HealthSupport
Queensland
Forensic and Scientific Services

Basics of DNA Profile Interpretation

1 Purpose

This procedure briefly describes the models behind the statistical analysis of DNA profiles within Forensic DNA Analysis. It also provides the concepts necessary to perform intuitive checking of STRmix analyses.

2 Scope

This procedure applies to all case managing and reporting scientists in Forensic DNA Analysis.

3 Definitions

PowerPlex®21 (PP21):	The amplification system in use within Forensic DNA Analysis consisting of 20 loci plus amelogenin
STRmix™:	A statistical analysis package used in Forensic DNA Analysis to perform deconvolutions and statistical calculations on single source and mixed DNA profiles
POI:	Person of interest.
LR:	Likelihood ratio

4 Principle of statistical interpretation

When a crime scene profile and a reference profile are found to match, a statistical analysis is performed to provide a weighting to this evidence. These statistical analyses are reported as a Likelihood Ratio (LR).

A Likelihood Ratio is the ratio of 'the probability of the DNA profile occurring if the DNA had come from the POI' to 'the probability of the DNA profile occurring if the DNA had come from someone other than, and unrelated to, the POI'.

The basis of all statistical calculations in DNA profile interpretation is the Product Rule. This allows the allele frequencies at a locus to be multiplied together to give the genotype frequency and the genotype frequencies at each locus to be multiplied together to give the combined genotype frequency.

The Product Rule model assumes that the population exists in Hardy-Weinberg Equilibrium (HWE) and Linkage Equilibrium (LE).

More information about the population genetic model can be found in Chapter 3 of Buckleton, Bright and Taylor, 2016 [10].

5 Theta (θ)

The theta value is also known as the co-ancestry co-efficient; it may also be referred to as F_{ST} (inbreeding co-efficient). Theta is used to account for population substructure or common ancestry between people in a subpopulation.

Forensic DNA Analysis uses the Balding-Nichols formulae [1] as described in NRC Recommendation 4.2 Equation 4.10 [2]:

Formula 4.10a

$$\Pr(AA|AA) = \frac{[2\theta + (1-\theta)p_A][3\theta + (1-\theta)p_A]}{(1+\theta)(1+2\theta)}$$

Formula 4.10b

$$\Pr(AB|AB) = \frac{2[\theta + (1-\theta)p_A][\theta + (1-\theta)p_B]}{(1+\theta)(1+2\theta)}$$

Theta is modelled by STRmix using beta distributions [3]. Theta is locus dependant and so different kits (with different loci) will have different theta distributions [4]. The theta distributions for PP21 are shown in Table 1.

Table 1 Theta distributions

	Caucasian	Aboriginal	Asian
PowerPlex 21	beta(1.5,242.3)	beta(1.5,80.7)	beta(1.5,83.2)

6 Population datasets

6.1 Routine population datasets

Population datasets are used to calculate an estimate of allele frequencies in the population of interest which are required for the calculation of the LR.

Australian datasets have been created for the PP21 loci using data obtained from jurisdictions in Australia and these datasets have been validated for use [5].

Forensic DNA Analysis routinely uses the Australian Caucasian, Australian Asian and Australian Aboriginal datasets, the sizes of which are detailed in Table 2.

Note: The Australian Aboriginal dataset includes Torres Strait Islander samples from Queensland.

Table 2 PP21 datasets

Racial group	Number of Alleles	Number of individuals (N)	Population proportion in Queensland
Australian Caucasian	3414	1707	87%
Australian Aboriginal	3556	1778	3%
Australian Asian	1980	990	10%

The policy of the laboratory is to report the LR that is stratified across all three datasets using the proportions detailed in Table 2 (obtained from 2011 Census data). This calculation is performed by STRmix and the stratified LR is provided in the STRmix report.

6.2 Use of alternative population datasets

An additional dataset is available to Forensic DNA Analysis (Table { SEQ Table * ARABIC }) but is not routinely used.

Table { SEQ Table * ARABIC } PP21 NT Pure Aboriginal dataset

Racial group	Number of Alleles	Number of individuals (N)
Australian Northern Territory Pure Aboriginal	1092	546

For the situations where the population required is not available within the datasets held by this laboratory, available datasets can be researched from published forensic journals or from other forensic laboratories.

A substantive Team Leader may give permission to use population data from other jurisdictions or sources provided it is validated and acceptable by NATA accreditation requirements. The use of combined or pooled population data from other jurisdictional laboratories is permissible provided the populations are compared for "goodness of fit" and are shown not to be statistically dissimilar.

In the event of this occurrence, a substantive Team Leader shall obtain a statistician's advice on the comparison of population data.

7 Sampling uncertainty

When providing a statistical weighting, the figure quoted can never be an exact answer but only ever a 'best estimate' of the true result. This is due, in part, to the fact that only a small proportion of the population is represented in the dataset. If a different dataset were to be produced, the LR would differ.

There are a number of different ways to account for sample variation; Forensic DNA Analysis uses the Highest Posterior Density (HPD) method using 1000 HPD iterations with a 99% one-sided significance value.

8 Basic principles of DNA profile interpretation

8.1 Background

A DNA profile provides both qualitative information (which alleles are present) and quantitative information (peak heights). The peak heights in a DNA profile are proportional to the amount of amplified DNA [6] and can be used to indicate the relative proportions of DNA of the contributors to a DNA profile.

When DNA is extracted from a sample that contains DNA from more than one individual, the result is often a mixed DNA profile. This consists essentially of one person's DNA profile superimposed on that of another (Figure 1).

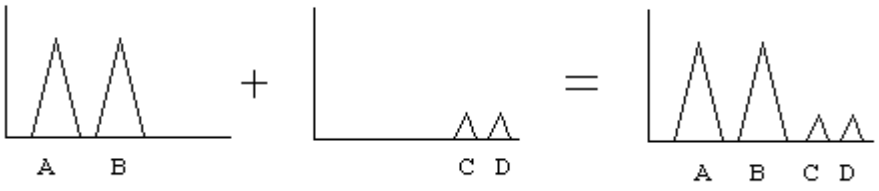


Figure 1 Mixed DNA profile

The analysis and interpretation of a mixed DNA profile follows a series of logical steps. It should be noted that interpretation of DNA profiles must be carried out before comparison with the reference profiles (with the exception of conditioning, see Section 8.10).

8.2 Single source DNA profiles

DNA profiles that have only one contributor are termed single source and are characterised by the presence of no more than two alleles at each locus (with the exception of mutations, see Section 0).

Testing within Forensic DNA Analysis has shown that where a single source DNA profile consists of 32 or more alleles the LR will always be greater than 100 billion regardless of the dataset used [7][8].

STRmix is used to calculate a LR for single source DNA profiles consisting of less than 32 alleles.

This LR is expressed as:

$$LR = \frac{\text{Probability of observing the DNA profile if the DNA had originated from the POI}}{\text{Probability of observing the DNA profile if the DNA had originated from someone other than, and unrelated to, the POI}}$$

Further information on the formulation of LRs can be found in Section 10.

8.3 Relatives values

On occasions in forensic casework, there may be a need to consider the alternative proposition that the DNA originated from a relative of the POI, such as a sibling, uncle/aunt, nephew/niece, or cousin.

In the first instance, a reference sample from that relative should be requested to enable a direct comparison to the crime scene profile. In the absence of a reference sample, a LR may be provided.

When providing a LR, the scientist should take care to ensure that the correct propositions are being addressed and be aware of transposing the conditional (see Section 11).

$$LR = \frac{\text{Probability of observing the DNA profile if the DNA had originated from the POI}}{\text{Probability of observing the DNA profile if the DNA had originated from a } \textbf{\textit{[relative]}} \text{ of the POI}}$$

The STRmix output provides LR values for the following relatives:

- Sibling
- Parent/Child
- Half Sibling
- Grandparent/Grandchild
- Uncle or Aunt/Niece or Nephew
- Cousin

8.4 Step 1: Identify the presence of a mixture

Mixed DNA profiles can be identified by the presence of additional peaks and/or the presence of allelic imbalance.

8.4.1 Additional Peaks

A mixed DNA profile is indicated by the presence of three (or more) alleles at any locus.

- It is considered unlikely that a two-person mixture would **not** be detected on the account of the lack of extra peaks in the whole profile.

- However, the propensity for this situation to occur increases when partial profiles are encountered, and / or when related persons are contributing to a mixed stain.

The presence of additional alleles at any particular locus is not necessarily diagnostic of a mixture because other circumstances can lead to extra peaks, for example pull-up, stutters and mutations, giving the (wrong) impression of a mixed DNA profile.

8.4.2 Allelic Imbalance

It is possible, especially in cases involving partial profiles and/or related people, for a mixed profile to exhibit no additional peaks at all. Therefore, the absence of additional peaks but the presence of asymmetric peaks could indicate a mixture. However, in case work samples, a significant peak asymmetry could be due to:

- differential amplification of the alleles.

  - It should be noted that in single source profiles the observed symmetry between peak heights is not exact, and that after amplification of reference DNA there is a degree of natural variation between the peak areas at a heterozygous locus.

- primer binding site mutations.

  - Primers are designed to attach to a particular base sequence that surrounds a target STR. These sequences are generally chosen for their relative uniformity across populations. A Primer Binding Site Mutation is a change in the base sequence (insertion or deletion) at a point where a primer has been designed to attach. This can cause inefficient primer binding to the target base sequence. The annealing reaction still occurs but to a lesser extent than normal and a radical imbalance in allelic height is observed for that particular locus. Alternatively, the mutation can totally prevent primer binding, and as such no there is no amplification product which can result in allelic dropout.

