders.edu.au/dspace/>

‘This is the peer reviewed version of the following article:

Bright, J.-A., Taylor, D., Kerr, Z., Buckleton, J., & Kruijver, M. (2019). The efficacy of DNA mixture to mixture matching. *Forensic Science International: Genetics*, 41, 64–71. <https://doi.org/10.1016/j.fsigen.2019.02.020>

which has been published in final form at

<https://doi.org/10.1016/j.fsigen.2019.02.020>

Crown Copyright © 2019 Published by Elsevier B.V. This manuscript version is made available under the CC-BY-NC-ND 4.0 license:

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Accepted Manuscript

Title: The efficacy of DNA mixture to mixture matching

Authors: Jo-Anne Bright, Duncan Taylor, Zane Kerr, John Buckleton, Maarten Kruijver

PII: S1872-4973(18)30588-X
DOI: <https://doi.org/10.1016/j.fsigen.2019.02.020>
Reference: FSIGEN 2057

To appear in: *Forensic Science International: Genetics*

Received date: 13 November 2018
Revised date: 15 January 2019
Accepted date: 25 February 2019

Please cite this article as: Bright J-Anne, Taylor D, Kerr Z, Buckleton J, Kruijver M, The efficacy of DNA mixture to mixture matching, *Forensic Science International: Genetics* (2019), <https://doi.org/10.1016/j.fsigen.2019.02.020>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Title

The efficacy of DNA mixture to mixture matching

Authors

Jo-Anne Bright¹, Duncan Taylor^{2,3}, Zane Kerr¹, John Buckleton^{1,4}, Maarten Kruijver¹

1. Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142 New Zealand
2. Forensic Science SA, GPO Box 2790, Adelaide SA
3. School of Biological Sciences, Flinders University, GPO Box 2100 Adelaide. SA 5001, Australia
4. University of Auckland, Department of Statistics, Auckland, New Zealand

Highlights

- We investigate a method to identify a common contributor in two mixed DNA profiles
- The discrimination power is limited by the smallest DNA contribution to the profiles
- We show good ability to find pairs of profiles with a common contributor
- This tool gives the ability to provide intelligence information

Abstract

Standard practice in forensic science is to compare a person of interest's (POI) reference DNA profile with an evidence DNA profile and calculate a likelihood ratio that considers propositions including and excluding the POI as a DNA donor. A method has recently been published that provides the ability to compare two evidence profiles (of any number of contributors and of any level of resolution) comparing propositions that consider the profiles either have a common contributor, or do not have any common contributors. Using this method, forensic analysts can provide intelligence to law enforcement by linking crime scenes when no suspects may be available. The method could also be used as a quality assurance measure to identify potential sample to sample contamination. In this work we analyse a number of constructed mixtures, ranging from two to five contributors, and with known numbers of common contributors, in order to investigate the performance of using likelihood ratios for mixture to mixture comparisons. Our findings demonstrate the ability to identify common donors in DNA mixtures with the power of discrimination depending largely on the least informative mixture of the pair being considered. The ability to match mixtures to mixtures may provide intelligence information to investigators by identifying possible links between cases which otherwise may not have been considered connected.

Keywords

Forensic DNA; investigative information; mixture comparison; Continuous models; Likelihood ratio.

Introduction

DNA databases are a powerful investigative tool, often identifying persons of interest in criminal investigations after comparing a crime scene profile with a large number of known reference profiles. The result of a database search may be provided to investigators as investigative intelligence to assess in conjunction with the wider case information. Typically, DNA databases consist of two sub databases: one containing profiles from known individuals who have either volunteered or been compelled to provide a sample (the database) and one containing profiles collected from samples associated with crime scenes [1] (the crime sample database). Profiles can be compared to link individuals with crime scenes by comparing the separate databases. In addition, crime samples can be compared with other crime samples within the crime sample database to link two or more crimes and recognise unidentified common contributors, who potentially are recidivist offenders. Based on our experience, database searches in most jurisdictions are currently limited to methods based on counting the number of concordant and non-concordant alleles between the samples. In addition, there may be a requirement for the profiles to be single-source or a single component resolved from a mixture, particularly when matching crime profiles.

In recent years, likelihood ratio approaches based on probabilistic genotyping have been advocated for matching crime profiles to individuals [2-5]. The likelihood ratio (LR) is generally accepted as the most powerful and relevant statistic that gives the weight of the DNA evidence [3] (which can be on a log10 scale). It is the ratio of the probability of the observed crime stain (O) given each of two competing hypotheses, H_1 and H_2 , and given all the available information, I . Mathematically, we express this as:

$$LR = \frac{\Pr(O | H_1, I)}{\Pr(O | H_2, I)}$$

Suitable propositions for database searching would be:

H_1 : the DNA profile has originated from the database individual and $N - 1$ unknown contributors,

H_2 : the DNA profile has originated from N unknown, unrelated contributors,

where N is the number of contributors assigned to the profile and the LR is computed for each database individual.

Most database search algorithms do not calculate an LR based on probabilistic genotyping but simply report the number of concordant and non-concordant alleles. For unresolvable or low-level mixtures, however, the use of probabilistic genotyping confers considerable advantages [2, 4], most importantly increased power of discrimination leading to more efficient database searches. More specifically, continuous models effectively take into account stochastic events such as heterozygote imbalance, allelic dropout, locus dropout, allelic drop-in, and stutter, a by-product of the PCR process, which can all complicate interpretation leading to uncertainty in the genotypes of the contributor(s) [6-8].

Recently Slooten [9] described a method to calculate an LR for whether or not a common donor exists between two profiles, M and M' , where there is no longer the requirement for one of the profiles to be unambiguous single-source (termed mixture to mixture matching due to its extension of standard searching). Although the treatment of Slooten is focused on the

drop model (semi-continuous), the treatment is general and, as we will show below, applying the theory to an approach based on a continuous model is straightforward. Both mixtures are first interpreted in isolation, that is, when using the drop model, Slooten proposes to deconvolute the mixtures with separate dropout and drop-in parameters. There is no requirement for the number of contributors to be the same between the mixtures. The propositions considering that the first donor of mixture M denoted D_1 is the same as the first donor of mixture M' , denoted D_1' are:

H_1 : $D_1 = D_1'$ and all other donors of the mixtures are unrelated

H_2 : all donors of both mixtures are unrelated, i.e. the mixtures do not have a donor in common

