A mixed DNA profile controversy revisited

Tim Kalafut Ph.D.¹, Simone Pugh M.S.², Peter Gill Ph.D.³⁴, Sarah Abbas M.Sc.⁵⁶, Marie Semaan M.Sc.⁵, Issam Mansour Ph.D.⁵, James Curran Ph.D.⁷, Jo-Anne Bright Ph.D.⁸, Tacha Hicks Ph.D.⁹¹⁰, Richard Wivell B.Sc. (hons)⁸, John Buckleton D.Sc.⁷⁸

1. Department of Forensic Science, College of Criminal Justice, Sam Houston State University, Huntsville, TX 77340
2. California Department of Justice, Redding, CA, USA
3. Forensic Genetics Research Group, Oslo University Hospital, Oslo, Norway
4. Department of Clinical Medicine, University of Oslo, Oslo, Norway
5. Department of Laboratory Science and Technology, Faculty of Health Sciences, American University of Science and Technology, Beirut, Lebanon
6. School of Criminal Justice, University of Lausanne, Lausanne, Switzerland
7. Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
8. Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142 New Zealand
9. Forensic Genetics Unit, University Center of Legal Medicine, Lausanne - Geneva, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
10. Fondation pour la formation continue Universitaire Lausannoise (UNIL-EPFL), 1015 Dorigny, Switzerland

Acknowledgements
This work was supported in part by grant NIJ 2020-DQ-BX-0022 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.

--

Previous to this submission, the problem described in this paper was presented by three of the authors (Semaan, Abbas, Mansour) in the Journal of Forensic Research (2020, 11:2), the 21st Triannual meeting of the International Association of Forensic Sciences in 2017, and at the 73rd American Academy of Forensic Sciences (AAFS) meeting in February 2021. Semaan, Abbas, and Mansour have agreed to join the rest of the authors in acknowledging the initial presentations were incorrect and this paper describes a more appropriate solution.

This work was partially funded as listed in the acknowledgements statement. There are no other disclaimers other than what is in the acknowledgements. There are no conflicts of interest.
A mixed DNA profile controversy revisited

**ABSTRACT**

Semaan et al. (1) discuss a mock case “where eight different individuals [P₁ through P₈] could not be excluded in a mixed DNA analysis. Even though ... expert DNA mixture analysis software was used.” Two of these are the true donors.

The $LR$s reported are incorrect due to the incorrect entry of propositions into LRmix Studio. This forced the software to account for most of the alleles as drop-in, resulting in $LR$s 60-70 orders of magnitude larger than expected. P₁, P₂, P₄, P₅, and P₈ can be manually excluded using peak heights. This has relevance when using LRmix which does not use peak heights.

We extend the work using the same two reference genotypes who were the true contributors as Semaan et al. (1). We simulate three two donor mixtures with peak heights using these two genotypes and analyse using STRmix™.

For the simulated 1:1 mixture, one of the non-donors’ $LR$s supported him being a contributor when no conditioning was used. When considered in combination with any other potential donors (i.e. with conditioning), this non-donor was correctly eliminated.

For the 3:1 mixture, all results correctly supported that the non-donors were not contributors. The low-template 4:1 mixture $LR$s with no conditioning showed support for all eight profiles as donors. However, the results from pair-wise conditioning showed that only the two ground truth donors had $LR$s supporting that they were contributors to the mixture.

We recommend the use of peak heights and conditioning profiles, as this allows better sensitivity and specificity even when the persons share many alleles.

**KEYWORDS**

Probabilistic genotyping, forensic DNA analysis, DNA mixtures, LRmix, STRmix™, relatives, exhaustive propositions
Semaan et al. (1) report a mock case with six non-donors falsely included in a two-person mixture.

The $L_R$s reported are incorrect.

We correct the LRmix results and provide the STRmix™ results.

All of the eight POI are related. This was not previously reported.

Using peak height information, or conditioning, all false donors are excluded.
In “A Mixed DNA Profile Controversy,” Semaan et al. (1) discuss the results of a mock case “where eight different individuals [two of whom were the source of the DNA] could not be excluded in a mixed DNA analysis. Even though relevant frequency datasets and an inbreeding coefficient were considered and expert DNA mixture analysis software was used.” This work was presented at the 21st Triannual meeting of the International Association of Forensic Sciences in 2017 (2) and was subsequently presented again at the 73rd American Academy of Forensic Sciences (AAFS) meeting in 2021 (3). In this work, Semaan et al. (1) produced a two-person mixture from two known contributors (P3 and P6 in Table 1) without dropout or drop-in.