8.5 Step 2: Identify the number of contributors to a mixture

Following the assessment conducted in Step 1, if a mixed DNA profile is the most likely scenario then the possible number of contributors must be assessed.

- The number of extra alleles at each locus, and their relative proportions are useful indicators. The maximum number of alleles that would be detected at a locus for a mixture of two heterozygotes is four.
- Peaks that are in stutter position but above the stutter threshold may indicate an additional contributor.
- Peaks between LOD and LOR should be considered (see Section 9).
- Five or six alleles at a locus are indicative of three or more contributors.
- More information about determining the number of contributors to a DNA profile can be found in the Appendix (Section 16.1)

For the purposes of demonstrating the steps required to interpret a DNA profile, mixtures from two contributors will be used. The same principles can be applied to higher order mixtures.

8.6 Step 3: Determine the approximate 'ratio' of the components in the mixture

This step is completed by STRmix in the deconvolution process; however, it is a requirement that the scientist performs an intuitive check of the STRmix output.

There are two ways in which the contributors to a profile can be represented:

1. The mixture proportion (M_x)
2. The mixture ratio (M_R)

Note: The mixture proportion and mixture ratio are not one and the same value, and therefore have different definitions.

The approximate value of M_x for a four-allele locus under an assumption of two contributors, where two minor component alleles a and b are present with two major component alleles c and d is:

$$M_x = \frac{\Phi_a + \Phi_b}{\Phi_a + \Phi_b + \Phi_c + \Phi_d}$$

where Φ_a is the *peak height* of the *a* allele.

The approximate value of M_R under the same conditions is:

$$M_R = \frac{\Phi_a + \Phi_b}{\Phi_c + \Phi_d}$$

Once the M_x parameter has been determined, the mixture ratio can be calculated using the equation below:

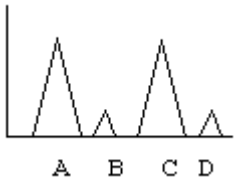
$$M_R = \frac{M_X}{1 - M_X}$$

Similarly, the M_x parameter can be calculated:

$$M_X = \frac{M_R}{1 + M_R}$$

A mixture can range from the two contributors being present in equal proportions to one contributor being greatly in excess. A contributor with a small contribution to the DNA profile (minor contributor) may reach a similar level to stutter and artefacts, and some peaks may be masked or have dropped out.

When there are no shared alleles at one or more of the loci (because both of the genotypes are heterozygous for different alleles), determining the approximate ratio of the mixture is a relatively straightforward task (Figure 2).



$$M_R = \frac{A+C}{B+D}$$

Figure 2

Determining the ratio when there are shared alleles is more complex because there may be more than one combination of alleles which could explain the observation. When there are shared alleles at a locus then a good understanding of the characteristics of heterozygous imbalance (H_b) are required.

- it is still possible to calculate a mixture ratio because the allele peak areas are additive for any shared alleles

- at any locus where there is a single shared allele, if the two DNA templates were admixed in a 1:1 ratio, then the peak area of the shared allele should be about twice those of the unshared alleles

An example of masking – three alleles at a locus

If the genotypes of two persons are *ab* and *bc*, then they share the *b* allele.

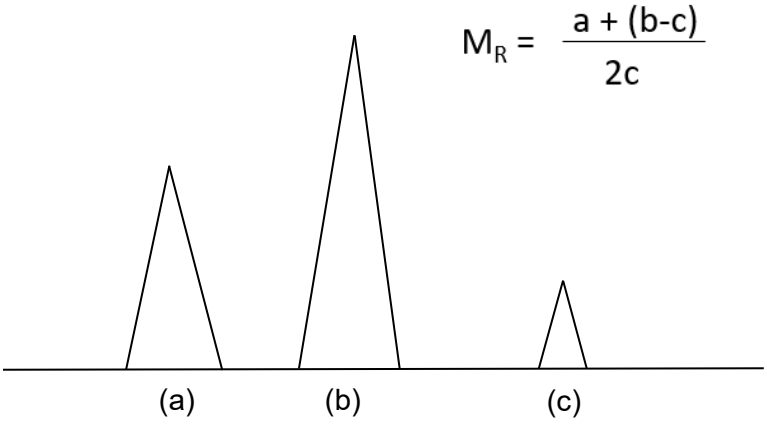


Figure 3 A three allele locus showing masking

It is important to remember though, that in casework profiles, the two peaks of a heterozygous locus will not necessarily be the same height. Therefore, when an allele is shared it may appear that there is not enough of that allele to share between the other two peaks, however when allelic imbalance is considered this shared allele may be able to be split.

Amelogenin is useful to assess the type of mixture present and in mixed sex samples can give a rough guide as to whether the male or female is contributing most of the DNA. However, it would seem that amelogenin is often not representative of the mixture as a whole and is therefore the locus least suitable for determining the mixture ratio.

8.7 Step 4: Determine the possible pair wise combinations for the components of the mixture

Do Not: At this point, do not determine whether or not the observed mixed profile is consistent with a mixture of DNA from the persons concerned by comparing the reference samples. Comparison to reference samples before Step 5 is not endorsed and it is recommended that interpretation of the mixed profile is conducted independently of knowledge of the results of reference samples.

This step is completed by STRmix in the deconvolution process; however, it is a requirement that the scientist performs an intuitive check of the STRmix output.

List all pair-wise combinations for allelic pairs at each locus.

- With four alleles at a locus, and allowing for reciprocal combinations, three pair wise combinations of alleles exist.
- With three alleles at a locus, there are six pair wise combinations and with two alleles, given that a full profile has been obtained for both components of the mixture, there are four possible combinations.
- With one allele, the only possibility is that both individuals are homozygous for the same allele.
- These combinations are listed in Table { SEQ Table * ARABIC }.

Table { SEQ Table * ARABIC } Pair wise combinations

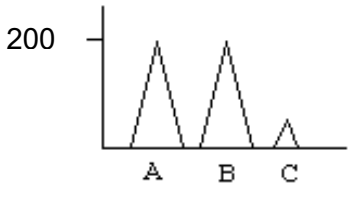
Four alleles (A,B,C,D)		Three alleles (A,B,C)		Two alleles (A,B)	
A, B	C, D	A, A	B, C	A, A	A, B
A, C	B, D	B, B	A, C	A, B	A, B
A, D	B, C	C, C	A, B	A, A	B, B
C, D	**A, B**	A, B	A, C	A, B	B, B
B, D	**A, C**	B, C	A, C	**A, B**	**A, A**
B, C	**A, D**	A, B	B, C	**B, B**	**A, A**
		B, C	**A, A**	**B, B**	**A, B**
		A, C	**B, B**		
		A, B	**C, C**		
		A, C	**A, B**		
		A, C	**B, C**		
		B, C	**A, B**		

Using the quantitative information drawn from the peak heights in the profile and the approximate ratio of the mixture, some of the pair wise possibilities can then be discounted based on AI and mixture ratios.

- This is straightforward where there are high ratio mixtures and the major component is the profile of interest as it invariably results in one combination per locus. This is because the peak areas will be such that there is an easily distinguishable allele pair for the main component and therefore no ambiguity in the resultant profile.

- When it is the minor component which is of interest, the reverse is usually true and the resultant profile is seldom unambiguous. This is a direct result of minor peaks being masked by major peaks and difficulty in deciding whether or not a small peak is a true allele or has some other explanation.

When considering genotype combinations, one must also consider the possibility of allelic drop-out. For example, with the three peaks seen below (Figure 4), the smallest peak is in the stochastic range, in this instance we must consider that its partner could have dropped out leaving the possible allele combination as A, B & C, - where "-" indicates that the partner to C could be any allele at that locus.



Combination A,B → C,-

Figure 4

8.8 Step 5: Compare the resultant profiles for the possible components of the mixture with those from the reference samples

If the profile from the reference sample of the POI matches one or other of the alternatives, then that person cannot be eliminated as a possible contributor to the mixed DNA profile.

It follows that the more ambiguity in the predicted profile for the component of interest, the lower will be the evidential significance of any subsequent match. Consequently, when the

minor component of a high-ratio mixture is under consideration, the evidential significance is often lower as a result.

Appendix Section 16.6 provides a guide for excluding reference samples from a DNA profile.

8.9 Factors which can affect mixture interpretation

The presence of degraded DNA can affect mixture interpretation.

- The higher molecular weight loci are the first affected.
- The electropherogram exhibits a gradual loss of signal.
- Depending on the extent of the degradation, the result can be a full profile with reduced signal at the higher molecular weight end or a partial profile at one or more loci or, in the extreme, no profile at all.

With a mixture it is possible that degradation of the DNA could affect the contributors unequally.

It is possible for one component of the mixture to exhibit degradation towards the higher molecular weight loci whilst the other component to the mixture does not exhibit degradation at all. If this does occur, then there is the possibility that the ratio of the profile will appear to 'flip' towards the higher molecular weight loci.

8.10 Conditioning

There are some items where one may assume that the presence of DNA from a particular person could be present. For example, obtaining a single source DNA profile from an intimate swab that matches the donor of the swab is not an unexpected result. In this instance, applying a statistical weighting to this match would not be of evidential value and is therefore not performed.

When a mixed DNA profile is obtained from an intimate swab, it would not be unreasonable to consider that the donor of the swab could be one of the contributors. In this instance, the presence of DNA from the donor could be factored into the interpretation of the mixed DNA profile in a process called 'conditioning'.

In Forensic DNA Analysis conditioning is routinely applied to samples taken from a person, including vaginal swabs, fingernail scrapings and swabs from an area of the body. Conditioning is also routinely applied to items of clothing that are known to have been worn by a particular person, but not to items of clothing that have been recovered from a scene such as a bedroom floor.