The proposition set above considers only the first contributor to each mixture. In order to be able to take donor 1, there has to be some ordering on the donors. We consider the donors to be in order of their modelled level of DNA contribution to a profile (from largest to smallest). We explain later how to assign an LR for two mixtures having any common contributor rather than a specific pair of contributor positions. The key formula from Slooten is replicated below:

$$LR(M, M') = \frac{P(M, M' | H_1)}{P(M, M' | H_2)} = \sum_g \frac{P_{\bar{d},c}^-(D_1 = g | M) P_{\bar{d},c}^-(D_1' = g | M')}{p(g)}, \quad \text{Equation 2.7 from [9]}$$

where the prior probability that a person chosen at random from the population has genotype g equals $p(g)$ and the posterior probability (after deconvoluting the mixtures in isolation) that donor 1 of mixture M has genotype g is denoted $P_{\bar{d},c}^-(D_1 = g | M)$, with the subscript \bar{d},c emphasising the dependence on the dropout parameters, $\bar{d} = (d_1, \dots, d_N) \in [0,1]^N$, and a drop-in parameter, $0 \leq c \leq 1$. This setup assumes that alleles within and between loci are drawn independently, i.e. the Hardy-Weinberg and Linkage Equilibrium model, also known as the Product Rule model.

Within this paper we apply Slooten's approach of mixture to mixture matching to a number of complex DNA mixtures interpreted using STRmixTM [8, 10] and demonstrate the efficacy. STRmixTM is a continuous method of DNA profile interpretation that uses the quantitative information from an electropherogram (epg) such as peak heights and molecular weights to calculate the probability of the observed profile (O) given possible genotype combinations (S_j). The genotype combination S_j describes the proposed genotypes of the N assumed contributors at a particular locus. A numerical value, or weight (w_j), is assigned to the normalised probability density $\Pr(O | S_j)$. STRmixTM assigns a relative weight to the probability of the epg given each possible genotype combination at a locus. The weights across all combinations at that locus are normalised so that they sum to one. Therefore, a single unambiguous genotype combination at any locus would be assigned a weight of one. Using STRmixTM nomenclature, genotype weights are proportional to the probability of the

profile given a proposed genotype combination i.e. $w_j \propto \Pr(O | S_j)$. To apply the formula above we need the reverse: $\Pr(S_j | O)$, which is the posterior probability of a proposed genotype combination and can be obtained using the general form of Bayes' Theorem:

$$\Pr(S_j | O) = \frac{\Pr(O | S_j) \Pr(S_j)}{\sum_k \Pr(O | S_k) \Pr(S_k)} = \frac{w_j \Pr(S_j)}{\sum_k w_k \Pr(S_k)}.$$

The posterior probabilities of genotype combinations (i.e. for all contributors considered jointly), $\Pr(S_j | O)$, can be used to compute posterior probabilities of genotypes for the contributors separately. Specifically, the posterior probability that donor 1 has genotype g is computed as:

$$\Pr(D_1 = g | M) = \sum_{j: \{S_{j,i} = g\}} \Pr(S_j | O),$$

where $S_{j,i}$ indicates the i^{th} genotype in genotype combination j . After computing the posterior probabilities, we can then apply Slooten's equation 2.7. The equation is explicitly stated for donor 1 though the extension to other donors is obvious.

The aim of this work is to demonstrate the efficacy of mixture to mixture matching using the method of Slooten and applied to continuous probabilistic genotyping interpretations. The results may be used to inform practices (e.g. what quality of mixture it should be applied to, and appropriate list management thresholds to use) when using the method in active casework. We also discuss the differences in the interpretation of mixture to mixture matching, compared with standard *LR* calculations where one of the profiles is a reference sample.

Method

Choice of profiles

Six each of two-, three-, four-, and five-person GlobalFiler™ mixtures from the publicly available PROVEDIt dataset [11] were analysed in GeneMapper® *ID-X* version 1.5 following Kelly et al. [12]. The mixtures were selected so that where possible half of them shared common contributors and the other half did not. Moreover, it was ensured that a range of mixtures had one, two, three, four, or five contributors in common. The chosen mixtures cover a range of complexities in regards to the number of contributors, mixture proportions, and total template DNA amounts, and also include several samples artificially degraded using DNase I. A summary of the profiles selected, the known contributors, mixture ratios, and total template DNA is given in the appendix (Table 1). An overview of the number of common contributors between the profiles is given in Figure 1.

LR calculation

Profiles were interpreted in STRmix™ version 2.6 using the known number of contributors (i.e. the experimentally designed number). All 24 deconvoluted mixtures were then compared with each other resulting in 276 mixture to mixture comparisons. When comparing two mixtures M and M' , we evaluate the following propositions:

H_{1A} : $D_i = D_j$ and all other donors of the mixtures are unrelated,

H_{2A} : all donors of both mixtures are unrelated,

where donors i and j are defined with respect to mixtures M and M' respectively. The order of the contributors in each mixture is by template contribution (from largest to smallest based on the deconvolution results of STRmix™). For these propositions, the likelihood ratio is denoted as $LR_{i,j}$. When considering the comparison of each contributor in each mixture with each contributor in every other mixture the result was a total of 3,366 LR s assigned.

Additionally, for each mixture to mixture comparison, an LR not specifying a specific contributor pair was computed, i.e. an LR was computed for the propositions:

H_{1B} : the two mixtures share one common donor,

H_{2B} : all donors of both mixtures are unrelated,

which is denoted simply as LR (that is without subscripts indicating specific contributors). Note that H_{2A} is the same as H_{2B} . To evaluate this LR , it is noted that the numerator proposition naturally unfolds into sub-propositions for each possible pairwise comparison. For instance, assuming mixture M was analysed as having N contributors and mixture M' was analysed as having N' contributors, we may write:

$$H_{1B} = \bigcup_{\substack{i=1,\dots,N, \\ j=1,\dots,N'}} D_i = D_j \text{ and all other donors of the mixtures are unrelated,}$$

i.e. as a union of mutually exclusive sub-propositions. Moreover, to not complicate the analysis too much, we have assumed a uniform prior on the sub-propositions so that the LR is simply equal to the average of all corresponding pairwise LR s.

$$LR = \frac{1}{NN'} \sum_{i=1,\dots,N} \sum_{j=1,\dots,N'} LR_{i,j}$$

This assumption means that there was no *a priori* belief that the common donor was more likely to be a particular contributor to the mixture. Although this may not be entirely realistic for all cases, the effect of specifying a different prior is likely to be small. Note that in our study we compare mixtures that have between 0 and 5 contributors in common. We do this as we wish to know the effect of more than one contributor in common between mixtures (a factor that will not be known in casework samples).