Additional background information has been kindly provided by Semaan et al. (1) for this article. Semaan et al. (1) describe that “517 samples were collected randomly from [seven] Lebanese villages of different religious backgrounds.” The samples from the one village central to this paper (Shiaa2) were collected to try to find a matching donor for a blood transfusion. These samples total 21 individuals and come predominantly or completely from two pedigrees.

All pairs of two persons from the 517 donors were virtually mixed to form two-person mixtures. These mixtures were compared with all 517 individuals composed, in each case, of two true donors and 515 false donors. One of these mixtures was identified with eight potential matching donors (POI), including the two DNA profiles used to simulate the mixture. This is the mixture reported in Semaan et al. (1) and focused on in this rework of that paper. All eight of the POI come from the village Shiaa2. In Figure 1 we give our best effort to construct the pedigrees for these eight individuals (in red) and the other persons sampled from this village but not matching this mixture (in blue). This is developed from the names of the donors, the donors’ parents and the donors’ grandparents. Semaan et al. (1) have already eliminated the remaining 509 individuals in the database including those marked in blue in the pedigree.
Despite the fact that these individuals are far from randomly sampled, they represent an especially useful experiment. The two true donors are from each of the two proposed pedigrees. The resulting virtual mixture was compared with a large number of close relatives. This provides an opportunity to examine how best to deal with the risk of false support when examining multiple close relatives of true donors.

Using LRmix Studio v. 2.1.3, the LRs assigned in Semaan et al. (1) for the eight potential donors ranged between $10^{81}$ and $10^{94}$, which are unreasonably high. We would expect LRs up to about $10^{34}$ at the most. Given their knowledge about the structure of the Lebanese villages, Semaan et al. (1) concluded:

“Consequently, in similar populations, we recommend to restrain from establishing an inclusion interpretation in cases of mixed DNA traces, even when 28-locus profiles are used and statistical analysis is performed by expert software. However, such traces could be definitely used for exclusion purposes.”

This conclusion appears to follow the logic that the observation of false inclusions renders all results using this evidence type unusable. This should not be the case; however, it is valuable to be able to identify situations, such as the presence of related individuals, that may be at risk of elevated inclusionary support. Empirical work has previously been reported assessing the risk of false support to a non-donor who is related to the true donor(s) (see for example (4)).

We repeat the LRmix work of Semaan et al. (1) and extend it to include STRmix™ analyses. We took advantage of the fact that, even if high allelic overlap is known to cause difficulties during the deconvolution, research (5) has shown that with proper strategies these difficulties can be mitigated against.

**Method**

**DNA profile interpretation**

We use Lebanese allele frequency data from El Andari et al. (6) and $\theta = 0.01$ in all the following analyses. When there are multiple possible candidates that could be contributors to the mixtures, it has been shown (7) that better discrimination can be obtained when examining their DNA profiles in
combination. The ISFG commission (8) also advises to check if the mixture can be explained by multiple persons of interest together.

In this case, manual examination shows that the only pair that can collectively explain the mixture is P₃ and P₆. P₁ is manually excluded on simple allele presence.

**Experiment 1: Blinded manual interpretation**

Genotypes used are shown in Table 1. We provided a simulated electropherogram of the 1:1 mixture with no additional information and the reference profiles of P₁-P₈ to five volunteers. Interpretation theory states that we should consider the probability of the evidence given the proposed POI. Historically the human operator has made a decision based on this comparison that if this probability is close to zero then it is considered sufficiently close to impossible that the POI is a donor. This is termed an exclusion. The word “exclusion” represents a decision not the probabilistic comparison that precedes it. If the comparison suggests that the probability is greater than zero then the decision is that the POI is not excluded. This is termed an inclusion. Volunteers were asked to assess and record their interpretations of any inclusions/exclusions and the effect of any conditioning profiles. Table 2 provides information on further loci that were analysed but unused here. We refer the reader who is interested in a summary of existing guidance on when to condition to Hicks et al. (7).

**Experiment 2: LRmix (a semi-continuous model using only presence and absence of alleles)**

LRmix Studio version 2.5.1 was downloaded from GitHub (9). This program is no longer under development, having been superseded by fully continuous models. Training courses are no longer held but users are supported. The probabilities of dropout and drop-in were set to 0 as that is consistent with how we believe the mixed profile was simulated. The rare allele frequency was set at 0.001. We first used the alleles at the 23 loci presented in Semaan et al. Figure 1 (1) and then repeated the experiment using only the 21 GlobalFiler loci.

We use different sets of propositions, which lead to different LRs:
$H_u$: The DNA comes from $P_i$ and an unknown unrelated person $U$

$H_{uu}$: The DNA comes from two unknown people $U, U$, unrelated to $P_i$ or each other.

This leads to $LR_{iu/uu}$ for person $i$.