Other examples may be specific to case circumstances and require the conditioning to be acceptable under both the prosecution and defence alternatives.

Conditioning is performed by STRmix during the deconvolution process.

8.11 Complex DNA profiles

A DNA profile may be deemed unsuitable for interpretation based on the following:

- Poor quality of the profile
- Limited information within the profile
- Degradation
- An inability to reliably assess the number of contributors to the profile
- Too many contributors to the profile such that any reference sample comparison to the DNA profile would be meaningless

The STRmix interpretation software requires that the scientist determines the number of contributors that best supports the DNA profile. If the number of contributors cannot be reliably assessed, then there is currently no method available for interpretation. In this instance the DNA profile is deemed unsuitable for meaningful interpretation, meaning that it is not possible to apply any statistical weighting to any findings in relation to this profile.

Forensic DNA Analysis has validated the use of STRmix for the interpretation of DNA profiles with 1-4 contributors. A DNA profile that displays evidence of more than four contributors is deemed too complex for interpretation.

9 Sub-threshold peaks

Sub-threshold peaks are defined as peaks that fall below the limit of reporting (LOR) and above the limit of detection (LOD).

Sub-threshold peaks can be used to inform the number of contributors and to exclude the donor of a reference sample, however they do not form part of the statistical interpretation of the profile and are not included in the LR.

10 Formulation of the likelihood ratio (LR)

In Forensic DNA Analysis, all statistical interpretations are reported in the form of a likelihood ratio which is the comparison of the probabilities of the evidence under two alternative propositions. These alternative propositions represent the position of the prosecution and the defence.

$$LR = \frac{Pr\,(E \mid H_p)}{Pr\,(E \mid H_d)}$$

H_p (or H_1) = prosecution hypothesis, usually inclusionary with respect to the POI
Numerator = probability of observing the evidence if H_p is true

H_d (or H_2) = defence hypothesis, usually exclusionary with respect to the POI
Denominator = probability of observing the evidence if H_d is true

If the LR is greater than one, then the evidence favours H_p but if it is less than one then the evidence favours H_d.

In the single-contributor case:

- the probability of the evidence profile under H_p (the POI is the contributor) is close to 1 and,
- the LR reduces to the reciprocal of the probability of the DNA profile occurring if the POI had not contributed.

10.1 Numerator

The numerator may or may not be completely defined by the prosecution hypothesis, and this will affect the value for the LR calculation.

10.1.1 Numerator completely defined by H_p:

If the prosecution hypothesis completely defines the numerator, for example, the mixture comes from the victim (V) and the POI:

H_p: V and POI

If all of the alleles present in the profiles of both the victim and the POI are present in the mixture, and if their genotype combinations are the best supported throughout the mixture taking into account the peak heights, then the numerator can be assumed to be one.

10.1.2 Numerator not completely defined by H_p:

If the prosecution hypothesis does not completely define the numerator then the numerator is not one; for example, the mixture comes from the POI and an unknown person (U):

H_p: POI and U

If all of the alleles present in the profile of the POI are present in the mixture and the relative proportions are maintained through the mixture, then the numerator depends on the genotype combinations after the POI's profile has been assumed. There may be single or multiple combinations at each locus.

10.2 Denominator

The aim for the denominator is to include all of the genotype combinations under the defence hypothesis. The denominator may or may not include conditioning.

10.2.1 Denominator including conditioning:

If the same contributor appears in both the numerator and denominator, then their profile may be used to condition the interpretation. The effect of this is to limit the number of genotype combinations in the denominator to only those pairs that include the genotype of the conditioned profile. For example, the mixture comes from the victim and an unknown person:

H_d: V and U

10.2.2 Denominator not including conditioning:

If conditioning is not possible, then all the genotype combinations for the two unknown contributors are assessed. The defence proposition is then that the mixture comes from two unknown people:

H_d: U1 and U2

10.3 Three families of LR

Note: The following explanations consider only the binary model of mixture interpretation for ease of understanding. Mixture interpretation is performed using STRmix and therefore LRs also include the weights assigned to the genotype combinations.

The pairs of propositions for two person mixtures used in casework can be divided into three families.

1. ONE unknown, e.g. victim and POI / victim and unknown
Case scenario: victim alleges rape by one man

For the propositions

H_p: DNA is a mixture of victim and POI
H_d: DNA is a mixture of victim and unknown

the calculation would be:

$$LR = \frac{\text{V and POI}}{\text{V and U}}$$

This pair of propositions requires a conditioning profile in the denominator. If the numerator can be reasonably assumed to be 1, and the denominator includes all the options remaining after the conditioning profile is allowed for, then this calculation simplifies to:

$$LR = \frac{1}{\sum \text{Genotype combinations for U}}$$

2. THREE unknowns, e.g. POI and unknown / 2 unknowns
Case scenario: victim alleges rape by two men but only one POI has been arrested. The profile obtained from her vaginal swab contains DNA from an unknown source as well as DNA which matches the POI.

$$LR = \frac{\text{POI and U}}{\text{U1 and U2}}$$

It is possible that the U in the numerator is the same as either U1 or U2 in the denominator but the key point is that the numerator is not 1, so combinations need to be considered for this as well as for the denominator. The calculation is:

$$LR = \frac{\sum \text{Genotype combinations for U}}{2 \times \sum \text{Genotype combinations for U1 and U2}}$$

3. TWO unknowns, e.g. POI1 and POI2 / 2 unknowns
Case scenario: victim alleges rape by two men; two POIs are arrested whose combined profiles can account for all of the alleles in the mixture.

$$LR = \frac{POI1 \text{ and } POI2}{U1 \text{ and } U2}$$

If all of the alleles present in the profiles of both POIs are present in the mixture, and if their genotype combinations are the best supported throughout the mixture taking into account the peak areas, then the numerator can be assumed to be one. The calculation simplifies to:

$$LR = \frac{1}{2 \times \sum \text{Genotype combinations for U1 and U2}}$$

This LR calculation **should not** be used unless specifically requested by the courts since it is based on the assumption that the DNA from both people (represented by the numerator) was deposited at the same time. When considering the suspects separately the LR is much lower than when they are considered together in the same proposition. The denominator will always be the same, but the numerator will have a big effect on the LR.

11 Transposed conditional

"The probability of the DNA profile (E) occurring if the DNA had come from Mr X (H_p)" is a conditional statement and is expressed mathematically as $Pr(E \mid H_p)$. In other words, the probability of the evidence given the prosecution hypothesis – the prosecution hypothesis is the condition.

When the condition is moved behind the bar $Pr(H_p \mid E)$, it is said to be transposed.

A simple example is as follows:

Evidence = I have 4 legs
Proposition = I am an elephant

Probability of the evidence given the proposition is going to be nearly 1 (probability of having 4 legs given that I am an elephant).

The transposed condition is 'the probability of being an elephant given that I have 4 legs' and that probability would be very small.

Examples of the transposed conditional in the forensic context can be found in Evett, 1995 [9].

12 Records

Nil

13 Associated Documentation

QIS 35007 Use of STRmix v2.7.0 Software
QIS 17117 Procedure for Case Management
QIS 33773 Procedure for Profile Data Analysis using the Forensic Register

14 References

1. Balding DJ and Nichols RA (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. For Sci Int 64: 125-140.
2. National Research Council Report "The Evaluation of Forensic DNA Evidence", 1996, National Academy Press (Washington D.C). Referred to as the second NRC report.
3. Taylor D, Bright J-A, Buckleton J, Curran J. An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations. Forensic Science International: Genetics. 2014;11:56-63.
4. Biology Specialist Advisory Group (BSAG) Statistics Scientific Working Group (StatSWG) recommendations for the interpretation of DNA, December 2016
5. Bright, J., Allen, C., Fountain, S., Gray, K., Grover, D., Neville, S., Poy, A., Taylor, D., Turbett, G. and Wilson-Wilde, L., 2014. Australian population data for the twenty Promega PowerPlex 21 short tandem repeat loci. Australian Journal of Forensic Sciences, 46(4), pp.442-446.
6. T. M. Clayton, J. P. Whitaker, R. Sparkes, P. Gill; Analysis and interpretation of mixed forensic stains using DNA STR profiling; For. Sci. Int.; 91 (1998), 55-70.
7. Parry R, Howes J, Allen C, The determination of the threshold number of alleles, above which single source DNA profiles can confidently be ascribed a likelihood ration in excess of 100 billion, 24 June 2014, located in I:\Change Management\Proposal#134 - Number of alleles for SS LR greater than 100billion
8. QIS 27923 - Risk Assessment: 32 alleles for SS profiles with STRmix v2.6.0
9. Evett IW (1995). Avoiding the transposed conditional. Science & Justice. 35: 127-131
10. Buckleton, J., Bright, J. and Taylor, D., 2016. *Forensic DNA Evidence Interpretation*. 2nd ed. CRC Press, pp.87-117.
11. Caunt, E. Morgan, R. Howes, J. Allen, C. Development of guidelines for the determination of number of contributors to a Powerplex®21 profile – Version 2, 2015

15 Amendment History

Version	Date	Author/s	Amendments
2	24 Feb 1999	V Ientile	
3	8 Oct 2001	V Ientile	
4	12 Jan 2004	P Clausen	Complete re-write of Statistical Calculations used in the Forensic Biology laboratory, including use of BRB Stats and Parentage Calculations.
5	19 July 2005	S Cave/V Ientile	Removed reference to use of effective individualization. Removed references to parentage testing documents, included additional explanations of suggested reporting phrases.
6	10 Mar 2008	S Cave, P Clausen, L Weston, T Gardam, &	Major Revision. Document spilt so as to cover single source profiles only. Definitions added, population database information updated, relatives values

