Interpretation of Mixtures Based on Peak Area - Identification of Genetic Anomalies, Stutters and Other Artefacts

P. Gill[1], B. Sparkes[1], T.M. Clayton[2], J. Whittaker[2], A. Urquhart[1], J.S. Buckleton[3]
[1] Forensic Science Service, Priory House, Gooch Street North, Birmingham, B56QQ, UK
[2] Forensic Science Service, Wetherby Laboratory, Sandbeck Way, Audby Lane, Wetherby, West Yorkshire LS22 4DN, UK
[3] ESR: Forensic, Mt Albert Centre, Private Bag 92-021, Auckland, New Zealand

INTRODUCTION

Powerful methods based on likelihood ratios have been developed by Evett *et al.* (1) and Weir *et al.* (2) to interpret mixtures. However, these models presuppose unambiguous identification of alleles prior to analysis and take no account of their relative peak areas. The use of automated sequencer technology makes it relatively simple to collect not only qualitative, but quantitative information (i.e. allele peak height and peak area). When mixtures are simple, originating from 2 individuals, and the ratio of admixture is <1:2 it is possible for the analyst visually to separate the alleles into major and minor components and to use this quantitative information in subsequent interpretation. This is normal practice in our laboratory casework. The method relies upon the experience of the expert who uses the rationale outlined by Clayton *et al.* (3).

Recently, Evett *et al.* (4) have suggested a model to assess mixtures, utilising the information in the peak area. However, the approach also assumes unambiguous identification of alleles. Gill *et al.* (5) have introduced a method to interpret mixtures against a background of artefacts (stutters in particular) and this begins the much needed work to model artefacts.

The purpose of this paper is to illustrate a framework to analyse simple mixtures in order to assist the reporting officer to make a preliminary assessment against a background of possible minor artefacts, prior to a full statistical evaluation. The process of interpretation suggested here can undoubtedly be improved by developing new models to form the basis of programmed expert systems.

NOMENCLATURE

Alleles are designated according to the recommendations of the ISFH (6). A generic designation of simple mixtures is also used based on the relative size of the alleles i.e. A-D in order of increasing molecular weight; a mixture of 2 individuals where the first (suspect) is HUMTH01 6, 7 and the second (victim) is HUMTH01 8, 9.3 is defined as AB, CD; a mixture where the first individual is HUMTH01 6, 9.3 and the second is 7, 8 is defined as AD, BC. A three allele mixture comprising HUMTH01 8, 9.3 and HUMTH01 8, 11 is designated AB,AC. Each locus in a simple mixture is separately designated and comprises 2-4 alleles (Table 1). The peak area is denoted by ϕA-ϕD. The frequency of the allele in the population is denoted *f*a- *f*d.

GUIDELINES

Interpretation of mixtures cannot proceed without an understanding about how non-mixtures behave. In multiplex systems such as that described by Sparkes *et al.* (7), mixtures are identified as multiple banded products at each locus. The bands at a locus will often appear imbalanced, with one or more peaks being markedly larger than others. The presence of artefacts such as stutters in the profile may affect the interpretation.

Before interpreting a potential mixture, it is important to understand the characteristics of heterozygotes and stutters in terms of peak areas and relative positions. To assist interpretation we use guidelines (8), although it is probable that different multiplexes will behave differently to those described. Loci will behave somewhat differently to each other (8) but it is possible to generalise:

- The smallest peak area of a heterozygote will usually be greater than 60% the size of its partner.
- Within the above guideline, the high molecular weight peak is often smaller than the low molecular weight peak since PCR amplifies the latter the most efficiently.
- Stutter peaks are usually less than 15% the area of the associated allelic peak
- Stutters are 1 repeat unit (e.g. 4 bp for tetramerics) less in size (bp) than the associated allele.
- The interpretation of mixtures follows a series of steps (3):

STEP 1: IDENTIFICATION OF A MIXED PROFILE

A mixture can only be identified if the alleles of the minor component are above the background noise. In practice the threshold is <1:10. A mixed profile consisting of more than one individual may be evident if a locus is observed with more than two peaks at a locus. However, extra bands or imbalanced peaks may be genetic or non-genetic. To make an objective assessment, all of the alternatives must be considered:

- Profiles with allelic artefacts e.g. stutters
- Non-specific artefacts
- Software - e.g. pull-up peaks
- Poor operator technique - e.g. lane to lane leakage
- Masking effect - when alleles from different individuals are superimposed
- Suppressed amplification of an allele - due to primer binding site mutation
- Promoted amplification of an allele - due to flanking region mutation
- Multiple banded or imbalanced profiles generated as a result of genetic phenomena such as trisomy, translocation and somatic mutation

Stutters

Whereas the great majority of artefacts can logically be excluded as non-allelic (8), it is not possible to exclude stutters since they are allelic products, differing structurally from the associated allele by just one repeat unit. It follows that a mixture with alleles from a low level contribution and stutters associated with the major contribution may be equivalent size and indistinguishable. Although stutters usually appear in pairs, this is not necessarily diagnostic. If an allele has non-consensus or partial sequences, then it will tend to stutter less than counterparts consisting of complete repeats (9).

Non-specific artefacts

Non-specific artefacts are generated as a result of priming from fragments of possibly degraded human, or bacterial DNA. When non-specific artefacts are found within the allelic region, the band shift test described by Gill *et al.* (10) is particularly useful to exclude non-specific artefacts; because they have a different sequence to STRs, they usually migrate atypically in the gel.