$H_{ij}$: The DNA comes from $P_i$ and $P_j$

$H_{ju}$: The DNA comes from $P_j$ and an unknown unrelated person $U$.

This leads to $LR_{ij/ju}$ for person $i$. More information can be found in (7, 10)

The DNA of a person that may be assumed to be present under both propositions is termed a conditioning profile. It is not a pre-requisite that the presence of the conditioning profile is certain or uncontested: it suffices that it is a reasonable possibility (see (7)). The conditioned $LR$s are of the form $LR_{ij/ju}$, and we assign these for the individuals with $LR_{iu/uu}$ larger than 1. All combinations of $i$ and $j$ (with $LR_{iu/uu}$ larger than 1) were trialled, not only those with the known donors as conditioning.

Conditioning has previously been shown to be highly valuable for all mixtures, especially those with high allelic overlap (5, 7, 10, 11) such as mixtures of close relatives.

**Experiment 3: STRmix™ (a continuous model using peak heights): moderate template, no dropout**

STRmix™ requires peak height information. We simulated the mixture from the genotypes of $P_3$ and $P_6$ using only the GlobalFiler loci for this experiment. We feel that restriction to one multiplex is more aligned with usual practice, and GlobalFiler is one of the most commonly used multiplexes.

For this experiment, we simulate two moderate template mixtures at ratios of 1:1 and 3:1, where $P_3$ is the major donor in the 3:1 mixture. A pseudocode description of the simulation is given in the supplementary material (File S1). The simulated mixture profiles appear in the supplementary material (Files S2 and S3). A mixture ratio of 3:1 is almost optimal in terms of demonstrating the benefit provided by peak
height information using a continuous interpretation model. Similarly, 1:1 mixtures demonstrate a considerable (but not complete) reduction in the benefit of including peak height since there is little distinction between the peak heights of each contributor. Forward and back stutters and stochastic effects were simulated using the variables from the PROVEDIt dataset (12).

We assign $LR_{aa/au}$ and $LR_{ij/ju}$ considering the same propositions as mentioned in Experiment 2.

A summary of the STRmix™ settings used in this experiment, which were previously determined for the PROVEDIt dataset, is given in the supplementary material (Table S1).

**Experiment 4: STRmix™ with low template and dropout**

We simulate a 4:1 mixture at low template (400:100 rfu, P3:P6). This profile has dropout of four alleles for the known minor donor. For person $i$, we assign $LR_{aa/au}$ and $LR_{ij/ju}$. The simulated mixture profile appears in the supplementary material (File S4).

**Results**

**Manual DNA profile comparisons (Experiment 1)**

In the blinded manual interpretation exercise for the 1:1 mixture, all volunteers successfully excluded donors P1, P2, P4, P5, and P8. A further three participants considered that two of the POIs could have both contributed to the DNA mixture and used conditioning. When conditioning profiles were used, P7 was found to be excluded, leaving only the ground truth donors P3 and P6 included.

**DNA profile comparisons using probabilistic genotyping systems (Experiments 2-4)**

The propositions considered and LR$\text{s}$ produced by LRmix and STRmix™ are given in Tables 3, 5 and 6.

Comparison with LRmix (Table 3):
The LR\textsubscript{s} calculated with LR\textsubscript{mix} by Semaan et al. (1) use the incorrect propositions. This occurred because of an error in entering propositions within the software. If the number of unknown persons is not changed from the default setting of 0, no unknown contributors will be considered. In Semaan et al. (1), the numerator proposition entered into LR\textsubscript{mix} was that a single POI, and no other donors, was the source of the DNA. All the peaks of the second donor thus need to be accounted for by drop-in. The alternative proposition considered that there were no donors, forcing LR\textsubscript{mix} to account for all of the alleles as drop-in.

In Table 4, we have assigned the inverse of the conditional profile probability (aka 1/RMP), which would be the LR we would expect (13, 14) when considering perfectly deconvoluted mixtures and sub-sub-source propositions (15, 16) such as POI is the major contributor or an unknown person is. We see that for high quality single source coming from those POIs, log(LR\textsubscript{s}) are not larger than 34.

For this study, we note that P\textsubscript{1} (who is not a donor) is excluded at D12S391 under the assumption of two donors unless drop-in is permitted. If an assumption of three donors is made under H\textsubscript{p} and two donors under H\textsubscript{a} we obtain $LR = 7.8 \times 10^{20}$. We do not, however, advise this, preferring the methods of (17)

When conditioned all \textit{LRs} support ground truth experiments seen in Table 3 row $LR_{ij/ju}$.