Population stratification

The FBI extended allele frequencies [13] were used in the LR calculations separately for Caucasian, African American, and South Eastern Hispanic populations, without subpopulation correction (i.e. assuming Hardy-Weinberg equilibrium and linkage equilibrium). The PROVEDIt dataset [11] was constructed to consist of mixtures with donors originating from different and unspecified ethnicities. In practice the true ethnic composition of crime scene samples is also not known and different contributors may have different or perhaps mixed ethnicities. Unrepresentative allele frequency databases may lead to inflated LR s since the probability of the alleles shared between samples is typically underestimated

[14]. Therefore, we use the minimum LR among the three populations with the intention of understating the weight of evidence:

$$LR^{\min} = \min \{ LR^{\text{Caucasian}}, LR^{\text{African American}}, LR^{\text{South Eastern Hispanic}} \}.$$

Note that LR^{\min} is a minimum of likelihood ratios and is not strictly a likelihood ratio itself.

A perhaps more accurate evaluation of the evidential value would involve extending the calculations to include the possibility of contributors from different populations, however no such extension was attempted in the current work.

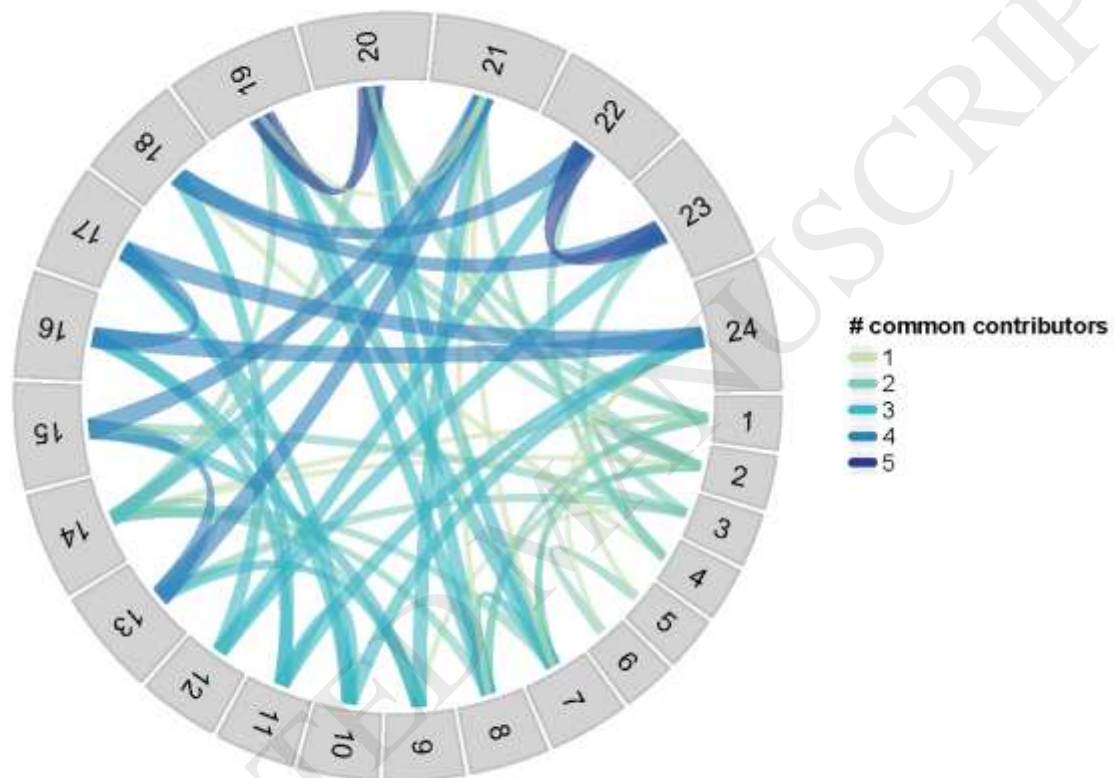


Figure 1 Overview of the number of common contributors between the 24 profiles that were analysed. The darker the colour, the greater the number of shared contributors.

Results

The efficacy of mixture to mixture comparison is illustrated by investigating the power of discrimination of the LR method as measured by sensitivity and specificity. Loosely speaking, sensitivity relates to how likely it is that the presence of a common donor between two mixtures leads to a large LR , while specificity relates to how unlikely it is that a large LR is obtained where there are no common donors between the mixtures. For single-source to mixture comparisons, sensitivity and specificity plots have been presented (e.g. [15, 16]) that display LR s for H_1 and H_2 true as a function of an explanatory variable such as the template DNA amount for a true contributor to the mixture or the average peak height of a true contributor. Similar plots can be produced for mixture to mixture matching. We first

examine the ability of the method to correctly identify true links between mixed DNA profiles.

An overview of the true links that were successfully identified and those that were missed using the method is given in Figure 2. A true link is assumed to be identified if LR^{\min} exceeds an arbitrary threshold of one million and otherwise it is incorrectly not identified. Inspecting Figure 1 and Figure 2, it is immediately clear that the majority of true links could be correctly identified. Specifically, 53 out of 61 true links led to a statistic exceeding one million. All missed links involve samples 14 or 21 (Appendix: Table 1). Sample 21 has one low level contributor (reference 30) who has dropped out at multiple loci. Review of the LR s for sample 14 across the three populations revealed that these links were correctly identified using the Caucasian allele frequencies (i.e. $LR^{\text{Caucasian}}$) however the African American and South Eastern Hispanic allele frequencies produced LR values below 1 million, hence LR^{\min} was also below 1 million.

The largest LR^{\min} for a non-associated pair (by this we mean a pair without a common contributor) was approximately 3. This was for the comparison of samples 3 (two person 1:4) and 19 (5 person 1:1:1:1:1). For this comparison, $LR^{\text{Caucasian}}$ was the largest being approximately equal to 6, so the choice of allele frequencies did not matter. Other comparisons, however, showed meaningful differences between the likelihood ratios calculated using different allele frequency sets. Most notably, LR^{\min} for the comparison of samples 1 and 22 (not sharing a common donor) equals about one, while $LR^{\text{Caucasian}}$ equals approximately 18 thousand. The small number of inclusionary likelihood ratios suggests the use of a match threshold smaller than one million could be considered for mixture to mixture matching. In the present study, use of a match threshold of 1,000 would have recovered the missed links for sample 14 whilst still avoiding adventitious links.