Software

One common problem is 'pull-up' (8). This is defined as a minor peak in a different colour directly below a major allele peak. Typically, a blue peak may pull up a green peak directly below it. This is only problematic if the minor peak is coincident with an potential allele. If there is a possibility either singleplexing or re-PCR are options to consider.

Poor operator technique

The commonest problem is leakage of a sample from one lane into the next. This is readily identified by reference to the scan data in ABD GS Analysis software.

Masking effect

A mixture may not always be evident by presence of multiple bands. This would occur in cases where the contributors to a mixture actually share alleles at a particular locus. Consider 2 individuals sharing the same alleles e.g. D18S51 14,14: 14,15; if the mixture ratio is 1:1 then the ratio of the 14:15 peak areas will be 3:1 respectively and pronounced peak asymmetry will be observed.

Suppression of amplification efficiency

Peak-area asymmetry outside the normal range for a non-mixture may occur because of a primer-binding site mutation. This has the effect of altering annealing and melting temperatures, which changes the amplification efficiency and decreases the resulting signal. If a substitution mutation occurs at the 5' end of the primer, a mismatch will result and amplification will fail completely, resulting in a null allele. The closer the substitution is to the 3' end of the primer binding site, the less the effect on the amplification efficiency.

Promotion of amplification efficiency

Sequence differences in flanking regions near to the PCR primer-binding site can actually improve amplification efficiency. At the HUMvWA locus, a sequence polymorphism is found associated with most HUMvWA 14 alleles, and to a much lesser extent with the HUMvWA 15 allele. The polymorphism consists of a substitution which is 3 bases from the 5' end of the primer binding site (in the amplification region). This appears to enhance amplification and may result in a peak area ratio >2:1.

Genetic anomalies

a) *Trisomy or translocations*: Both chromosomal duplication or gene duplication affect all cells in an individual. In practice it is impossible to tell the difference without resorting to genetic analysis. If duplication is

accompanied by a deletion or insertion of a repeat unit then 3 bands of similar size are generated (Fig. 1).

If a gene is duplicated without additional mutation, then just 2 bands are visible in a 2:1 ratio. In the example in Fig. 2, an XYY individual has a double dose of the Y gene. Note that other loci are balanced and this argues against possibility of a mixture. In the multiplex described by Sparkes *et al.* (7) trisomy or translocation was observed in 1 in 2000 at each locus.

Somatic mutation:

If a somatic mutation occurs during embryological development then two types of cells with different genotypes may coexist and this leads to a three-banded profile (Fig. 3). The peak areas will be dependent upon the relative proportion of the mutant cell and will not be equivalent. This is arguably the most difficult condition to elucidate since it is possible that not all tissues will demonstrate somatic mutation. On the National DNA database of England and Wales, the incidence of somatic mutation is variable - out of 120,000 samples none have been observed at the HUMTH01 locus, whereas the incidence is approximately 1/5000 at D18S51 and HUMFIBRA loci. It is possible that some somatic mutations will not be distinguishable from stutters, hence these figures are probably underestimates since they are only recorded if unambiguous.

The genetic phenomena described (trisomy, translocation, primer binding site mutations) can be verified by analysis of the reference sample, which should also demonstrate the same anomaly unless a tissue specific somatic mutation has occurred. In the latter case confirmation may depend upon a reference sample which has the same origin as the case stain, although perhaps we cannot completely rule out the possibility that appearance of somatic mutations could vary over time within tissues as the buccal lining which consists of rapidly dividing cells.

Can a simple mixture ever be confused with a non-mixture?

Although most simple mixtures can be identified by the presence of 3 or more alleles at several loci, it is relevant to ascertain if it is possible for a mixture to consist of no more than 2 alleles per locus. This would happen if masking occurred at every locus in a multiplex.

To estimate this chance, we carried out more than 212,000 pairwise comparisons[1] of our Caucasian frequency database to artificially generate mixtures from unrelated individuals (8). Most mixtures will show 15-22 bands across the 6 STR loci (Fig. 4). There were only four examples where only 1 or 2 alleles were observed at each locus, and even in these cases a mixture would be suspected since peak heights would be noticeably imbalanced. An example is given in Table 1. Only at the HUMTH01 locus, where both individuals share the same genotype will the locus appear to be balanced. Masking is more likely to occur with mixtures of relatives.

STEP 2: DESIGNATION OF ALLELIC PEAKS

Once an assessment of the mixed profile has been made, designation of allelic peaks or possible allelic peaks follows the principles described by Gill *et al.* (10). The positions of peaks (bp) are compared with allelic ladders and scored only if within guidelines:

- within 0.5 bp of the designated control allelic ladder marker.
- band shift (correlation) is consistent[2]

STEP 3: IDENTIFICATION OF THE POTENTIAL NUMBER OF CONTRIBUTORS

Once the most likely explanation for multiple allelic peaks and asymmetry has been attributed to the profile being a mixture, the next step is to estimate the number of contributors. The maximum number of alleles at any locus for a simple two-person mixture is 4 (given that no genetic phenomena are involved). In our experience more complex mixtures are relatively rare, but care must be taken not to confuse stutters with true alleles. Consideration of the circumstances of a case are often important in assessing the number of potential contributors - for example in a vaginal swab it would not be surprising to find a minor component from the victim[3].