\textit{Comparison with STRmix\textsuperscript{TM}:}

The \textit{LRs} for STRmix\textsuperscript{TM} are shown in Tables 5 to 6. Except for one non-donor of the mixture 1:1 and the low-level mixtures, STRmix\textsuperscript{TM} provided \textit{LRs} that supported ground truth experiments. For the low-level DNA mixture, the non-donors (four of whom were the siblings of a true donor) have \textit{LRs} supporting they are donors, when this was not the case. It is important to note that when conditioned on any other POI, all \textit{LRs} support ground truth experiments. The assigned $LR_{ij/ju}$ for Experiment 4, the low-level 4:1 mixture with dropout, were all 0 except when considering the true donors P\textsubscript{3} and P\textsubscript{6}, who had an

$$LR_{3,6/6,U} = 2.32 \times 10^{26} \text{ and } LR_{3,6/3,U} = 4.88 \times 10^{12},$$ respectively. Note that the STRmix\textsuperscript{TM}
deconvolutions gave weights of 5-10% at many loci for genotypes with dropout for the minor donor, and some non-zero (but much smaller) weights at a few loci for the major donor (data not shown).

**Discussion**

The *LR*s calculated with LRmix by Semaan et al. (1) use the incorrect propositions. This is a useful reminder that 1) *LR*s depend on propositions that should be meaningful in the context of the case, and 2) one should scrutinize the output produced by the software. The warning about careful setting of the propositions has been given previously (16, 18). An *LR* for mixtures cannot be larger than 1/RMP. (There can be minor departures from this due to the exact details of the population genetic model employed.) In future development of LRmix or similar software, it would be beneficial to set error messages that would alert the user to such input errors.

The *LR*s produced by LRmix in this study are on the order of $10^{20}$, which is more within the expected range for true contributors to a two-person mixture.

**Peak height**

For laboratories using LRmix or other semi-continuous software, it is expected that mixtures will be inspected by an experienced analyst for exclusionary information based on peak heights. It is likely that the peak height information in the 3:1 mixture would have led to human exclusion for all comparisons giving *LR*s larger than one for non-donors. Manual exclusions were also possible for four of the five profiles giving false inclusionary support for the 1:1 mixtures. Presumably, expert exclusion would preclude the use of LRmix in the first place. The concept of the software expert pair (SEP) discussed in (19) would be useful when using a semi-continuous system such as LRmix. Inspection of results by an experienced analyst is recommended for any probabilistic genotyping system.

The effect of peak height can be seen in Table 5. $LR_{\text{true}}$ for the 3:1 mixture is greater than one only for the two true donors and again is of a magnitude more familiar. For the 1:1 mixture, unconditioned *LR*s
greater than one were produced for the true donors, P3 and P6, and also for P7, a non-contributor who is a sibling of P6. It is worthwhile to consider how a POI may be excluded from a 1:1 mixture when all his alleles are present. Consider a 1:1 mixture of two donors with genotypes ab and bc, or ac and bb. These are expected to produce peaks a, b, and c in the ratio 1:2:1. If a POI has a homozygote aa or cc genotype, then the expected peak height will differ from the observed and, if the template is moderate or high, then the POI will be excluded by a continuous system. The loci causing these exclusions are marked in bold in Table 1. P7 has no exclusionary loci and accordingly the unconditioned LR_uu/uu falsely supports P7 being a contributor to the mixture. At lower template (see Table 6), peak height is less informative. This results in no exclusions for unconditioned LR_uu/uu in Experiment 4.

**Considering an exhaustive set of propositions by conditioning on the profile of possible donors**

The use of a conditioning profile allows all meaningful propositions to be considered. LR_ij/ju in Tables 3 and 5 give the results of the use of conditioning for Experiments 2 and 3. For Experiment 4, the assigned values of LR_ij/ju were all 0 except for the combinations of P3 and P6, where LR_{3,6/6,1} = 2.32 \times 10^{26} and LR_{3,6/3,1} = 4.88 \times 10^{12}. All false donors are excluded when using conditioning, and all LRs therefore support the ground truth. Inspection of the STRmix™ results for LR_uu/uu in Table 6 suggests that conditioning on P3 may be especially helpful, as that LR is 11 to 15 orders of magnitude larger than for all other donors.

Conditioning is sufficiently beneficial that it should be undertaken whenever reasonable (5, 7, 10), as it is a powerful tool in avoiding adventitious inclusions. This can, and should be done, even if there is no specific proposition aligning with a conditioned LR, as the court might find value in knowing that two POI cannot be included in a mixture together. The value of conditioning is at its absolute maximum in unresolvable mixtures with high allelic overlap. There is a considerable amount of allelic overlap in the eight POI.
Accounting for relatedness in propositions

Modern probabilistic genotyping systems have the ability to calculate LRs based on alternate propositions such as: “the POI’s brother (unavailable for analysis) is the source of the DNA.” This should be used if the defence or the case information indicates that this is a meaningful proposition (20). However, generally there is no reliable knowledge of the possible alternate perpetrator. If there is concern that the population from which the perpetrator might have come contains relatives, or if a database containing relatives is searched, the use of alternate propositions that account for relatives might be appropriate. The recommended method here is to assign an exhaustive LR (21) accounting for the possibility that the alternate source might, or might not, be related.