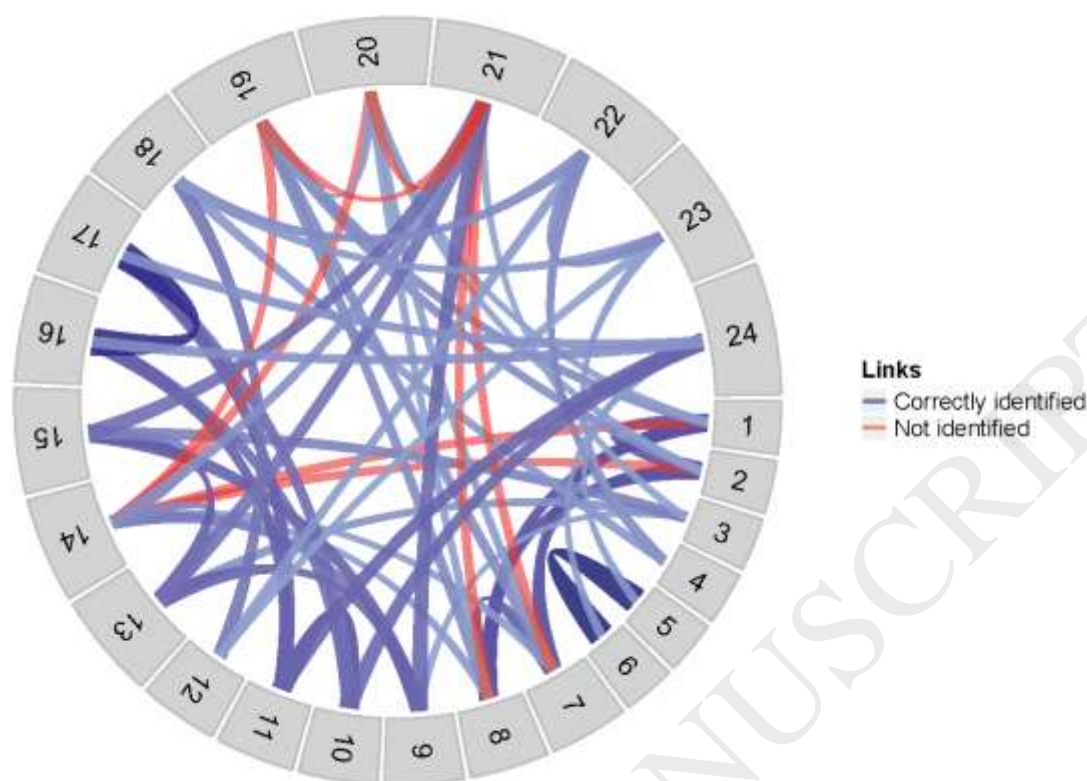


Figure 2. Overview of the true links that were correctly identified and the ones that were missed when links were assumed to be present if LR^{\min} exceeded one million. There were no false associations. The largest LR^{\min} for a non-associated pair was about 3. The darker the colour, the higher the LR .

The graphical overviews in Figure 1 and Figure 2 allow for a side by side comparison because the number of samples in this study is not too large. In a practical investigative or quality assurance setting, however, we suspect that mixture to mixture comparisons are potentially implemented across too many deconvolutions for such a visualisation to be meaningfully interpreted. In such cases it is perhaps more revealing to present a table of comparisons that yielded an LR greater than some threshold set for list management. Alternatively, one could construct an undirected graph with mixtures as nodes and edges between nodes whenever the LR exceeds some threshold.

We present the pairwise LR s and compare those with the LR s not computed for specific contributor pairs in Figure 3. A number of observations can be made from the plot. Most importantly, there is good separation between the LR s obtained for comparisons where there is a common donor (green) and those produced where there are no common donors between the mixtures compared (red). This separation demonstrates again that the sensitivity and specificity of the method is sufficient to discriminate mixtures that do share a donor from those that do not.

To better understand the differences observed between the LR s and the pairwise $LR_{i,j}$ values, we need to acknowledge that there is often limited power to distinguish between different

donors in the same mixture when the template is more or less evenly divided between the donors. For example, consider two mixtures, each originating from two contributors in equal amounts, and assume that there is one common donor between these two mixtures. The LR considers the following sub-propositions in the numerator:

- Contributor 1 from mixture 1 is contributor 1 from mixture 2
- Contributor 1 from mixture 1 is contributor 2 from mixture 2
- Contributor 2 from mixture 1 is contributor 1 from mixture 2
- Contributor 2 from mixture 1 is contributor 2 from mixture 2

Since the mixtures are from equally contributing donors, there will be an approximately equal weight attached to the genotype sets comprising the same genotype pairs but in the reverse order. In other words, each genotype is about as good an explanation for contributor 1 as for contributor 2 in both mixtures. Therefore, even though there is only one contributor in common in our example, it would be expected that all four $LR_{i,j}$ values (for the four sub-propositions listed above) would yield equivalent pairwise LR s (i.e. $LR_{1,1} \approx LR_{1,2} \approx LR_{2,1} \approx LR_{2,2}$).

The LR (without specifying specific contributors) considers the mixture comparison as a single comparison rather than a number of comparisons for each component pair, akin to the difference between sub-sub-source and sub-source LR s as described in Taylor et al. [17]. By assuming a uniform prior on which contributor pair is the common donor, the LR can be obtained by averaging the pairwise $LR_{i,j}$ values. By doing this we expect to see two trends in LR produced by mixture comparisons that mean the results differ slightly from standard specificity and sensitivity experiments.

- First, the comparison of mixtures that do not have any contributors in common will tend to be closer to $LR=1$ than seen in standard specificity and sensitivity tests (particularly when both mixtures have a low-level contributor). The reason for this is that there are multiple component pairs being compared, and (again, particularly if there is a low-level contributor in at least one of the mixtures) then the largest $LR_{i,j}$ (which will be the bounding value for LR) will dominate the average and is expected to be approximately neutral. Therefore, if we compared the $LR_{i,j}$ values from the contributor component pairs to the average across all pairs, we would expect a shift of the large exclusions to a value of $LR=1$.
- Secondly, for those mixtures with multiple contributors in common, we expect LR to more closely align with the largest of the pairwise LR s, resulting in LR showing, in general, a more noticeable shift to higher values from the corresponding pairwise $LR_{i,j}$ values.

Figure 3 shows two sets of data; the green points show the comparison of $LR_{i,j}$ to LR for H_{1A} true and H_{1B} true values. The red points show the comparison of $LR_{i,j}$ to LR for H_{2A} true and H_{2B} true values. Note that we do not plot comparisons of $LR_{i,j}$ to LR for H_{2A} true and H_{1B} true values as this data is not helpful to the interpretation of trends (note that this means a plot such as in Figure 3 can only be produced knowing the donors that the mixtures have in common). In Figure 3 all LR s for are calculated using the African American population database. The choice of the African American database (as opposed to the Caucasian or South Eastern Hispanic) was arbitrary. A minimum of the three databases was not used as we could consider a minimum on either variable, and these may not be the same database. In

any event, the results of Figure 1 are mainly produced to observe a general trend of results from the production of LR from $LR_{i,j}$. The two trends just described can be seen in Figure 3. Note that for the pairs with a true common donor(s), only those that have more than one contributor in common show a significant shift from the $x = y$ line.