STEP 4: ESTIMATION OF THE RELATIVE PROPORTION/RATIO OF THE INDIVIDUALS CONTRIBUTING TO THE MIXTURE

A mixture can range from the contributors being in equal proportion to each other, to one being in great excess. It is helpful to classify the mixture:

- Type A: No clearly defined major contributor
- Type B: Clearly defined major and minor contributors

We have demonstrated that if DNA templates are mixed (e.g. in the ratio 2:1) then this ratio will be approximately preserved throughout all loci compared (Gill *et al.* (11). Furthermore, the ratio can be estimated relatively easily when there are no shared alleles as follows:

Consider the GS Analysis profile of D18S51 in Fig. 5. It is possible to pair the alleles into minor components (14,15) and major components (16,18). The mixture proportion can be calculated ($\phi A+\phi B)/(\phi A+\phi B+\phi C+\phi D$) = 2840/7988 =0.36. Alternatively, a mixture ratio can be calculated ($\phi A+\phi B)/(\phi C+\phi D$) = 2:1. Once the mixture proportion or ratio is known then it is possible to estimate the peak areas for any given genotype combination[4]:

The amelogenin result from the same mixture (Fig. 6) can be used to independently check estimates of the mixture ratio or proportion. In addition, it is used to determine whether the major proportion of the mixture is male or female. In the above example, the relative peak area proportion of $\phi Y/(\phi X+\phi Y)$ = 0.17, suggesting a disproportionate male/female mixture where the male contribution forms the minor part. Mx[5] is estimated (3) as $2\phi Y/(\phi X + \phi Y)$=0.35; in agreement with that obtained previously from D18S51. We can conclude that the evidence supports the suggestion that the 14,15 phenotype has originated from a male. Additional confirmations of Mx can be carried out if other loci in the multiplex are 4-banded

Next, the analysis of loci having fewer than 4 bands follows under the assumption of Mx=0.35. Taking the three-band profile of D8S1179 (Fig. 7) as an example; the peak areas which support contentions for the likeliest male: female phenotype combinations respectively are 13,15: 13,14; 13,14:13,15 or 14,15:13,13. Other combinations such as 15,15:13,14 are excluded because the 13,14 heterozygote would be markedly imbalanced after comparison to the 60% guideline referred to previously[7].

STEP 5: CONSIDERATION OF ALL POSSIBLE GENOTYPE COMBINATIONS

Interpretation and analysis

At this stage, it would be relatively simple to compare the crime profiles derived from reference samples. However, this approach is not recommended - interpretation of the mixed profile should be derived independently, and without the possibility of introducing bias by prior knowledge of the suspects genotype. If we consider a four-banded profile, in the absence of peak area then the possibilities in Table 2 are considered (1,2).

There are 3 possible genotype combinations (and their reverse options) - provided that the mixture can be visually separated into major and minor components (Fig. 5) then normally only 1 possibility (or the reverse) remains. Under the assumption that the contributors to the mixture are the suspect (CD) and an unknown person (AB) in the numerator and two unknown people in the denominator, ignoring the peak area information the likelihood ratio is 1/12*fcf*d (2). Taking into account the peak area information in the denominator, only AB,CD or CD,AB are possible, hence the LR= 1/4 *fcf*d. Furthermore, if it is possible to condition on the victim's profile e.g. suppose that the crime sample is a vaginal swab and the victim is AB (the minor component). The major profile may be deduced as originating from a male provided that the relative proportion of the amelogenin Y peak supports this contention. The female genotype can effectively be removed to leaving just one possibility for the suspect in the denominator - the LR is 1/2*fcf*d.

The process is not as straight forward when 2 or 3-banded profiles are observed, since masking has occurred and this frequently gives rise to several possibilities after considering peak areas and conditioning on the victim's phenotype. Stutters only have an effect when the minor profile is the one of interest, and alleles and stutters are approximately the same size as each other.

Returning to the above example in Fig. 7, the possibilities to consider (conditioning on individual 2 as the major (female) contributor to the profile) are as follows:

individual 1 (male)	individual 2 (female)
14,15	13,13
13,15	13,14
13,14	13,15

By evaluating the evidence of the peak areas it is possible to rank the phenotype combinations in order of their strongest support. It is necessary to emphasise that this does not mean that the actual genotype will necessarily rank first, but such a procedure is useful to assist the reporting officer as part of the overall interpretation of a mixture.

We can use the proforma in Fig. 8 to work out expected peak areas of alleles 13,14 and 15 given the estimated Mx and phenotype combinations 14,15:13,13 or 13,15: 13,14 or 13,14:13,15.

By comparing the 3 different scenarios tested in Table 3, the contention that the mixture is phenotype 14,15:13,13 has the best support. The same principles can be used to evaluate any scenario, including those involving stutters.

Evett *et al.* (1998) describe a complex model to evaluate the evidence based on peak areas. Few assumptions are made about the phenotype combinations since all possible combinations (within the set of subjectively designated alleles) simultaneously contribute to the likelihood ratio. It has the advantage of allowing evaluation of the denominator across all possible mixture combinations. However,

absolute identification of alleles is assumed *a priori*, and it is also assumed that loci are free from amplification suppression, hence it is envisaged that the model will be primarily used when ambiguities such as stutters are absent or discounted using the guidance previously described.

STEP 6: COMPARE REFERENCE SAMPLES

Once all possible phenotypes have been evaluated and recorded by the reporting officer. a comparison is made with the reference samples, and where possible conditioning on the circumstances of the case follows. At this stage genetic anomalies such as trisomy or primer binding site mutation may be evident in the reference samples and this may in turn influence the concluding statement.