The situation in the case considered in Semaan et al. (1) is even more complex. Most software deal with dyadic relationships, that is relationships between two individuals. This could be a POI and an untyped brother of the POI. The genotype of the POI informs the probability distribution for the genotypes of the untyped brother. Neither STRmix™ nor LRmix deal with triadic situations or higher, although DBLR™ does (22).

There seems to be a common perception that there is a 'system failure' if an LR < 1, based on unrelatedness, is not achieved when the ground truth is that the donor is in fact related to the POI. The most likely effect of considering relatedness is to lower the LR, but it will most probably remain above 1. The only way that we are aware of to produce an LR < 1 for a sibling with otherwise false support is to use conditioning profiles.

Conclusions

The main difference in LRs in this study compared to Semaan et al. (1) is due to the application of propositions in LRmix Studio. If the number of unknown contributors is not changed from its default setting of zero, none will be considered. Thus, propositions for a two-person mixture are that the DNA is only from the POI or is from no one. This forces the majority of the alleles to be accounted for by drop-in,
which explains the $L_R$s in the order of $10^{80}$ instead of $10^{20}$. A valuable message from this study is that the output of any software should be scrutinized. A useful check for mixtures is that $L_R$s cannot be higher than approximately $1/R_{MP}$.

Although the persons that were compared in Semaan et al. (1) share many alleles and comprise at most two sets of close relatives, we have shown strategies to achieve better specificity. In such cases, the use of conditioning and peak height information is key. When conditioning was employed, all $L_R$s supported the ground truth proposition (although this must not be taken as a proof that adventitious matches are impossible).

$LR_{mix}$ is a semi-continuous software, meaning it does not use peak height information. It was an early and effective model designed primarily to deal with dropout. It has a pedigree back to 2000 (23, 24), is robust and widely used, and operates as advertised. Development has now ceased for $LR_{mix}$ (25).

Neither $LR_{mix}$ nor any other software or interpretation method can claim that the rate of false support is zero. This is not due to the software, but to the nature of DNA evidence itself: there will always be uncertainty about the source of the DNA, as we cannot know who left the DNA trace. A false support rate of zero cannot be achieved by any method, nor is it necessary: this explains why DNA (or any evidence) should not be solely relied upon to reach a conclusion, but instead must be considered in combination with the other elements of the case.

Allelic overlap and populations of closely related persons can present difficulties for profile deconvolution. However, these challenges are not insurmountable, even for close relatives. When peak height is informative or if profiles of known persons who could have contributed to the DNA are available, the rate of $LR > 1$ for false donors can be greatly reduced (7, 10). For simpler mixtures this reduction in false inclusions can be realized by applying quantitative genotyping by the expert, even if the probabilistic system does not use peak height. When there is a possibility that a close relative is the
source of the DNA, it is very desirable to obtain a sample from first order relatives and genotype them. They will either be eliminated or, if not, one will be alerted to the possibility of false support.