		R Smith.	added, References reviewed and updated to include only those related to single source profiles, Appendix of formulae added.
7	April 2008	QIS2 Migration Project	Headers and Footers changed to new CaSS format. Amended Business references from QHSS to FSS, QHPSS to CaSS and QHPS to Pathology Queensland
8	24/02/2009	T Nurthen	Added Kinship references, re-word Random match probability statement to include unrelated data
9	10/05/2012	E Caunt	Removed reference to A & G Stats, BRB Stats and DNA View. Match probability changed to random match probability and definition changed. References to DNA Analysis laboratory changed to DNA Analysis Unit. Interpretation profiles with NR peaks changed to sub-threshold peaks. Sections 8 and 9 removed. Removed archived associated documents.
10	13/05/2013	E Caunt	Changed to HSSA template, removed reference to DNA Analysis 'Unit', changed to refer to P+ profiles only
11	February 2015	E Caunt	Rewrite
12	January 2017	E Caunt	Typographical errors fixed, definition of LR changed (pg 2), added reference to Torres Strait Islander samples in dataset, removed "all" jurisdictions provided data (section 5.2); changed permission for relatives calculations from Managing Scientist to substantive Team Leader
13	Feb 2020	E Caunt	
14	June 2020	E Caunt	Major revision, combined with QIS 25302v5, added number of contributors guidelines changing stutter from locus specific to allele specific, added intuitive exclusion guide

16 Appendices

16.1 Assessment of the Number of Contributors for Mixed PowerPlex® 21 DNA Profiles within Forensic DNA Analysis

16.1.1 Introduction

The aim of these guidelines is to assist in the assessment of the number of contributors for mixed DNA profiles obtained using the PowerPlex® 21 system. These guidelines should be used in conjunction with the training and experience of the scientist. There may be features within the DNA profile other than those detailed in this document that may inform the number of contributors. If the scientist observes information/behaviours within the DNA profile that override these guidelines, it is acceptable for these observations to be used in the determination of the number of contributors. There are also certain reworks that may be required, for example for quality reasons, before a reasonable assessment of the number of contributors can be made and these should be performed separately to the guidelines provided. Background information has also been presented in this document as this has been considered in the development of the recommendations provided. References to stochastic effects relate to peaks which may drop out or be imbalanced.

16.1.2 Background

Bright et al [1] showed that the assumption of an increased number of contributors to a mixed profile does not affect the likelihood ratio (LR) assigned to the stronger contributor to the profile. However, the assumption of an increased number of contributors may significantly decrease the LR assigned to known low level contributors. This is definitely the case if all of the alleles in the low level component of the profile match those of the known contributor under the assumption of two contributors. Consider however a mixed profile that consists of three contributors (with two of those contributors being in the low level component), due to the stochastic effects of low level profiles, the low level contribution appears to only have originated from one person. In this case, a known contributor to the mixture could be excluded from the low level contribution if alleles from both low level contributors are displayed in the profile. Additionally, the low level contribution could be considered to be an unknown profile and be loaded to NCIDD possibly leading to adventitious links. This has been demonstrated in constructed mixtures where known contributors were not related [2].

It is noted that the true number of contributors to a casework mixture can never be known with certainty. If a three contributor casework mixture (with two low level contributions) was incorrectly assessed as having originated from two contributors, then the risk is that a true contributor to the low level contribution would be excluded. Exclusion is a statement of certainty rather than a LR favouring non-contribution and false exclusion is not a desirable outcome. Conversely, if a two contributor casework mixture was incorrectly assessed as having originated from three contributors, then there is an increased chance of a non-contributor being falsely ‘included’. An incorrect assessment of this type is most likely to occur with a mixture with a stronger contribution and one/two low level contributions, with a non-contributor being falsely ‘included’ in the low level contribution. In this instance, it is expected that the non-contributor would be a poor fit to the profile and would produce a LR favouring non-contribution or a LR approaching 1.

16.1.3 Initial assessment

Before assessing the number of contributors to a mixed DNA profile, a general overview of how the profile is behaving should be obtained. The following things should be looked for in profiles generated using the PowerPlex®21 system:

- Preferential Amplification (PA) – this appears as an arc shape in the blue and/or red dyes (i.e. increasing peak heights from D3 to D6 before decreasing towards Penta E) and indicates too much input DNA in the amplification [3]. This pattern has the potential to mask the true number of contributors and the sample should be reworked by decreasing the amount of input DNA.
- Degradation – this is expected to occur in some samples and is displayed as decreasing peak heights moving from left to right across a profile in accordance with an increase in fragment size [3]. This pattern has the potential to misinform the number of contributors to a profile. Degradation is not usually able to be rectified by reworking, but should be factored into the interpretation.
- Bad baseline – this can be displayed in different ways, but is often due to excessive pull-up. Bad baseline has the potential to mask any low level contributions. Samples displaying bad baseline should be reworked according to the cause e.g. too much input DNA.

16.1.4 Stutter

Before assessing how many contributors there may be to a profile, the scientist should examine the profile for the presence of artefacts. Kit specific artefacts and -2 and +1 repeat stutters should be removed at plate reading stage. Some -2 and +1 repeat stutter peaks may still be present at interpretation stage for reasons including peaks above one threshold but below a combined threshold, or they fall below the limit of reporting.

In-house stutter thresholds determined during the validation of the PowerPlex®21 system are provided as a guideline to assist the scientist to determine the number of contributors. They are calculated using the mean peak height + 3 standard deviations meaning that, for each threshold there are outliers. Some of the thresholds have been calculated using very few data points (see Appendix Section 16.5) and therefore it would be expected that stutter peaks above the threshold may be observed.

The stutter threshold should be used to inform the scientist of the expected height of stutter peaks at a particular locus, however an allele that falls slightly above threshold may not necessarily be indicating an additional contributor e.g. with a stutter threshold of 14.6%, a peak does not necessarily become an allele at 14.7%. It is appropriate for a scientist to use their experience in conjunction with the thresholds to determine whether a peak should be considered to be a stutter or an allele.

Please refer to see Appendix Section 16.5 for information regarding the designation of -2, -1 and +1 rpt stutter peaks and guidelines for combining thresholds.

16.1.5 Limited information

In some instances there may be very little information within the low level contribution of the profile and further investigation may not assist the scientist to provide meaningful interpretation of this contribution. Regardless of how many contributors there are to this low level component, interpretation will be limited. If the low level contribution, where it is expected that the impact of stochastic variation is heightened, consists of four or fewer alleles (including sub-threshold peaks) it is acceptable to consider that there is only one

contributor to this low level component[1]. It is at the discretion of the scientist whether to rework the sample or not. There may be scenarios where this number is not appropriate, however for the majority of profiles in this category this approach is considered suitable and is favoured over increasing the number of contributors to this limited component.

The reasoning behind using a cut-off of four alleles uses experience from Profiler Plus interpretations, an understanding of LRs that are generated by components with limited information as well as the success of reworking samples. It has also been developed based on the understanding of the risks of false inclusion and exclusion.

When considering a profile interpreted using a binary approach, such as with Profiler Plus, scientists within Forensic DNA Analysis would generally consider a minor profile to be unsuitable for meaningful interpretation if approximately four, or less, alleles are obtained. This is due to the increased probability of more than one person having these four alleles in their DNA profile. With a PP21 profile and a continuous interpretation approach, if a person matched all four alleles in a low level contribution, it is likely that the LR at best would provide low support for contribution.

When dealing with so few alleles, any interpretation of this information will be limited. It is likely that reworking the profile will not provide any additional information to either inform the number of contributors or to add significant weighting to the interpretation. It is therefore suggested that reworking a sample such as this will not be beneficial and that the profile should be interpreted as it stands.

The limitation of interpreting these low level components as single source is that a person could be falsely excluded if the alleles do originate from two different people. However, as previously mentioned it is expected that any LR generated under the assumption of an increased number of contributors would provide limited information.

16.1.6 Sub-threshold peaks

Forensic DNA Analysis has validated a limit of reporting for the purpose of confidently distinguishing true allelic peaks from background noise [4]. This means that only those peaks above the limit of reporting can be used in the statistical analysis of DNA profiles. It is noted however that there is a chance that peaks below this level could be from DNA and the closer these peaks are to the limit of reporting, the more likely they are to be from DNA. Where there is a low level contribution to the profile it is expected that these sub-threshold peaks could interfere with the interpretation of the allelic peaks above the limit of reporting and therefore should be considered in the determination of the number of contributors. The results of the testing have demonstrated that it is appropriate to use sub-threshold peaks during the interpretation of a DNA profile. Following on from this, if these sub-threshold peaks are used in the determination of the number of contributors it is expected that they would also be used for exclusionary purposes.

Note: Sub-threshold peaks should only be considered if they are distinct from baseline, above the LOD, below the LOR and not potential stutter peaks.

16.1.7 Conditioning

Before interpreting a profile, the scientist should assess from the case circumstances/sample type whether it is reasonable to assume the presence of DNA from a

[1] Unless three or more of these alleles are at one locus, in which case there is evidence that there is more than one contributor to the low level component.

person. If the presence of DNA from a person can be assumed and there is information within the profile supporting this assumption, then the profile can be conditioned. There is no minimum number of alleles required before this assumption can be made.

If the profile has one or more contributors in the stochastic range, then the presence of one or more 'matching' alleles in that contribution is enough to assume the presence of DNA from the conditioned reference sample.

If the low level contribution can be accounted for by the conditioned reference sample in its entirety, then the low level contribution can be assumed to have originated from one contributor (the conditioned reference sample).