Incorporating stutters into the statistical analysis of mixtures

The treatment of stutters is considered in detail by Gill *et al.* (11). Whereas it is expected that the development of improved multimixes will result in reduced artefacts, it is unlikely that any electrophoretic system will ever be totally free. The main artefact to consider is the stutter. Stutters are 1 repeat unit less than the 'parent' allele. Generally, in the multimix discussed in this paper, they are less than 15% the size of the associated allele. However, different multimixes will have different characteristics which must be taken into account. An understanding of the characteristics of stutters is an important facet of mixture interpretation. Generally, stutters only become significant when the minor profile is the one of evidential significance.

Consider the following scenario: *Suspect matches both minor components one of which is in a stutter position and the major components match the victim.*

If there is a profile consisting of minor components A,B and major peaks C,D (Fig. 10) such that they are easily distinguished. If B is 1 repeat unit less than C and <15% the size of C then B may be a stutter or an allele. The likelihood ratio is:

[1]

$$LR = \frac{p(\phi A \phi B \phi C \phi D \mid H, S)p(S) + p(\phi A \phi B \phi C \phi D \mid \bar{H}, S)p(\bar{S})}{p(\phi A \phi B \phi C \phi D \mid \bar{H}, S)p(S) + p(\phi A \phi B \phi C \phi D \mid \bar{H}, \bar{S})p(\bar{S})}$$

$p(\phi A \phi B \phi C \phi D \mid H, S)$ is the probability of peak areas A,B,C,D given the hypothesis H and that the area of B is at least in part a stutter.

$p(\phi A \phi B \phi C \phi D \mid H\bar{S})$ is the probability of peak areas A,B,C,D given the hypothesis H and that the area of B is at least in part an allele.

$p(\phi A \phi B \phi C \phi D \mid \bar{H}, S)$ is the probability of peak areas A,B,C,D given the hypothesis $\bar{H}$ and that the area of B is at least in part a stutter.

$p(\phi A \phi B \phi C \phi D \mid \bar{H}, \bar{S})$ is the probability of peak areas A,B,C,D given the hypothesis $\bar{H}$ and that the area of B is at least in part an allele.

When the suspect is AB and the victim is CD

$$[2]\ LR = \frac{1}{[fa^2 + 2fafb + 2fafc + 2fafd]p(S) + [2fafb]p(\bar{S})}$$

If p(S) is unknown then the lower bound of the LR can be calculated:

$$[3]\ LR \geq \frac{1}{fa^2 + 2fafb + 2fafc + 2fafd}$$

i.e. the LR calculation above is always conservative compared to formula [2] if B must be an allele in the numerator.

Given the profile in Fig. 10, consider the condition that the suspect is AC and the victim is CD; now the probability that B is a stutter must be included in the numerator.

Suspect matches one minor component and one major component; the second minor component is in a stutter position

$$[4]\ LR = \frac{p(S)}{[fa^2 + 2fafb + 2fafc + 2fafd]p(S) + [2fafb]p(\bar{S})}$$

Under the condition where B must be a stutter in the numerator, p(S) remains. This means that it would not be conservative to use formula [3]. The probability of the stutter can be derived from a probability density function based upon the distribution of stutter areas of non-mixtures.

CONCLUSION

Interpretation of mixtures takes place against a background of artefacts and genetic phenomena which must be assessed beforehand. The most important to consider are stutters since they can be indistinguishable from alleles derived from the minor component of a mixture. Provided

a series of logical steps are followed, the possible phenotypes can be elucidated based on analysis of peak areas and these can be incorporated into the likelihood ratio. To avoid bias, it is recommended that reference samples are not assessed until the case stain has been evaluated and the possible phenotypes noted. Mutations such as trisomy or translocation which affect the profile of the case stain will also be seen in the reference sample unless a tissue-specific mutation has occurred.

A case will usually comprise several stains, not all of these may be mixtures. In addition the proportions of a mixture will often vary across the stain itself. Ambiguity may be solved by running additional samples. Interpretation is based on the case in its entirety and the circumstances of the case may allow conditioning to take place.

ENDNOTES

[1] From a frequency database of size *n*, simulate a mixture of first and 2nd samples. Then simulate admixture of first with the third sample and so on until the end of the database is reached. Then return to the second sample and proceed until mixtures have been simulated across the entire set. This results in *n*(n-1)/2* comparisons.

[2] If the distance between the allelic ladder markers and the sample alleles are δ_1 and δ_2 respectively, the band shift is calculated as c=δ_1 and δ_2. If the band shift is greater than 0.5 then the sample is examined for possibility of rare allele, or abnormal electrophoretic conditions resulting in anomalous band shift.

[3] Strictly we should allow for all possible combinations of the number of contributors and the genotypes of contributors regardless of whether a peak exists or not at any given allelic position. For instance the combination BB;CD should be considered even when the peaks present are at ABCD. These are then weighted according to how well they explain the peak areas. In the example cited above the explanation would require modelling the possibility that the area at A is stuttering from B. We do not attempt to implement here this purist approach. Rather we assign the weight 1 to all reasonable explanations of the evidence. We believe that this subjective approach is "safe" whenever the prosecution hypothesis explains the peak areas as well.

[4] This portion of our suggested algorithm is to assist the operator in determining whether the peak areas are likely under the hypothesis in question, in which case they are assigned subjectively a weight of 1, or unlikely in which case they are assigned a weight of 0.

[5] We point out that estimating Mx is not an endpoint in the interpretation. rather it is an intermediate step to assist the operator in subjective assessment as to whether the peak areas (at either suggested allelic or stutter positions) are likely.