References


Table 1. The alleles detected in the mixed profile, and the genotypes of the eight potential candidates. Genotypes causing exclusions for the 1:1 moderate template mixture using STRmix™ are marked in bold.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Mixed profile</th>
<th>P₁</th>
<th>P₂</th>
<th>P₃</th>
<th>P₄</th>
<th>P₅</th>
<th>P₆</th>
<th>P₇</th>
<th>P₈</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>15,17,18</td>
<td>17,18</td>
<td>15,17</td>
<td>17,18</td>
<td>15,18</td>
<td>15,17</td>
<td>15,17</td>
<td>15,17</td>
<td>15,17</td>
</tr>
<tr>
<td>D19S433</td>
<td>13,16</td>
<td>13,13</td>
<td>13,16</td>
<td>13,16</td>
<td>13,16</td>
<td>13,16</td>
<td>13,16</td>
<td>13,16</td>
<td>13,16</td>
</tr>
<tr>
<td>D8S1179</td>
<td>8,12,13,15</td>
<td>12,13</td>
<td>12,15</td>
<td>12,13</td>
<td>8,12</td>
<td>12,15</td>
<td>8,15</td>
<td>8,13</td>
<td>12,15</td>
</tr>
<tr>
<td>D5S818</td>
<td>11,13,14</td>
<td>11,14</td>
<td>11,13</td>
<td>13,14</td>
<td>11,13</td>
<td>11,13</td>
<td>11,11</td>
<td>11,14</td>
<td>11,14</td>
</tr>
<tr>
<td>TH01</td>
<td>7,9,10</td>
<td>9,9</td>
<td>9,10</td>
<td>7,10</td>
<td>9,10</td>
<td>9,9</td>
<td>9,10</td>
<td>9,10</td>
<td>9,10</td>
</tr>
<tr>
<td>vWA</td>
<td>14,20</td>
<td>20,20</td>
<td>14,20</td>
<td>14,20</td>
<td>14,20</td>
<td>14,20</td>
<td>20,20</td>
<td>20,20</td>
<td>14,14</td>
</tr>
<tr>
<td>D21S11</td>
<td>29,30,30.2</td>
<td>29,30</td>
<td>29,30</td>
<td>30,30.2</td>
<td>30,30</td>
<td>29,30</td>
<td>29,30</td>
<td>29,30</td>
<td>29,30</td>
</tr>
<tr>
<td>D13S317</td>
<td>10,12</td>
<td>10,12</td>
<td>12,12</td>
<td>10,12</td>
<td>10,12</td>
<td>12,12</td>
<td>12,12</td>
<td>12,12</td>
<td>10,12</td>
</tr>
<tr>
<td>TPOX</td>
<td>8,10,11</td>
<td>10,11</td>
<td>8,10</td>
<td>8,11</td>
<td>8,10</td>
<td>10,10</td>
<td>8,10</td>
<td>8,11</td>
<td>8,10</td>
</tr>
<tr>
<td>FGA</td>
<td>22,23</td>
<td>22,23</td>
<td>22,23</td>
<td>23,23</td>
<td>22,23</td>
<td>22,23</td>
<td>22,23</td>
<td>22,23</td>
<td>22,23</td>
</tr>
<tr>
<td>D7S820</td>
<td>9,10,12</td>
<td>10,12</td>
<td>10,10</td>
<td>9,10</td>
<td>9,10</td>
<td>9,12</td>
<td>10,12</td>
<td>10,12</td>
<td>9,10</td>
</tr>
<tr>
<td>D16S539</td>
<td>11,12</td>
<td>11,11</td>
<td>11,11</td>
<td>11,12</td>
<td>11,11</td>
<td>11,11</td>
<td>11,11</td>
<td>11,11</td>
<td>11,11</td>
</tr>
<tr>
<td>D18S51</td>
<td>12,19</td>
<td>12,12</td>
<td>12,19</td>
<td>12,12</td>
<td>12,12</td>
<td>12,19</td>
<td>12,19</td>
<td>12,19</td>
<td>12,19</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>11,12</td>
<td>11,11</td>
<td>12,12</td>
<td>11,12</td>
<td>11,12</td>
<td>11,12</td>
<td>11,12</td>
<td>11,11</td>
<td>11,12</td>
</tr>
<tr>
<td>Locus</td>
<td>18,19,20</td>
<td>19,20</td>
<td>18,19</td>
<td>20,20</td>
<td>19,20</td>
<td>19,20</td>
<td>19,20</td>
<td>19,20</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>D2S1338</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penta E</td>
<td>5,12</td>
<td>5,12</td>
<td>5,12</td>
<td>5,5</td>
<td>5,12</td>
<td>5,12</td>
<td>5,12</td>
<td>5,12</td>
<td></td>
</tr>
<tr>
<td>Penta D</td>
<td>8,11,12</td>
<td>8,11</td>
<td>11,11</td>
<td>8,11</td>
<td>8,11</td>
<td>11,12</td>
<td>11,12</td>
<td>8,11</td>
<td></td>
</tr>
<tr>
<td>SE33</td>
<td>19,25.2</td>
<td>28.2</td>
<td>29.2</td>
<td>25.2</td>
<td>25.2</td>
<td>25.2</td>
<td>25.2</td>
<td>25.2</td>
<td></td>
</tr>
<tr>
<td>D22S1045</td>
<td>11,15</td>
<td>11,15</td>
<td>11,15</td>
<td>11,15</td>
<td>11,15</td>
<td>11,15</td>
<td>11,15</td>
<td>11,15</td>
<td></td>
</tr>
<tr>
<td>D1S1656</td>
<td>12,16,17</td>
<td>12,17</td>
<td>16,17</td>
<td>12,17</td>
<td>16,17</td>
<td>12,17</td>
<td>12,17</td>
<td>12,17</td>
<td></td>
</tr>
<tr>
<td>D10S1248</td>
<td>12,14</td>
<td>12,12</td>
<td>12,14</td>
<td>14,14</td>
<td>12,14</td>
<td>12,14</td>
<td>12,14</td>
<td>12,12</td>
<td></td>
</tr>
<tr>
<td>D2S441</td>
<td>11,14</td>
<td>11,14</td>
<td>11,14</td>
<td>14,14</td>
<td>11,14</td>
<td>11,14</td>
<td>11,14</td>
<td>11,14</td>
<td></td>
</tr>
<tr>
<td>D12S391</td>
<td>19,21,22,23</td>
<td>19,19</td>
<td>19,22</td>
<td>22,23</td>
<td>19,22</td>
<td>21,22</td>
<td>19,21</td>
<td>19,21</td>
<td>19,22</td>
</tr>
</tbody>
</table>
Table 2: Additional loci that are available but not used in the studies reported here.