If the low level contribution can only be partly accounted for (one or more matching alleles) by the conditioned reference sample, then the number of contributors should be adjusted to allow for the alleles, in addition to the conditioned contribution, that are present.

Given sub-threshold peaks are considered in the assessment of the number of contributors these must also be consistent with the conditioned contribution where a low level component is assumed to be single source. If this is not the case, there is no need to add a contributor to account for these sub-threshold peaks, rather, the profile can be reported as having possible sub-threshold peaks not interfering with the interpretation.

16.1.8 Reproducibility

When a profile has a low level contributor, it is expected that this contributor will be subjected to stochastic effects. If there are two low level contributors to the mixture, then both of those contributors will be subjected to these stochastic effects. What could potentially happen, and has been observed [2], is that different peaks from each contributor could be displayed in the profile; however other peaks have dropped out. This could make the profile appear as though there is only one low level contributor rather than two.

If the profile has been reworked such that there are, for example three amplifications, then these three profiles should be considered together. When considered together, the following observations may be made:

1. The profiles have different alleles (either above or below threshold) in the low level contribution, however it appears that overall there is only one low level contributor.
2. The profiles have the same alleles (either above or below threshold) in the low level contribution and it appears that there is only one low level contributor.
3. There is additional information in subsequent amplifications and it is evident that there are two low level contributors.

If the observation is consistent with option 1, then it may be the case that there are two low level contributors, however the stochastic effects are such that different alleles are appearing in the different amplifications. In this instance, there is no certainty that there is only one contributor to the low level contribution and a contributor should be added.

If the observation is consistent with option 2, then the fact that the low level contribution is reproducible should give confidence to the notion that there is only one contributor to the low level component.

If the observation is consistent with option 3, then the profile should be considered as a three person mixture (two low level contributors).

Thee amplifications is the suggested requirement for profiles with a suspected single source low level component because testing has demonstrated this increases the opportunity to identify where peaks are reproducible [2].

16.1.9 Guidelines

1. **Initial assessment of number of contributors**
   This should involve an assessment of the number of peaks (including those which are sub-threshold), the height of the peaks and the presence of stutters and other kit specific artefacts to identify the minimum number of potential contributors to the mixed DNA profile.

2. **Stochastic range**[2]
   This should involve assessing whether any one potential contribution may be subject to stochastic effects. This may be indicated by the presence of peaks less than ~300RFU or by a quantitation value which has resulted in an input of any one potential contribution to be less than ~150pg. If there is no contribution falling within the stochastic range further investigation may not be required and interpretation can continue based on the original assessment of the minimum number of contributors (see Appendix Section 16.3 for further information).

3. **Conditioning**
   Where the presence of a reference sample can be assumed this should be conducted before further investigation into the potential number of contributors. If this assumed contributor can account for the low level contribution in its entirety then any stochastic variation can be accepted. In this case no further investigation is required and interpretation may continue based on the original assessment of the minimum number of contributors. If this assumed contributor can account for only part of the low level contribution, it is reasonable to increase the original assessment of the number of contributors to account for this assumed contribution.

4. **Allelic Imbalance (AI)**
   Whilst there is no defined AI threshold for casework profiles, it is reasonable to see imbalance as low as 40% [4]. If there is imbalance below this figure at one or more loci, this sample should be reworked to confirm the ratio of these peaks as this imbalance could indicate an additional contributor to the low level contribution. If there is one imbalance between 40%-60% this may be accepted without further investigation, however if imbalance within this range is present at more than one locus the sample should also be reworked to confirm the ratio of these peaks. Imbalance, or variation in peaks heights, between loci within the same dye should also be considered, and imbalance between loci within the same dye not consistent with the expected degradation should be investigated further by reworking the sample[3].

5. **Degradation**
   Degradation is characterised as a decrease in peak heights moving from left to right across a profile as the molecular weight increases. Laboratory observations [3] have shown that some variation can be seen however larger peaks at larger fragments may be an indication of more than one contributor to a low level component. If this is observed, the sample

[2] The stochastic range has been developed by the project officers based on observations made in the PowerPlex 21 validation [4].
[3] AI is only confirmed if that same pattern is observed across runs, i.e. 14 (1000rfu) and 16(400rfu) in one run and 14 (1200rfu and 16 (480rfu) in a second run = AI confirmed; if the AI flips to 14 (400rfu) and 16 (1000rfu) in the second run or if the partner drops out this is an assessment of the stochastic effects within the profile.

should be reworked to assist the scientist to determine whether this pattern is reproducible or due to expected variation.

6. Ratios

It is noted that the peak heights may differ between dye lanes, however the approximate ratio of the mixture should hold across the profile [5]. This means that if the mixture appears to have a stronger contribution, a mid range contribution and a low level contribution in the blue dye, then this should also be represented in the other dyes. If the green dye then displays a stronger contribution and two low level contributions, this may indicate that there are four contributors to the mixture. If there is uncertainty regarding the mixture ratio, a rework should be performed to confirm, or otherwise, how the profile is behaving.

7. Reworking

The aim of the rework should be to confirm how a profile is behaving, assess the reproducibility of a component(s) for which the number of contributors is unclear or to potentially provide additional information in the form of additional peaks. Where more extract can be included in the amplification without overloading or increasing baseline noise, this should be done. For mixed DNA profiles the input can be increased above 0.5ng where it is suitable to do so based on the peak heights and the complexity of the profile [3]. In other cases where an increase in template is not possible a repeat amplification is sufficient. Where reworks have been performed, the minimum number of contributors may need to be reassessed based on the reproducibility of peaks or additional information that may have been obtained. It is recommended that, where a contribution in the stochastic range is thought to be single source, two reworks are performed so that final assessment can be made with a total of three amplifications. Depending on whether the input template is being increased or kept the same[4], these reworks may be ordered at the same time. It is also recommended that no more than three amplifications are performed for the determination of the number of contributors, unless there is an issue with one or more of these runs, since more amplifications may increase the complexity of the interpretation.

8. Reproducibility[5]

This should be used to confirm a single source contribution within the stochastic range and the following guidelines have been developed to address this. The level of reproducibility will depend on the rework that was ordered, that is, if there is a change in input template it is possible that the number of peaks and peak heights will differ between amps and potentially make the reproducibility assessment more difficult. For single source contributions within the stochastic range:

a. Labelled peaks – it is expected that not all peaks within the stochastic range will be reproduced due to the nature of DNA profiling, however a single source contribution in the stochastic range with greater than 12 labelled alleles should have 80% of alleles reproduced twice out of a possible three runs [2] (see Appendix Section 16.4 for further information). If this reproducibility is not achieved the number of contributors should be increased by one. If there is a suspected single source contribution within the stochastic range with less than 12 labelled alleles a lower level of reproducibility may be expected. This has been demonstrated to be around 60% but may fall as low as 30%, therefore whether this contribution can be called single source should be at the discretion of the scientist in conjunction with other factors.
b. Peak heights – it is expected that peak heights within the stochastic range will fluctuate due to the nature of DNA profiling, however the ratio should remain

[4] Where the input template is changed it may be appropriate to order one rework at a time as there is a chance the sample may become overloaded.
[5] Reproducibility is defined as the number of reproduced peaks divided by the total number of peaks in this contribution.

consistent across runs. When there is a change in the ratio at one or more loci it may be appropriate to increase the number of contributors by one.

c. Allelic imbalance – if any AI less than 40% is confirmed at one or more loci then increasing the number of contributors by one should be considered.
d. Degradation – it is expected that peak heights will decrease when moving from left to right and if any apparently significant deviation from this is confirmed on rework the number of contributors should be increased by one.

16.1.10 Reference comparisons

Following the determination of the number of contributors to a mixture, reference samples (if available) can be compared. Unless conditioning, it is important that reference samples are not compared until an assessment of the number of contributors has been completed.

Although the LRs are calculated separately for each reference sample in the case, the manual comparison should include a check of all reference samples together, particularly for strong profiles with low mixture ratios. For example, if the casework sample gives a two person mixture and both reference samples could have contributed, it is important to check that both reference samples could have contributed together. If they could not then multiple scenarios are possible:

- Person 1 and an unknown (not person 2) have contributed
- Person 2 and an unknown (not person 1) have contributed
- Neither person 1 nor person 2 have contributed
- Person 1 and person 2 have contributed but the profile actually has more than two contributors

It is not possible for the scientist to determine which of these scenarios is more likely leaving only two options for reporting the result to the court. The first option is to present all of the scenarios to the court and let the court decide which scenario is more likely; the second option is to concede that the number of contributors cannot be determined and that the profile is unsuitable for meaningful interpretation. It is considered that the second option is the most appropriate approach. Taylor et al [6] acknowledge that presenting these options to a court places the decision with the court and that if it is not possible for an expert to distinguish between these scenarios then it is asking a lot for the court to do so. If this scenario is encountered the number of contributors should not be increased to allow for person 1 and person 2 to have contributed together without further investigation into the reassessment of the number of contributors including reworks.

This approach may not be relevant for all mixture types. If a profile has contributors in the stochastic range there may only be limited information within the profile with which to make an assessment regarding whether or not a person could be a potential contributor. These assessments would generally result in a LR closer to 1. It is important to note that there is little difference between two different values which only slightly fall either side of 1 and that it is not expected that it is the role of the scientist to determine the level of support for either proposition where a figure falls in this range. It is expected in these instances that the LR provided would speak for itself in terms of the level of support or fit to the profile for that reference comparison.