[6] In the example above both n and 'n+1' peaks are observed (these are combined to give the total peak area).

[7] Bearing in mind the (unlikely) possibility that allele 14 might be suppressed because of primer binding site mutation - however this can be verified by comparison to the reference sample.

REFERENCES

1. Evett I.W., Buffery C., Willot G and Stoney D.A., (1991) A Guide to Interpreting Single Locus Profiles of DNA Mixtures in Forensic Cases. *J. Forensic Sci.*, **31**:41-47.

2. Weir B.S., Triggs C.M., Starling L., Stowell K.A.J., Buckleton, J. (1997) Interpreting DNA Mixtures. *J. Forensic Sci.*, **42**: 213-222.

3. Clayton T.M, Whitaker J.P., Sparkes R., Gill P., (1998) Analysis and Interpretation of Mixed Forensic Stains Using DNA STR Profiling. *Forensic Sci. Int.*, **91**: 55-70.

4. Evett I.W., Gill P., Lambert J.A. (1997) Taking Account of Peak Areas When Interpreting Mixed DNA Profiles. *J. Forensic Sci.*, **43**:62-69.

5. Gill P., Sparkes R., Pinchin R., Clayton T., Whitaker J. and Buckleton J., (1998) Interpreting Simple STR Mixtures Using Allele Peak Areas. *Forensic Sci. Int.*, **91**:41-53.

6. Bar W., Brinkmann B., Budowle B., Carracedo A., Gill P., Lincoln P., Mayr W., Olaisen B., (1997) DNA Recommendations. Further Report of the DNA Commission of the ISFH Regarding the Use of Short Tandem Repeat Systems. International Society for Forensic Haemogenetics. *Int. J. Legal Med.*, **110**(4): 175-176.

7. Sparkes R., Kimpton C., Watson S.K., Frazier R.R.E., Johnson P.E., Millican E.S., Gill P., (1997) The Validation of a 7-Locus Multiplex Test for Use in Forensic Casework (I) Mixtures, Ageing, Degradation and Species Studies. *Int J. Leg. Med*, **109**:186-194.

8. Gill P., Sparkes R., Kimpton C., (1997) Development of Guidelines to Designate Alleles Using an STR Multiplex System. *Forensic Sci. Int.* **89**;185-197.

9. Walsh P.S., Fildes N.J., Reynolds R., (1996) Sequence Analysis and Characterization of Stutter Products at the Tetranucleotide Repeat Locus vWA. *Nucleic Acids Res*, **24**: 2807-2812.

10. Gill P., Urquhart A., Millican E., Oldroyd N., Sparkes R and Kimpton CP., (1996) A New Method of STR Interpretation Using Inferential Logic - Development of a Criminal Intelligence Database. *Int. J. Legal Med.*, **109**:14-22.

11. Gill P., Sparkes B., Buckleton J.S., (1998, in press) Interpretation of Simple Mixtures When Artefacts Such as Stutters are Present - with Special Reference to Multiplex STRs Used by the Forensic Science Service. *Forensic Science International*.

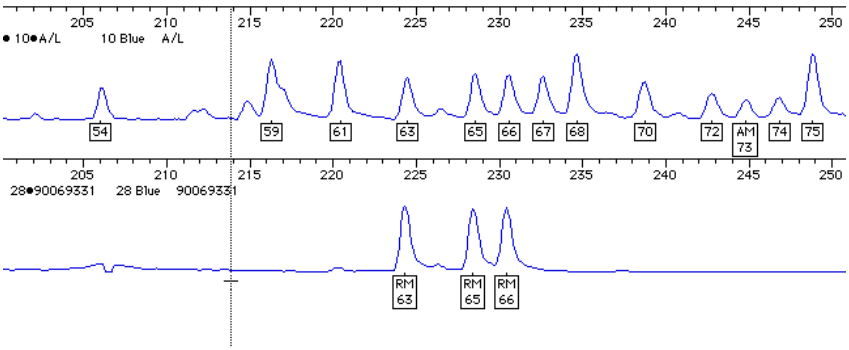


Figure 1:*D21S11 trisomy or translocation in the lower pane. Note that the bands are equivalent in size. Allelic ladder in the upper pane.*

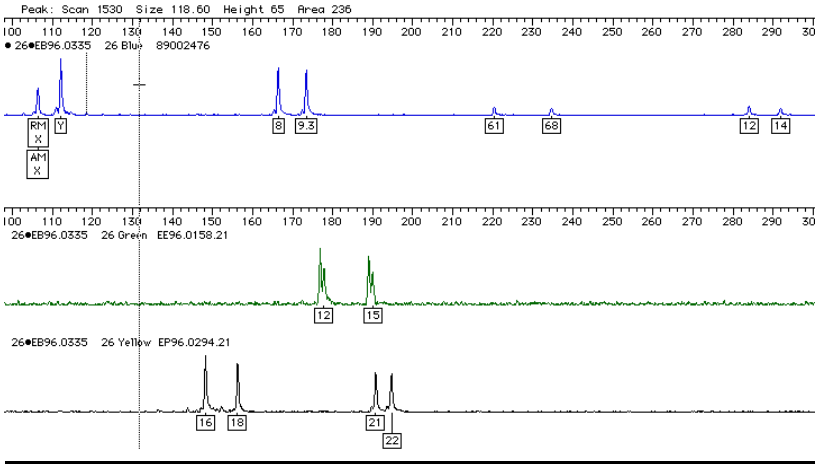


Figure 2: *XYY individual, upper pane left, showing a Y peak twice the size of the X peak. The remaining loci of the multiplex are balanced.*