<table>
<thead>
<tr>
<th>Locus</th>
<th>P₁</th>
<th>P₂</th>
<th>P₃</th>
<th>P₄</th>
<th>P₅</th>
<th>P₆</th>
<th>P₇</th>
<th>P₈</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPL</td>
<td>9,10</td>
<td>10,10</td>
<td>9,9</td>
<td>9,10</td>
<td>9,9</td>
<td>10,10</td>
<td>9,10</td>
<td>10,10</td>
</tr>
<tr>
<td>F13B</td>
<td>8,9</td>
<td>8,8</td>
<td>7,10</td>
<td>8,9</td>
<td>8,8</td>
<td>8,10</td>
<td>9,10</td>
<td>8,9</td>
</tr>
<tr>
<td>FESFPS</td>
<td>10,12</td>
<td>10,12</td>
<td>11,11</td>
<td>10,11</td>
<td>10,12</td>
<td>10,12</td>
<td>10,12</td>
<td>11,12</td>
</tr>
<tr>
<td>F13A01</td>
<td>6,6</td>
<td>5,6</td>
<td>5,5</td>
<td>5,6</td>
<td>5,12</td>
<td>6,6</td>
<td>6,6</td>
<td>6,6</td>
</tr>
<tr>
<td>Penta C</td>
<td>11,14</td>
<td>11,11</td>
<td>11,11</td>
<td>11,14</td>
<td>11,14</td>
<td>11,11</td>
<td>11,11</td>
<td>11,11</td>
</tr>
</tbody>
</table>
Table 3: LRs assigned using \( LR_{\text{mix}} \) (23 loci and GlobalFiler 21 loci) as reported in SAM considering that the two-person mixture is from the POI only versus no one, and the LRs obtained here using \( LR_{\text{mix}} \) are \( LR_{iu/uu} \) and \( LR_{ij/ju} \).

<table>
<thead>
<tr>
<th></th>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( P_3 )</th>
<th>( P_4 )</th>
<th>( P_5 )</th>
<th>( P_6 )</th>
<th>( P_7 )</th>
<th>( P_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported LR in SAM (1) (23 loci)</td>
<td>1.3 ( \times 10^{81} )</td>
<td>9.9 ( \times 10^{85} )</td>
<td>1.8 ( \times 10^{84} )</td>
<td>4.0 ( \times 10^{84} )</td>
<td>2.3 ( \times 10^{89} )</td>
<td>2.0 ( \times 10^{90} )</td>
<td>9.5 ( \times 10^{90} )</td>
<td>2.3 ( \times 10^{89} )</td>
</tr>
<tr>
<td>( LR_{iu/uu} ) obtained in this work using ( LR_{\text{mix}} ) (23 loci)</td>
<td>0 ( \times 10^{18} )</td>
<td>2.0 ( \times 10^{21} )</td>
<td>1.9 ( \times 10^{21} )</td>
<td>1.3 ( \times 10^{21} )</td>
<td>2.1 ( \times 10^{17} )</td>
<td>1.7 ( \times 10^{19} )</td>
<td>5.5 ( \times 10^{19} )</td>
<td>1.4 ( \times 10^{21} )</td>
</tr>
<tr>
<td>( LR_{ij/ju} ) obtained in this work using ( LR_{\text{mix}} ) (21 loci)</td>
<td>0 ( \times 10^{16} )</td>
<td>4.7 ( \times 10^{18} )</td>
<td>4.8 ( \times 10^{18} )</td>
<td>4.1 ( \times 10^{18} )</td>
<td>2.0 ( \times 10^{15} )</td>
<td>1.7 ( \times 10^{17} )</td>
<td>5.2 ( \times 10^{17} )</td>
<td>3.6 ( \times 10^{18} )</td>
</tr>
</tbody>
</table>
Table 4. 1/RMP values for the genotype of the POI and $\theta = 0.01$. Only $P_i$ was used in conditioning. This treats the POI as a single source, and these values are not appropriate for interpretation of a mixture. They do, however, give an approximate upper bound for the LR for a mixture with a single additional unknown in the denominator compared with the numerator. This is approximate because the alleles assumed to have been observed from the sub-population may differ. When the LRs for mixtures are significantly higher than 1/RMP, this is an indication that there is an issue with the calculation. RMP was calculated as \( \frac{2\theta + (1-\theta) p_i}{(1+\theta)(1+2\theta)} \) for a homozygote of genotype \( a_i a_i \) and \( \frac{2(\theta + (1-\theta) p_i)(\theta + (1-\theta) p_j)}{(1+\theta)(1+2\theta)} \) for a heterozygote of genotype \( a_i a_j \). There is good logic to include all of $P_1$ to $P_8$ in the conditioning in this case because they all come from the same subpopulation that is also supposed to be the subpopulation of the offender. We have done this experimentally but not reported it here for brevity. As expected, these unreported LRs are lower.