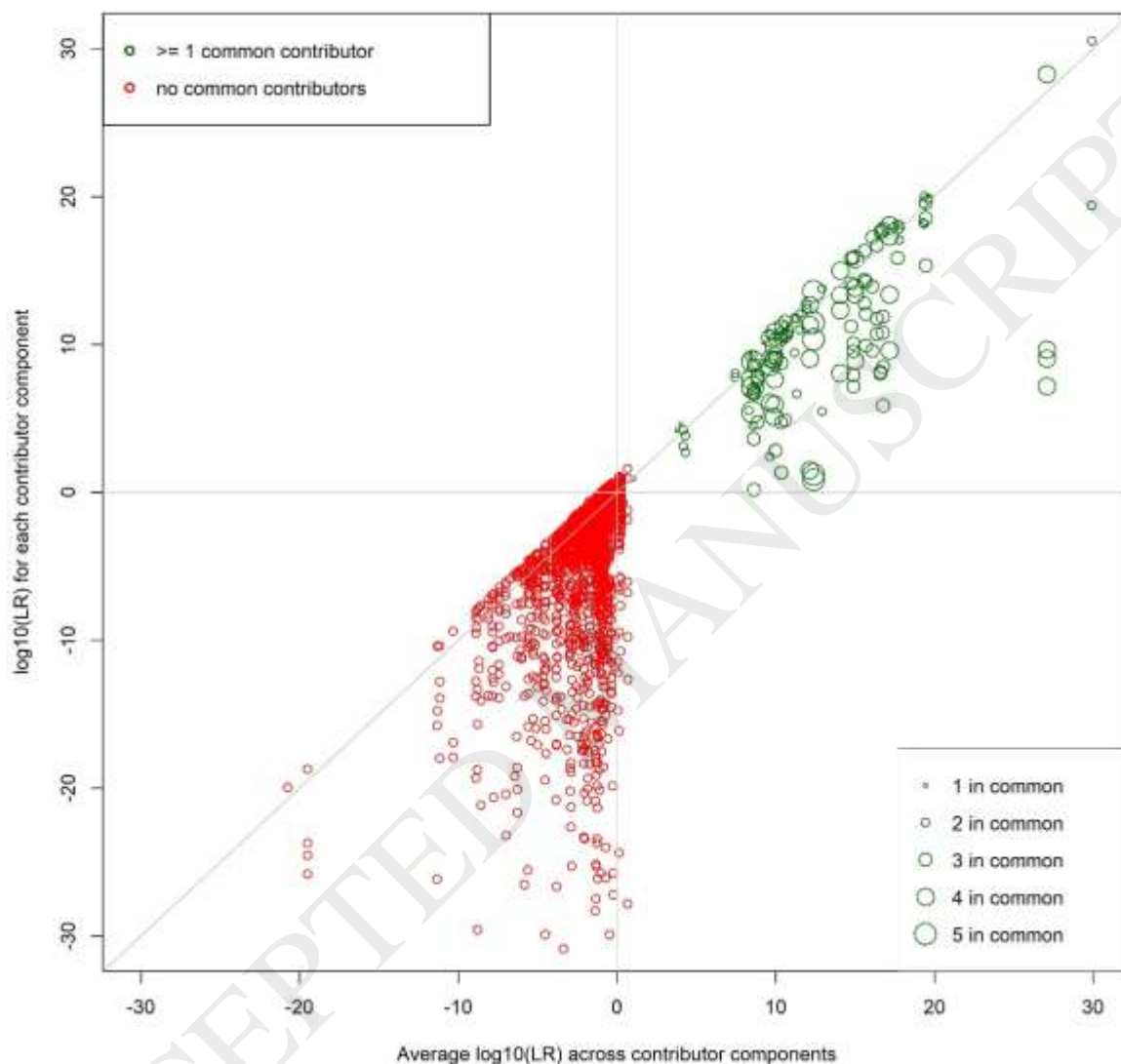


Figure 3: Comparison of $\log_{10}(LR_{i,j})$ (from contributor component pairwise comparisons), compared with the $\log_{10}(LR)$ (across all components). The size of the green symbols represents the number of contributors in common. The green points show the comparison of $LR_{i,j}$ to LR for H_{1A} true and H_{1B} true values. The red points show the comparison of $LR_{i,j}$ to LR for H_{2A} true and H_{2B} true values.

The second consideration when plotting sensitivity and specificity for mixture to mixture matching is the appropriate dependant variable. Taylor [16] plotted H_1 true LR s against the amount of input DNA from a known contributor used to construct the mixtures examined. This was relatively simple as there was only one DNA amount to consider. DNA template amount was found to significantly affect the ability to discriminate between true and false donors, with LR s trending towards one as template amount decreased. For the mixtures used

in the present study the targeted input amount of DNA for each contributor can be determined using information regarding the total template amount and mixture ratio for each sample. When comparing mixtures, we consider the same limiting factor of input DNA on the LR^{\min} , i.e. the lowest DNA donation by the contributor to the two mixtures will be the limiting factor in the discrimination power, and hence minimum donation is the value that should be used for displaying the results of sensitivity and specificity plots. This is further complicated when there are multiple contributors in common. In this instance there will be multiple pairs of donations, each with their own minimum and maximum template DNA amounts. In these cases, the discrimination power of the LR^{\min} will be the maximum of the minimums in the pairs of DNA donation amounts¹. This will now be the appropriate value used for displaying the results of sensitivity and specificity plots LR^{\min} values for pairs with a common contributor. For pairs without a common contributor, we use the minimum amount of DNA for any contributor in either mixture being compared (as this is the most likely contributor position to give rise to false inclusions).

Additionally, Taylor [16] demonstrated the effect of profile complexity on sensitivity and specificity by separating the plots produced according to the known number of contributors, N , of each of the mixtures examined. When considering mixture to mixture comparisons each mixture will have its own N which may or may not equal that of the other mixture. As such, plotting the LR^{\min} values produced according to the number of contributors is not straightforward and we have instead plotted all data on a single graph regardless of profile complexity. We would expect however that profile complexity will affect the discriminatory power of mixture to mixture comparisons with the more complex of the two mixtures under comparison being the limiting factor.