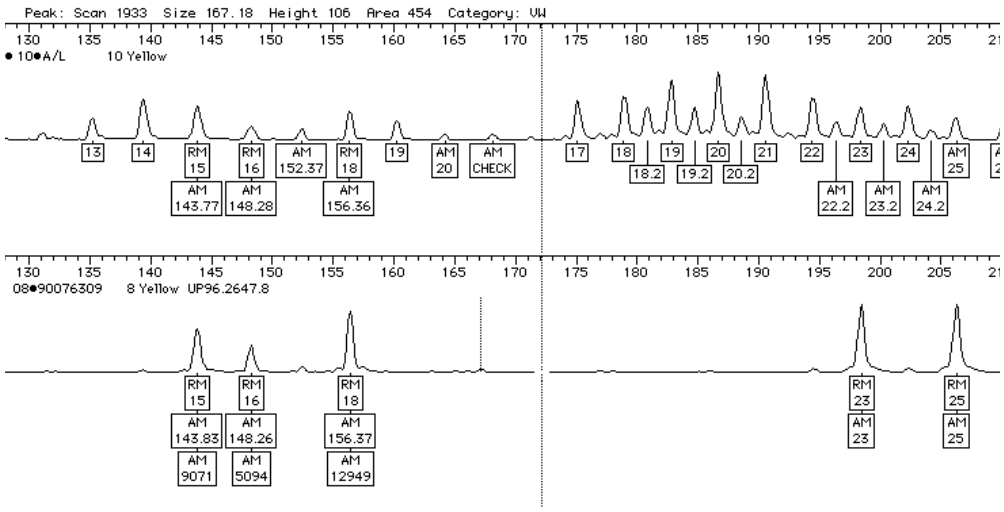


Figure 3: *Somatic mutation of HUMvWA, lower left pane. Note three peaks are present of different sizes. HUMFIBRA/FGA peaks are shown on the right side. The upper pane shows HUMvWA and HUMFIBRA allelic ladders.*

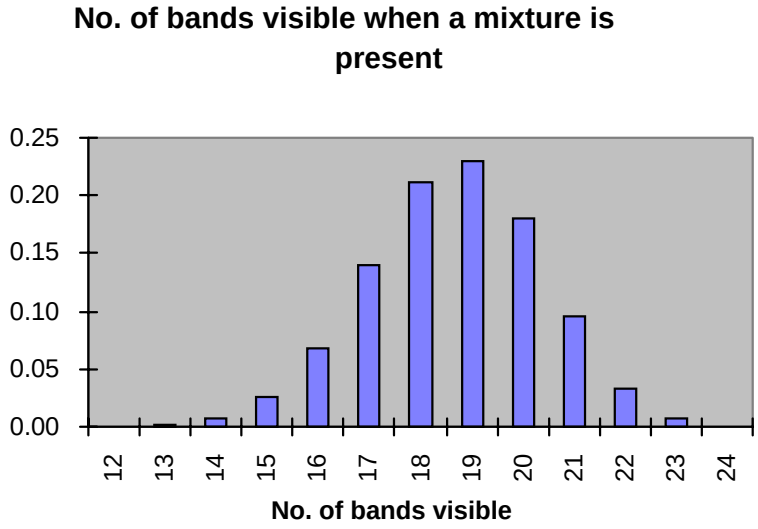
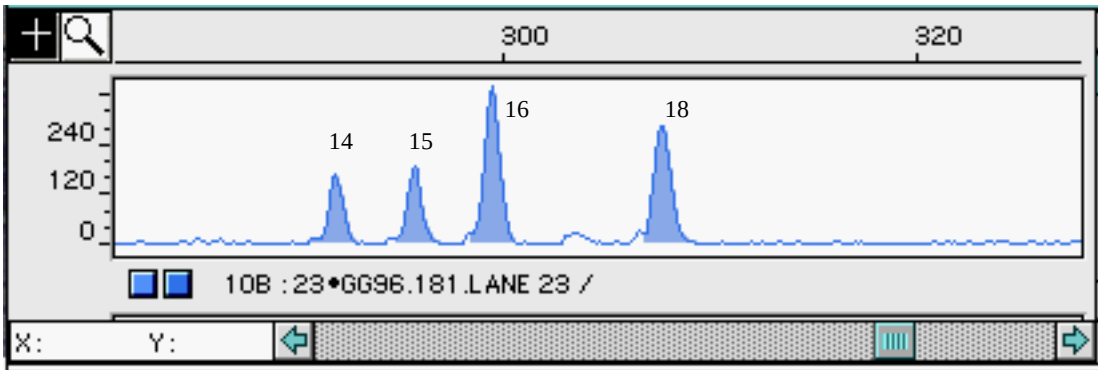


Figure 4: *Result of pairwise comparison simulation of simple mixtures. For example, approximately 23% of mixtures will show 19 bands (excluding amelogenin).*