<table>
<thead>
<tr>
<th>RMP for POI expressed as 1 in $P_i$</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
<th>$P_5$</th>
<th>$P_6$</th>
<th>$P_7$</th>
<th>$P_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.36 $\times 10^{33}$</td>
<td>6.16 $\times 10^{30}$</td>
<td>3.06 $\times 10^{32}$</td>
<td>2.38 $\times 10^{34}$</td>
<td>4.98 $\times 10^{29}$</td>
<td>9.74 $\times 10^{30}$</td>
<td>2.52 $\times 10^{31}$</td>
<td>5.74 $\times 10^{33}$</td>
</tr>
</tbody>
</table>
Table 5: LRs assigned using STRmix™ (GlobalFiler 21 loci) using allelic frequencies from El Andari et al. (6) and $\theta = 0.01$. The conditioned LRs were assigned for person $P_i$, considering the results given the proposition that either “The DNA comes from $P_i$ and $P_j$ or from $P_i$ and an unknown unrelated person” and the proposition that “The DNA comes from $P_j$ and an unknown unrelated person or two unknown unrelated people”. Only $P_3$ and $P_6$ explain the mixture together. $P_7$ (a sibling of $P_6$) tested with either $P_3$ or $P_6$ gives $LR = 0$. When conditioned on non-donors, $LR_{ij/uu} = 0$, even for the true donors (e.g., $LR_{3,7/7, U}$).

| $LR_{iu/uu}$ | STRmix™ (GlobalFiler 21 loci) | 1:1 | 0 | 0 | 1.65×10^{19} | 0 | 0 | 2.84×10^{18} | 2.86×10^{18} | 0 |
| 3:1 | 0 | 0 | 6.40×10^{27} | 0 | 0 | 1.07×10^{27} | 0 | 0 |
| $LR_{ij/uu}$ | STRmix™ (GlobalFiler 21 loci) | 1:1 | 0 | 0 | 1.43 × 10^{28} | $LR_{3,6/6, U}$ | 0 | 0 | $LR_{3,6/3, U}$ | 0 | 0 |
| 3:1 | 0 | 0 | 1.53 × 10^{28} | $LR_{3,6/6, U}$ | 0 | 0 | $LR_{3,6/3, U}$ | 0 | 0 |
Table 6. Unconditioned *LR*\(_s\) for Experiment 4 (4:1 low-level mixture) using STRmix\(^\text{TM}\) and data from El Andari et al. (6) and \(\theta = 0.01\). All eight references give values that support the first proposition compared to the alternative when no conditioning profiles are used.

<table>
<thead>
<tr>
<th>i</th>
<th>(P_1)</th>
<th>(P_2)</th>
<th>(P_3)</th>
<th>(P_4)</th>
<th>(P_5)</th>
<th>(P_6)</th>
<th>(P_7)</th>
<th>(P_8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(LR_{iu/uu})</td>
<td>(1.46 \times 10^{10})</td>
<td>(9.14 \times 10^{11})</td>
<td>(5.86 \times 10^{23})</td>
<td>(1.58 \times 10^{11})</td>
<td>(7.40 \times 10^{8})</td>
<td>(4.61 \times 10^{11})</td>
<td>(1.80 \times 10^{10})</td>
<td>(4.07 \times 10^{12})</td>
</tr>
</tbody>
</table>
FIG. 1. The proposed pedigrees (a) and (b) for those individuals sampled from the village Shiaa 2 that we have been able to place on the pedigree. Green symbols represent the two POI who are the true donors. Red symbols represent the six POI who are non-donors. Blue symbols signify those individuals that are not one of the POI but which were sampled and we think are part of the pedigree. As usual males are square, females are circles and unisex are diamonds. The numbers are a code to the identifiers.