Figure 4 shows the results of the sensitivity and specificity plot, constructed within the limitations just described. The $\log_{10} LR^{\min}$ values are the minimum obtained from calculations in all three population databases. In order to provide some information on profile complexity, the size of the symbols in Figure 4 represent the summed number of contributors between the two mixtures being compared. The lines shown on Figure 4 are loess lines [18, 19] and are present to assist in observing visual trends. The disjunction in the loess lines comes from the categorical nature of the DNA amounts (particularly in the green points). As expected, the highest LR^{\min} values for the lower three matching contributor pair DNA amounts are from the smallest summed N , i.e. four. This is not the case for higher DNA

¹ For example, samples 9 and 14 have two common contributors (29 and 50). Contributor 29 donates approximately 0.031 ng to sample 9 (obtained by the ratio of 1:4:1 in sample 9 where contributor 29 represents the terminal 1, and a total input of 0.186 ng, i.e. $0.186 \times 1/(1+4+1) \sim 0.031$) and 0.125 ng to sample 14. The minimum of these two values (0.031 ng and 0.125 ng) is 0.031ng, which will likely be the donation that limits the discrimination power for contributor 29. Contributor 50 donates 0.124 ng to sample 9 and 0.125 ng to sample 14. The minimum amount of DNA for contributor 50, and limiter of $LR_{i,j}$ size, is 0.124 ng. The minimum contribution by contributor 29 is 0.031 ng and the minimum contribution by contributor 50 is 0.124 ng. The contribution that will limit the size of the LR comparing samples 9 and 14 will be the maximum of these minimum values, which in this case is 0.124 ng. Therefore, the comparison between samples 9 and 14 for the LR is plotted at 0.124 ng. To put this another way, if we renumber the donors such that M and M' have the first k donors in common, then LR will be determined by the strongest $LR_{i,j}$. For a particular common donor, j , the expected $LR_{i,j}$ correlates with the smallest contribution of donor j to either M or M' . Therefore, the explanatory variable is set as the maximum value over all i , of the minimal template contribution of donor i to M or M' .

amounts as there were no two-person to two-person comparisons with the maximum of minimum DNA from common contributors this high.

Finally, we compute the average LR s obtained for profile pairs without a common contributor as a sanity check. As with any likelihood ratio, we expect the likelihood ratios produced from mixture to mixture comparisons to adhere to Turing's lemma that "*the expected factor for a wrong hypothesis in virtue of any experiment is 1.*" (from Good [20], quoting Turing) as demonstrated for standard profile comparisons in [21]. We do not strictly sample from the denominator proposition, because we have assumed the Product Rule model for the separate LR calculations, which is not realistic for this dataset. In this instance, the average of the LR across all mixture comparisons that had no common contributors was 291, 3 and 0.16 using the Caucasian, South Eastern Hispanic, and African American population databases, respectively. These numbers are plausibly close to 1 and suggest that the system is not overstating the evidence to a great extent.

We note that users of probabilistic genotyping systems are likely to be familiar with the concept of adventitious matching (although the term 'matching' does not sit well in the context of mixed DNA profiles), in two ways; the results of multiple comparison between a reference and numerous profiles in a database [22], and also the multiple comparisons of a reference to different components to a mixture [23]. The latter issue is also present in mixture to mixture comparisons, with the difference that both profiles being mixed will lead to more comparisons than the more traditional comparison of a reference to a mixture. The first issue is also relevant here, since comparing every mixture to every other mixture will rapidly lead to a large number of LR s as the number of comparisons grows quadratically with the number of mixtures. For the single source analogue, it has been demonstrated that as databases grow larger there will eventually be pairs of profiles sharing a number of alleles that is surprisingly large to some [24]. When conducting very many mixture to mixture comparisons, one will ultimately find strong adventitious results even though those are very rare when viewed in isolation.

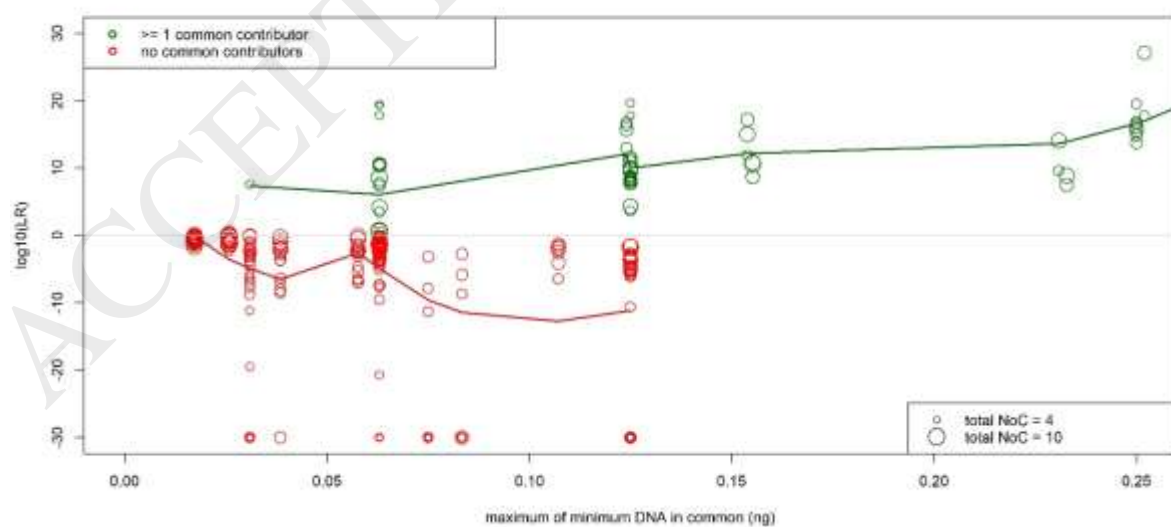


Figure 4: $\log_{10}(LR^{\text{min}})$ values for mixture comparisons across the varying input DNA amounts. The size of the symbol represents the summed number of contributors shared between the

two mixtures. Red points represent LR values for comparisons of mixtures for which there are no contributors in common. Green points represent LR values for comparisons of mixtures for which there is at least one contributor in common.