When the reference sample(s) is compared, it may be that the reference sample is a poor fit to the profile under the assumed number of contributors, however increasing the number of contributors will make the reference sample a good fit to the profile. It is not acceptable to increase the number of contributors based on the fit of the reference sample. If this situation is encountered, then the scientist may decide to rework the sample to reinforce, or

otherwise, the assessment of the number of contributors. Should the rework support the change in the number of contributors (without the influence of the reference sample), then this may occur. Should the rework back up the original assessment of the number of contributors, then this assessment should stand. If the scientist considers that the fit of the reference sample to the profile is not intuitively correct then the scientist has the option of deeming the profile unsuitable for interpretation.

16.1.11 References

[1] J. Bright, J.M. Curran, J.S. Buckleton, The effect of uncertainty in the number of contributors to mixed DNA profiles on profile interpretation, Forensic Sci. Int.: Genet. 12 (2014) 208-214

[2] E. Caunt, R. Morgan, J. Howes, C. Allen, Development of guidelines for the determination of number of contributors to a PowerPlex®21 profile 2015

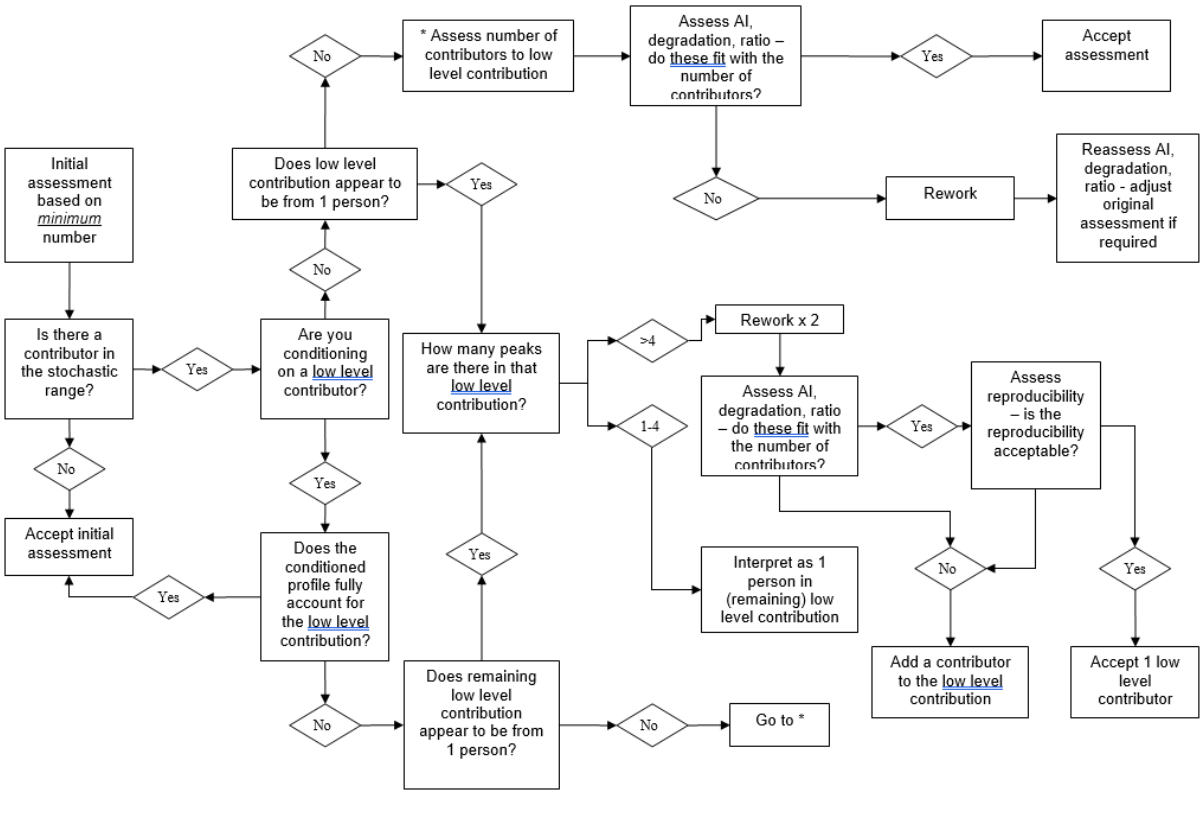
[3] Powerpoint presentation, PowerPlex 21 Observations, October 2014 (located in G:\ForBiol\AAA Forensic Reporting & Intel\AAA_Reporting guidelines\PP21 and STRmix™ case mgt)

[4] T. Nurthen, M. Mathieson, C. Allen, PowerPlex 21 – Amplification of extracted DNA validation v2.0 2013

[5] T.M. Clayton, J.P.Whitaker, R. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Sci. Int. 91 (1998) 55-70

[6] D. Taylor, J. Bright, J. Buckleton, Interpreting forensic DNA profiling evidence without specifying the number of contributors, Forensic Sci. Int.: Genet. 13 (2014) 269-280

16.2 Workflow



16.3 Determining contributors within the stochastic range

A contribution within the stochastic range is defined as one with peaks heights less than 300RFU and/or an input of less than 150pg. It may be difficult to determine whether all of the peaks of a contribution that is potentially in the stochastic range are below 300RFU due to the possibility of sharing with other allelic peaks or stutters. If all of the peaks associated with one (or more) low level contribution(s) are below 300RFU then that contribution is clearly in the stochastic range. However, if some of the peaks are below 300RFU and some of the peaks are above 300RFU it may be necessary to use the following additional information to assist the determination of whether this contribution falls within the stochastic range:

- Quantitation value and approximate input template of contribution
- The proportion of peaks above and below 300RFU
- For any peaks below 300RFU, how much below 300RFU they are, e.g. 298RFU vs 56RFU

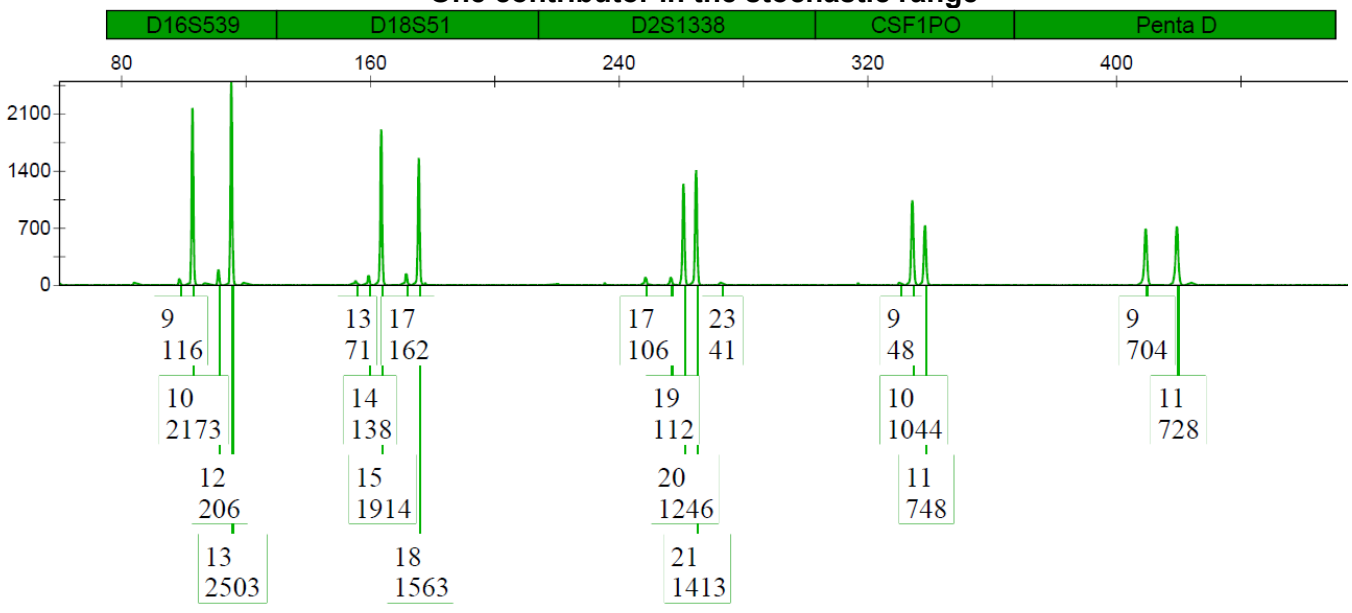
When determining the approximate theoretical input of a contribution the total input template and the approximate proportion of the questioned contribution in the profile should be used. It is important to note that the values of 300RFU and 150pg are approximate and that a contribution at 299RFU and/or 149pg is not much different from a contribution at 301RFU and/or 151pg.