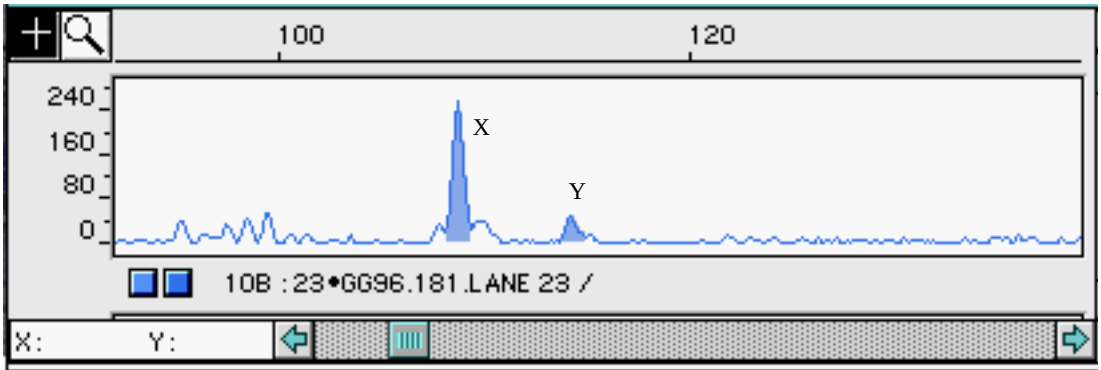
Table 1: *In this mixture, the components do not give rise to more than 2 bands at any locus. However, only the HUMTH01 locus will appear balanced since there are identical phenotypes from each contributor*

Locus	D18S51	D18S51	D21S11	D21S11	HUMTH01	HUMTH01	D8S1179	D8S1179	HUMFIBRA	HUMFIBRA	HUMVWA	HUMVWA
Allele	1	2	1	2	1	2	1	2	1	2	1	2
Allele designations (1)	14	14	61	63	8	9.3	8	13	23	23	18	19
Allele designations (2)	14	17	63	63	8	9.3	13	13	20	23	18	18



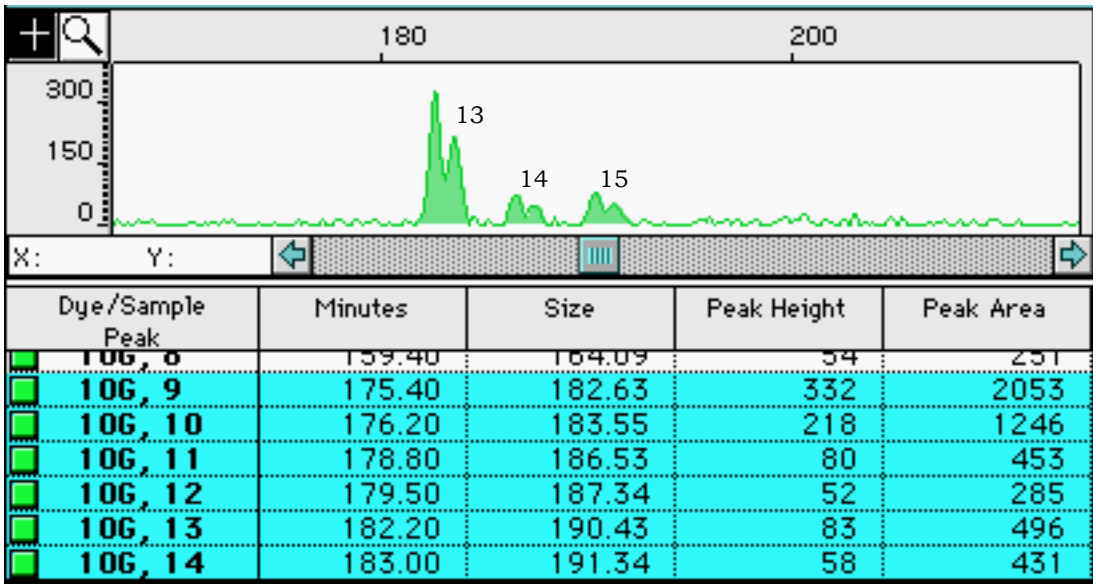
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area
10B, 23	270.80	291.87	171	1375
10B, 24	274.30	295.71	192	1465
10B, 25	277.70	299.45	390	2867
10B, 26	281.10	303.49	31	226
10B, 27	283.70	306.60	33	142
10B, 28	284.60	307.68	291	2281

Figure 5: *ABD GS Analysis of D18S51 mixture showing heterozygotes from two different individuals.*



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area
10B, 7	114.10	107.70	36	176
10B, 8	114.90	108.69	266	1265
10B, 9	115.90	109.93	44	314
10B, 10	119.40	114.25	52	268
10B, 11	150.10	152.99	46	167

Figure 6: *Amelogenin, showing imbalanced X:Y peaks, typical of a male/female mixture.*



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area
10G, 8	159.40	164.09	54	251
10G, 9	175.40	182.63	332	2053
10G, 10	176.20	183.55	218	1246
10G, 11	178.80	186.53	80	453
10G, 12	179.50	187.34	52	285
10G, 13	182.20	190.43	83	496
10G, 14	183.00	191.34	58	431

Figure 7: *Three banded profile of DS1179.*

Table 2: *A list of all possible combinations of a four-banded profile, ignoring peak area*