Conclusion

We have shown that mixture to mixture comparisons function in the same way as standard comparisons of reference to evidence profiles, i.e.:

- As the amount of DNA decreases, the LR^{min} contracts around one. This trend is less pronounced than for the comparison of reference profiles to mixed samples, due to the complicating nature of the mixture to mixture comparisons as outlined in our work. Nevertheless, a trend can be seen in Figure 4 in the datapoints where there are one or more contributors in common from $\log_{10}(LR^{min}) \approx 20$ at 0.25 ng, decreasing to $\log_{10}(LR^{min}) \approx 5$ at 0.05 ng. This behaviour is known from previous studies comparing reference profiles to mixed profiles [15, 16, 25]. Our study simply shows the strength of that trend for comparison of two mixed profiles.
- As the complexity of the comparison (i.e. the combined number of contributors in the two mixtures being compared) increases, the LR^{min} contracts around one. This conclusion may need to be drawn (at least in part) by a thought experiment. Consider a scenario where two matching profiles being compared are complete and single source (the simplest of comparisons). We can calculate an LR in the way that has been possible for decades, which would be equal to the inverse of the match probability. Now consider adding a contributor to one of these profiles, and we have a standard comparison of a reference to a two person mixture, we know that the LR must be equal to or less than the LR from the single source example. There will be instances when the LR s are virtually the same (i.e. when we are talking about a case of the two person profile having a clear matching major contributor), but the general trend (across all two person mixtures) would be a lower LR than the single source example. We would expect (and has been shown [16]) this trend of yielding less informative LR s to continue if we kept adding more contributors to the mixed sample. By extension we would also expect the same trend to continue if we then started adding contributors to the remaining single source profile, with the general trend being less informative LR s as more contributors are added. Again there are properties of profiles that would lead to extremes within this trend such as a common major donor would be less affected by adding contributors, and the greatest effect on informativeness will be obtained by adding contributors of approximately the same intensity as the common contributor. The trend that has just been described is only weakly observable in our study. This is most easily seen in Figure 4 for the DNA amounts where comparisons of two-person mixtures to two-person mixtures are carried out (~0.06 ng and 0.125 ng). In these two areas the comparisons arising from two-person mixtures to two-person mixtures can be seen at $\log_{10}(LR^{min}) \approx 20$, whereas the comparisons of more complex mixtures tend to be below this, and in some instances at $\log_{10}(LR^{min}) \approx 0$.

There are a couple of differences that must be kept in mind:

- The comparisons cannot be carried out between components of mixtures and designated as being H_{1A} or H_{2A} true tests in a meaningful manner. Instead we recommend the average LR across all contributor component comparisons be used.
- When LR is used there is a strong drive for H_{2B} true LR s towards one (more pronounced than for standard H_2 tests comparing reference profiles to evidence samples). For mixtures with a single contributor in common, the difference between $LR_{i,j}$ for the known common contributor component to LR is minimal. When there is more than one contributor in common, the smaller $LR_{i,j}$ value(s) for the known common contributor component comparisons will be driven towards the largest $LR_{i,j}$ value for the known common contributor component comparison. Again, this is not surprising behaviour as the expectation for all individual H_2 comparisons is $LR = 1$, hence the average of multiple H_2 LR s will tend to contract around this mean.

We note that the mathematics published by Slooten considers the propositions:

- H_{1B} : the mixtures have one common contributor
- H_{2B} : the mixtures have no common contributors

A more general form would be to consider propositions:

- H_{1C} : the mixtures have M common contributor(s)
- H_{2C} : the mixtures have N common contributors(s)

where $N \geq 0$ and $M > N$, and test multiple values for M and N as required by the case circumstances. We do not attempt the generalisation of the mathematics of Slooten to accommodate such propositions, but the general structure would be the same, simply with multiple genotype sets being considered. However, the results of our study show that the benefit of this added complexity may not be warranted, as the effect of multiple contributors in common, over a single contributor in common is only slight, if noticeable at all. Any trend to increase the LR with increased common contributors is likely to arise from the averaging process alone. For example, consider comparing two four-person mixtures, and for simplicity consider that all four components in both mixtures have completely and unambiguously resolved genotypes. First, consider that the two mixtures have a single common contributor, and comparison of the mixture components that hold this donor yields a $\log_{10}(LR)$ of 20. All other comparisons lead to $LR = 0$ and so the total (by averaging across all component comparisons) is $10^{20} / 16$ i.e. $\log_{10}(LR) \approx 18.8$. Now consider the situation that all four contributors between the two mixtures are in common. Therefore four component yield $\log_{10}(LR) \approx 20$ and 12 LR s of 0. Now the total (averaged) LR is $4 \times 10^{20} / 16$ i.e. $\log_{10}(LR) \approx 19.4$. Taking the effect of different levels of contribution out of consideration, this thought experiment demonstrates the very minor effect that adding more common contributors to a mixture to mixture comparison will have on the component average LR .

In this paper we have demonstrated the efficacy of assigning likelihood ratios for common contributors when comparing mixed DNA profiles. Possibly the most obvious use for mixture to mixture comparison is to provide intelligence to law enforcement on potentially linked crimes, for which no suspect has yet been identified. Another use would be in a quality assurance context, i.e. searching samples within laboratory processing batches to determine whether sample to sample contamination may have occurred. The method could

also be used when comparing weak and partial single-source reference samples to mixed DNA profiles. Traditionally, a common practice has been to ignore loci that are partial and only carry out the comparison using loci that are complete in the reference.

Acknowledgements

This work was supported in part by grant NIJ 2017-DN-BX-0136 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 their organizations.

References

- [1] Walsh SJ, Buckleton JS. DNA Intelligence Databases. In: Buckleton JS, Triggs MC, Walsh SJ, editors. Forensic DNA evidence interpretation. Boca Raton: CRC Press; 2005.
- [2] Bright J-A, Taylor D, Curran J, Buckleton J. Searching mixed DNA profiles directly against profile databases. *Forensic Science International: Genetics*. 2014;9:102-10.
- [3] Berger CEH, Vergeer P, Buckleton JS. A more straightforward derivation of the LR for a database search. *Forensic Science International: Genetics*. 2015;14:156-60.
- [4] Benschop CCG, van de Merwe L, de Jong J, Vanvooren V, Kempnaers M, van der Beek CP, et al. Validation of SmartRank: A likelihood ratio software for searching national DNA databases with complex DNA profiles. *Forensic Science International: Genetics*. 2017;29:145-53.
- [5] Bleka Ø, Dørum G, Haned H, Gill P. Database extraction strategies for low-template evidence. *Forensic Science International: Genetics*. 2014;9:134-41.
- [6] Caddy B, Taylor G, Linacre A. A review of the science of low template DNA analysis. UK Home Office report, May 2008.
- [7] The Forensic Science Regulator. Response to Professor Brian Caddy's Review of the Science of Low Template DNA Analysis. 2008.
- [8] Bright J-A, Taylor D, Curran JM, Buckleton JS. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. *Forensic Science International: Genetics*. 2013;7:296-304.
- [9] Slooten K. Identifying common donors in DNA mixtures, with applications to database searches. *Forensic Science International: Genetics*. 2017;26:40-7.
- [10] Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*. 2013;7:516-28.
- [11] Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM. A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Science International: Genetics*. 2018;32:62-70.
- [12] Kelly H, Bright J-A, Kruijver M, Cooper S, Taylor D, Duke K, et al. A sensitivity analysis to determine the robustness of STRmix™ with respect to laboratory calibration. *Forensic Science International: Genetics*. 2018;35:113-22.
- [13] Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, et al. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. *Forensic Science International: Genetics*. 2016;25:175-81.