Multiple contributors in the stochastic range

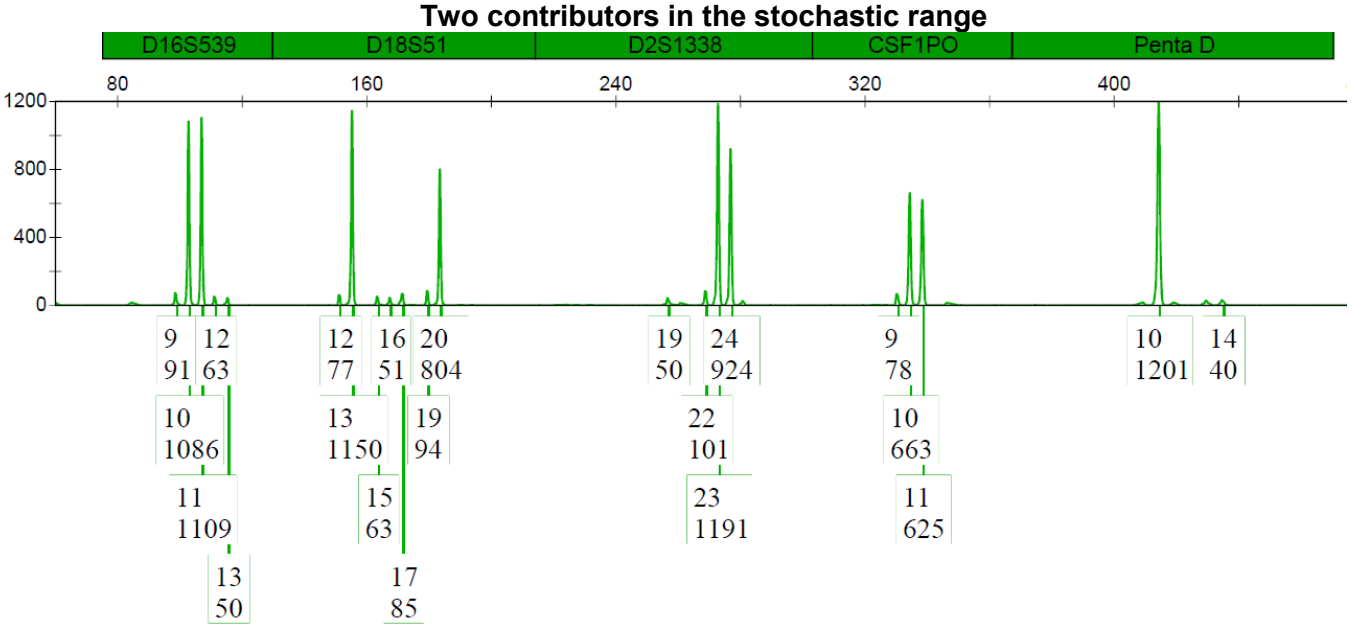
The contributors to a mixture can range in ratio from distinct strong contributions and low level contributions to all strong contributions or all low level contributions. If the profile has more than one low level contribution that falls either within or close to the stochastic range the risk of false exclusion and/or uploading a false profile to NCIDD decreases. In this instance the STRmix™ deconvolution should be flexible enough (in most situations) to allow the LR relating to the low level contribution to speak for itself, i.e. reference samples with a poor fit will produce a low LR and reference samples with a good fit will produce a higher LR. Unless there are a lot of labelled alleles at a locus, exclusion is unlikely (as allele,NR and NR,NR combinations will be considered) and false exclusion more unlikely still. With this type of profile it is not expected that a profile for upload to NCIDD will be deconvoluted. Therefore, profiles with more than one low level contribution are not required to have as thorough an investigation as those with only one low level contribution.

Examples

One contributor in the stochastic range



Two contributors in the stochastic range



16.4 Reproducibility calculation

Reproducibility is only used to confirm a single source component in the stochastic range. It is not possible to determine the reproducibility for a mixed component due to the increase in variability and the inability to determine which alleles belong to a single contributor. To determine whether a profile falls into the <12 or >12 allele category only labelled peaks should be included (this would be the total number of different labelled peaks across the three runs). This number has been selected as it is the upload to NCIDD threshold. Once the category has been selected, sub-threshold peaks may then be considered.

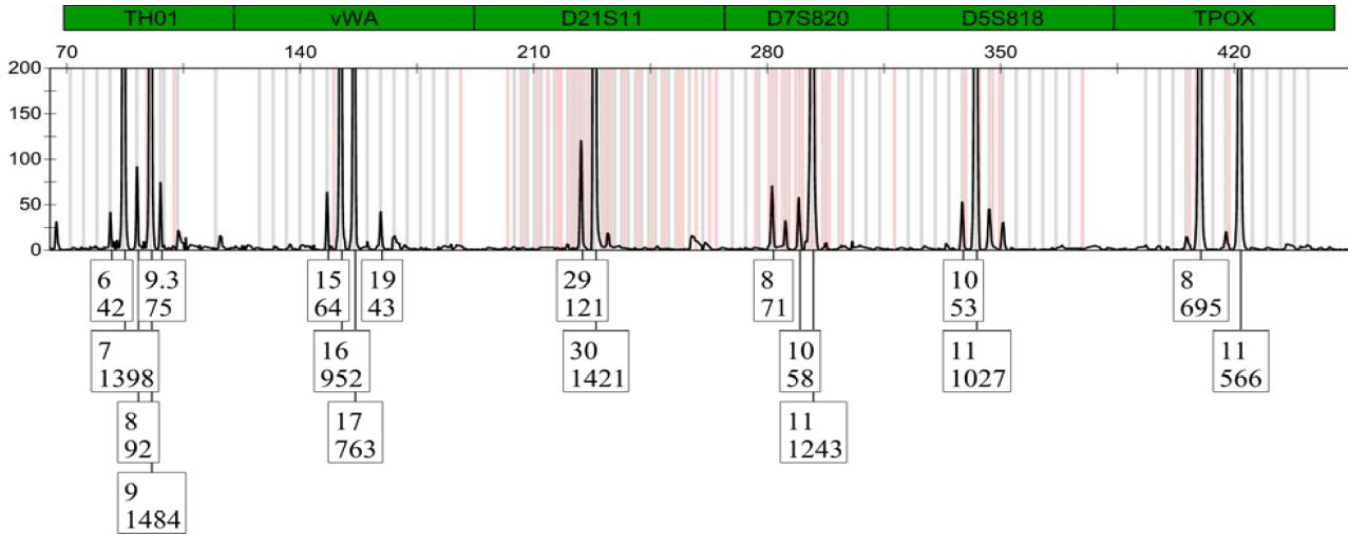
To determine whether a peak is reproduced peaks above and below threshold are considered. If any peak above or below threshold is only seen once across all of the three runs this peak is considered not reproduced. If a peak is seen above or below threshold two or three times across the three runs this peak is considered reproduced. To determine the percentage reproducibility the number of reproduced alleles (both above and below threshold) should be divided by the total number of different alleles seen (reproduced and not reproduced both above and below threshold) and then multiply by 100.

Three runs are required to determine reproducibility due to the stochastic nature of low level DNA.

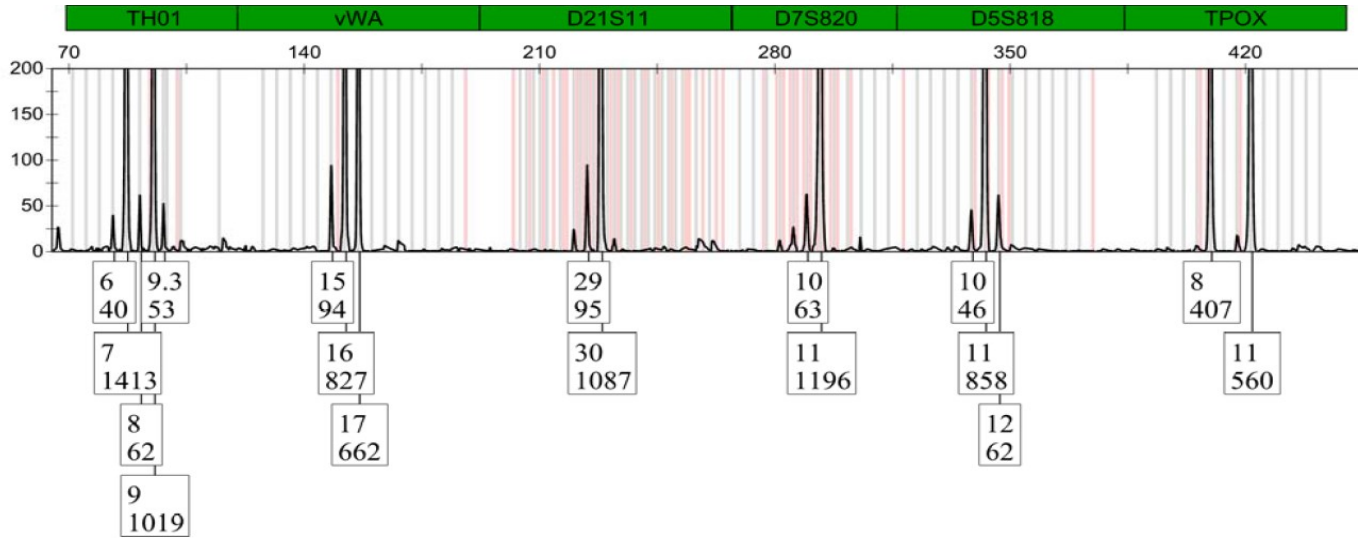
Example reproducibility calculation

The diagrams below show three runs of the same profile. The table details the alleles observed in the stochastic range (including sub-threshold alleles above the LOD shown in brackets) for each of the three runs. The final column shows whether the allele is reproduced.