Individual 1	Individual 2
A,B	C,D
A,C	B,D
A,D	B,C
C,D	A,B
B,D	A,C
B,C	A,D

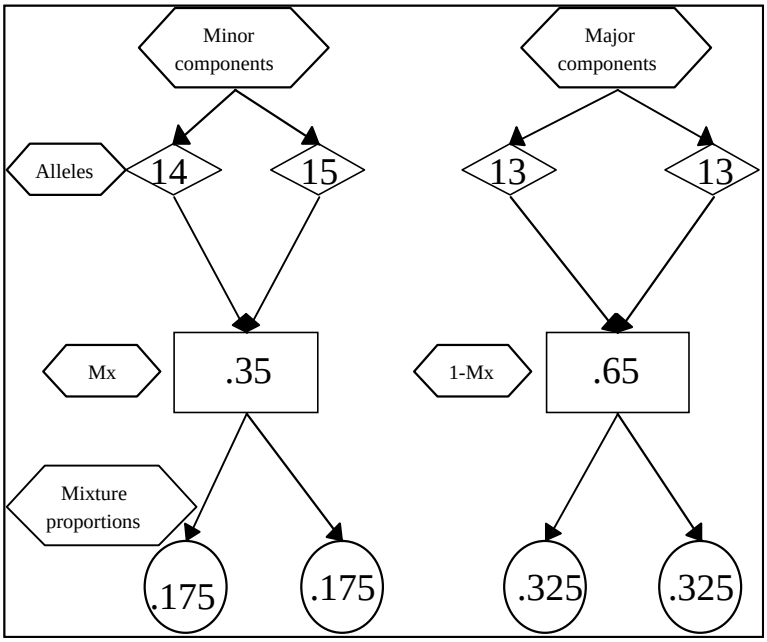


Figure 8: *If the minor:major phenotypes are 14,15 : 13,13 respectively, Mx is 0.35 then the expected relative proportions of 13:14:15 are 0.65:0.175:0.175*

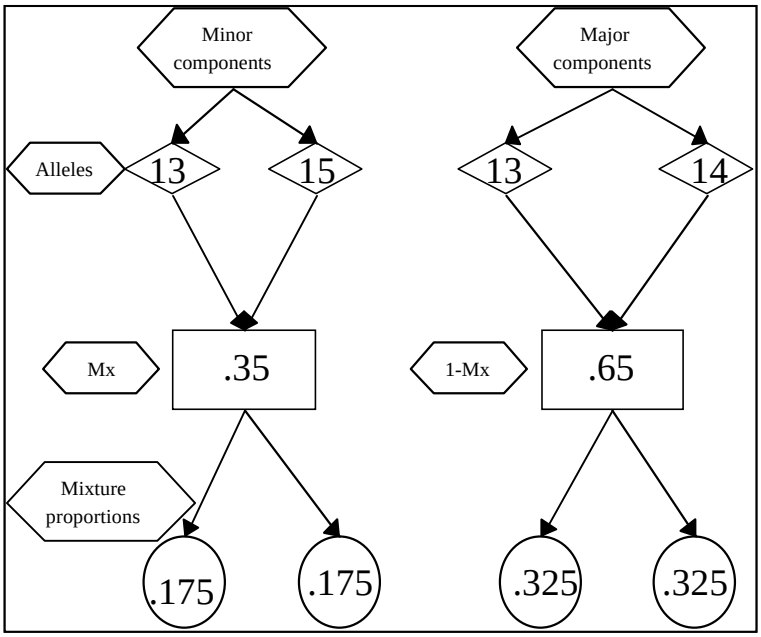


Figure 9: *If the minor:major phenotypes are 13,15:13,14 and Mx=0.35 then the expected proportions of 13:14:15 are 0.5:0.325:0.175 (expected proportion of the allele13 peak area is 0.175 + 0.325 = 0.5).*

Table 3: *For each allele the observed peak areas are taken from ABD GS Analysis software; expected peak areas are derived from expected allele proportions from Figs 8 and 9. The differences between the two sets of figures are absolute (ignore sign).*

Allele	**13**	**14**	**15**	**If the phenotype is:**
Observed	**3299**	**738**	**927**	
Expected	**3226**	**869**	**869**	**14,15:13,13**
Difference	**74**	**131**	**58**	
Allele	**13**	**14**	**15**	**If the phenotype is:**
Observed	**3299**	**738**	**927**	
Expected	**2482**	**1613**	**869**	**13,15:13,14**
Difference	**817**	**875**	**58**	
Allele	**13**	**14**	**15**	**If the phenotype is:**
Observed	**3299**	**738**	**927**	
Expected	**2482**	**869**	**1613**	**13,14:13,15**
Difference	**817**	**131**	**686**	

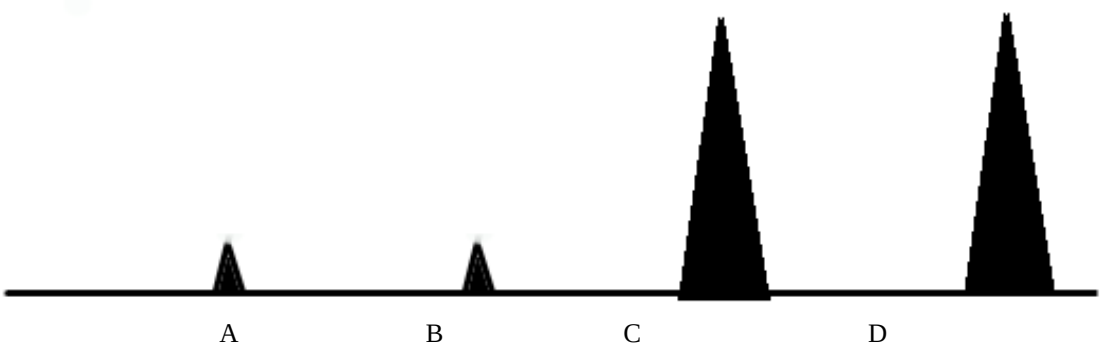


Figure 10: *Schematic diagram of a profile comprising two minor bands A,B and two major bands C,D. The minor bands are <15% the area of the major bands and the distance between B and C is 1 repeat unit and could be a stutter.*