- [14] Gittelsohn S, Moretti TR, Onorato AJ, Budowle B, Weir BS, Buckleton J. The factor of 10 in forensic DNA match probabilities. *Forensic Science International: Genetics*. 2017;28:178-87.
- [15] Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright J-A, et al. Internal validation of STRmix for the interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*. 2017;29:126-44.
- [16] Taylor D. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Science International: Genetics*. 2014;11:144-53.
- [17] Taylor D, Bright JA, Buckleton J. The 'factor of two' issue in mixed DNA profiles. *Journal of Theoretical Biology*. 2014;363:300-6.
- [18] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Amer Statist Assoc*. (1979; 74:829-36.
- [19] Cleveland WS. LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*. 1981;35:54.
- [20] Good IJ. *Probability and the Weighing of Evidence*. London: Charles Griffin & Company Limited; 1950.
- [21] Taylor D, Buckleton J, Evett I. Testing likelihood ratios produced from complex DNA profiles. *Forensic Science International: Genetics*. 2015;16:165-71.
- [22] Storvik G, Egeland T. The DNA Database Search Controversy Revisited: Bridging the Bayesian–Frequentist Gap. *Biometrics*. 2007;63:922-5.
- [23] Taylor D, Bright J-A, Buckleton J. The 'factor of two' issue in mixed DNA profiles. *Journal of Theoretical Biology*. 2014;363:300-6.
- [24] Weir BS. The rarity of DNA profiles. *Annals of Applied Statistics* 2007;1:358-70.
- [25] Bright J-A, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, et al. Internal validation of STRmix™ – A multi laboratory response to PCAST. *Forensic Science International: Genetics*. 2018;34:11-24.

Appendix:

Table 1: Summary of profiles interpreted, number of contributors, known contributors, mixture ratios, and total template DNA added to PCR

No.	Sample File	<i>N</i>	Contributors	Mixture ratio	Total template (ng)
1	B01_RD14-0003-31_32-1;1-M1a-0.25GF-Q1.2_02.15sec.hid	2	31, 32	1:1	0.25
2	C01_RD14-0003-31_32-1;1-M2a-0.126GF-Q1.2_03.15sec.hid	2	31, 32	1:1	0.126
3	G05_RD14-0003-48_49-1;4-M2c-0.625GF-Q0.7_07.15sec.hid	2	48, 49	1:4	0.625
4	C10_RD14-0003-39_40-1;2-M2c-0.375GF-Q1.0_03.15sec.hid	2	39, 40	1:2	0.375
5	C04_RD14-0003-42_43-1;9-M2a-0.75GF-Q0.5_03.15sec.hid	2	42, 43	1:9	0.75
6	E04_RD14-0003-42_43-1;9-M2a-0.31GF-Q0.5_05.15sec.hid	2	42, 43	1:9	0.31
7	A10_RD14-0003-30_31_32-1;4;4-M2d-0.567GF-Q0.6_01.15sec.hid	3	30, 31, 32	1:4:4	0.567
8	C09_RD14-0003-30_31_32-1;4;4-M2a-0.75GF-Q0.6_03.15sec.hid	3	30, 31, 32	1:4:4	0.75
9	A08_RD14-0003-49_50_29-1;4;1-M3a-0.186GF-Q0.5_01.15sec.hid	3	49, 50, 29	1:4:1	0.186
10	B06_RD14-0003-44_45_46-1;2;2-M2c-0.625GF-Q0.8_02.15sec.hid	3	44, 45, 46	1:2:2	0.625
11	F05_RD14-0003-44_45_46-1;2;2-M2a-0.625GF-Q0.6_06.15sec.hid	3	44, 45, 46	1:2:2	0.625
12	F01_RD14-0003-36_37_38-1;2;1-M2c-0.5GF-Q0.6_06.15sec.hid	3	36, 37, 38	1:2:1	0.5
13	H10_RD14-0003-48_49_50_29-1;4;4;4-M2d-0.75GF-Q0.7_08.15sec.hid	4	48, 49, 50, 29	1:4:4:4	0.75
14	B01_RD14-0003-50_29_30_31-1;1;2;1-M2a-0.625GF-Q0.5_02.15sec.hid	4	50, 29, 30, 31	1:1:2:1	0.625
15	E10_RD14-0003-48_49_50_29-1;4;4;4-M2c-0.5GF-Q0.6_05.15sec.hid	4	48, 49, 50, 29	1:4:4:4	0.5
16	C03_RD14-0003-44_45_46_47-1;1;4;1-M2d-0.75GF-Q0.8_03.15sec.hid	4	44, 45, 46, 47	1:1:4:1	0.75
17	D03_RD14-0003-44_45_46_47-1;1;4;1-M2d-0.441GF-Q0.8_04.15sec.hid	4	44, 45, 46, 47	1:1:4:1	0.441
18	C04_RD14-0003-36_37_38_39-1;2;2;1-M2a-0.75GF-Q0.7_03.15sec.hid	4	36, 37, 38, 39	1:2:2:1	0.75
19	C06_RD14-0003-30_31_32_33_34-1;1;1;1;1-M2a-0.315GF-Q0.7_03.15sec.hid	5	30, 31, 32, 33, 34	1:1:1:1:1	0.315
20	F06_RD14-0003-30_31_32_33_34-1;1;1;1;1-M1d-0.625GF-Q1.5_06.15sec.hid	5	30, 31, 32, 33, 34	1:1:1:1:1	0.625
21	D03_RD14-0003-48_49_50_29_30-1;1;2;4;1-M3a-0.567GF-Q0.8_04.15sec.hid	5	48, 49, 50, 29, 30	1:1:2:4:1	0.567
22	A08_RD14-0003-36_37_38_39_40-1;9;9;9;1-M2a-0.5GF-Q0.7_01.15sec.hid	5	36, 37, 38, 39, 40	1:9:9:9:1	0.5
23	C09_RD14-0003-36_37_38_39_40-1;9;9;9;1-M1e-0.75GF-Q2.7_03.15sec.hid	5	36, 37, 38, 39, 40	1:9:9:9:1	0.75
24	B01_RD14-0003-43_44_45_46_47-1;1;2;1;1-M2a-0.75GF-Q0.8_02.15sec.hid	5	43, 44, 45, 46, 47	1:1:2:1:1	0.75