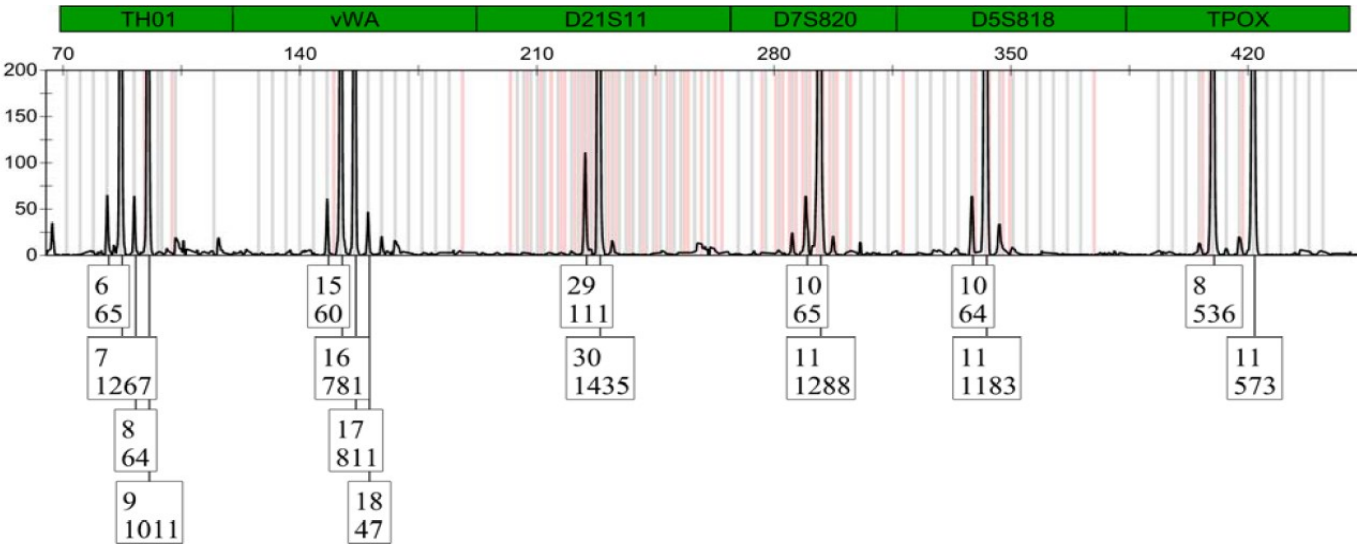
Run 1



Run 2



Run 3



Run 1	Run 2	Run 3	Reproduced
TH01 9.3	TH01 9.3		Yes
		vWA 18	No
vWA 19		(vWA 19)	Yes
D7 8			No
	D5 12		No
(D5 13)			No

A total of six alleles were observed in the stochastic range, of these two were reproduced. The reproducibility is therefore 2/6 = 33%.

Note: If a peak falls in a stutter position and is considered part of the low level contribution as it is above the stutter threshold it can be included in the reproducibility calculation.

16.5 Stutter

Forensic DNA Analysis has implemented allele specific -1 rpt stutter thresholds and locus specific +1 rpt and -2 rpt stutter threshold based on national data and data from the validation of the PowerPlex®21 kit. These stutter thresholds are calculated by comparing the height of the stutter peak to the height of the allele peak. The -1 repeat stutter thresholds are used during case management for the determination of number of contributors, whilst -2 repeat and +1 repeat stutter thresholds are used at plate reading to determine which peaks should be removed before the profile is suitable for interpretation.

As STRmix™ cannot model -2 repeat and +1 repeat stutter peaks, a binary approach is applied in that those peaks which are equal to or less than the threshold are removed before the profile is interpreted. It should be noted that where this has resulted in a large peak being removed, STRmix™ users need to check the deconvolution to ensure the removal of this artefact is covered by the consideration of dropout of this allele if required based on the rest of the profile. Where there is a -2 repeat or +1 repeat stutter peak over threshold which may result in an increase in the number of contributors reworking to confirm that this peak is not stutter may also be performed.

As STRmix™ can model -1 repeat stutter, and does so up to a value of 30% of the main peak, these thresholds are only used to determine number of contributors and the % value is only used as a guideline. Where there is a peak below the +1 repeat stutter threshold but also below the -1 repeat stutter threshold, this peak is left labelled to be modelled by STRmix™ as -1 repeat stutter. There are a few scenarios encountered in casework for which guidelines have had to be developed:

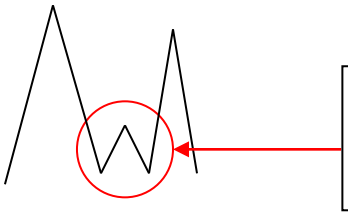


Figure 1:
Peak is above -1rpt stutter threshold (in-house) but below 30%
Peak is below +1rpt stutter threshold

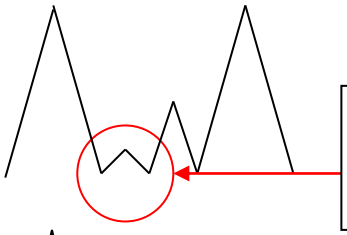


Figure 2:
Peak is above +1rpt and -2rpt stutter thresholds
Peak is below cumulative +1rpt and -2rpt stutter thresholds

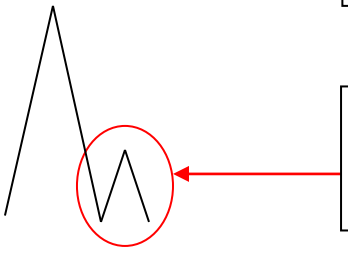


Figure 3:
Peak is below +1 stutter threshold
Peak is above dropout range eg.400RFU

In Figure 1 the highlighted peak is above the -1 repeat stutter threshold but is below the +1 repeat stutter threshold. In this example the peak would be removed at plate reading stage because if it was left on STRmix™ may model this peak as -1 repeat stutter but it is above the in-house stutter threshold and therefore may lead to the number of contributors being increased. The deconvolution would be checked to ensure the removal of this artefact is

covered by the consideration of drop-out of this allele if required based on the rest of the profile.

In Figure 2 the highlighted peak is above the +1 repeat and -2 repeat stutter thresholds. In this example the peak would not be removed at plate reading stage but the reporting scientist may consider adding the two thresholds to remove the peak if it doesn't fit the rest of the profile. There are no strict guidelines around when this approach can be used however it would not be used often. The reporting scientist may also choose to rework the sample to aid in their decision making.

In Figure 3 the highlighted peak is below the +1 repeat stutter threshold but is above the RFU range at which we would reasonably expect STRmix™ to model dropout. As a result, a reference sample which may have contributed to the profile and has the allele that falls in that position may be excluded simply because of the height of the main allele. If dropout is not modelled even after increasing the number of iterations this sample would be reworked to see if the peak falls above the threshold on another run.

16.6 Guide for performing intuitive exclusions

Once the scientist has made an assessment of the number of contributors to a DNA profile, the next stage is to compare DNA profiles obtained from reference samples associated to the case to determine whether these people could be potential contributors. For a single source DNA profile, this is a simple process since the reference sample must match all alleles present within the crime scene profile in order for it to be considered a potential contributor. This process is more complex for mixed DNA profiles as there are more allelic peaks, and combinations of these peaks, to be considered.

As an example, for a two person mixture where there are four alleles at a locus, the reference sample must match two of those alleles to be considered as a potential contributor. If this condition is not met then, under the assumption of two contributors, the reference profile is excluded.

As the number of contributors increase and/or the contributor proportions change, the decision process for excluding a reference profile also changes.

The table below provides a guide for when a reference sample can be excluded as a potential contributor to a DNA profile when comparing the reference sample to an individual locus. It describes the following:
the mixture type that may be encountered;
the component of the mixture to which the reference sample is being compared (for an even mixture this would be the entire mixture, for the minor component of a major/minor mixture this would be only the minor alleles);
the type of match to the reference sample (one or two alleles matching and whether the reference sample is homozygous or heterozygous at that locus).

Where there is an ‘S’, this means that the reference sample may not be able to be excluded intuitively and that a LR should be calculated using STRmix™. Where there is an ‘x’, this means that the reference sample can be excluded without the need to calculate a LR in STRmix™.

The number of alleles being compared does not include peaks that could be -2, -1 or +1 repeat stutter. These alleles can be either above or below the LOR but must be above the LOD.

The ‘component’ of the mixture being compared refers to whether the scientist is comparing to the whole mixture as in an even mixture, or to part of the mixture e.g. major, minor, remaining.

Mixture Type	Number of alleles at locus being examined in component of mixture being compared	Reference sample			
		Doesn't match any alleles	Homozygous matching 1 allele	Heterozygous matches 1 allele	Matches 2 alleles
2 person even mixture - drop-out is unlikely to have occurred	1	x	S	x	n/a
	2	x	S	x	S
	3	x	S	x	S
	4	x	x	x	S
2 person even mixture - drop-out is likely to have occurred	1	S	S	S	n/a
	2	S	S	S	S
	3	x	S	S	S
	4	x	x	x	S
2 person major/minor mixture **- comparison to minor** (major is considered single source)	1	x	S	S	n/a
	2	x	x	x	S
2 person conditioned mixture **- comparison to remaining**	1	x	S	S	n/a
	2	x	x	x	S
3 person even mixture - drop-out is unlikely to have occurred	1	x	S	x	n/a
	2	x	S	x	S
	3	x	S	x	S
	4	x	x	x	S
	5	x	x	x	S
	6	x	x	x	S
3 person even mixture - drop-out is likely to have occurred	1	S	S	S	n/a
	2	S	S	S	S
	3	S	S	S	S
	4	S	S	S	S
	5	x	S	S	S
	6	x	x	x	S
3 person major/minor mixture - 2 people in major - drop-out of major is unlikely to have occurred **- comparison to major**	1	x	S	x	n/a
	2	x	S	x	S
	3	x	S	x	S
	4	x	x	x	S
3 person major/minor mixture - 2 people in major **- comparison to minor** (minor single source)	1	x	S	S	n/a
	2	x	x	x	S
3 person major/minor mixture - 2 people in minor **- comparison to minor** (major is considered single source)	1	S	S	S	n/a
	2	S	S	S	S
	3	x	S	S	S
	4	x	x	x	S
3 person conditioned mixture - 2 people in remaining **- comparison to remaining**	1	S	S	S	n/a
	2	S	S	S	S
	3	x	S	S	S
	4	x	x	x	S

3 person conditioned mixture - remaining is 2 person major/minor **- comparison to remaining minor** (single source) (major is considered single source)	1 2	x x	S x	S x	n/a S