



Guide to STR Interpretation - mixtures and allelic artefacts

Peter Gill



Dont do it!

- Cases often have multiple stains
- select those which are not mixtures
- ensure that any mixtures are consistent with case work circumstances.
- Dont feel you have to do a statistical analysis



Statistical analysis of mixtures

- Assume alleles have been identified unambiguously ($P=1$).
- Is this realistic?



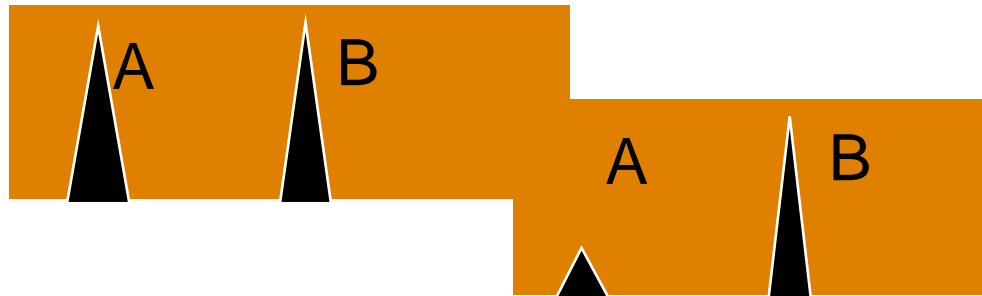
Nomenclature

	Mixture components	TH01	D21	D18	D8	VWA
Mixture 1	MT	8,9.3	67,70	11,11	9,17	16,18
	NO	9.3,9.3	59,65	17,19	9,11	17,19
	Mixture code	AB,BB	CD,AB	AA,BC	AC,AB	AC,BD
Mixture 2	EM	9.3,9.3	70,70	11,19	13,15	17,18
	NO	9.3,9.3	59,65	17,19	9,11	17,19
	Mixture code		CC,AB	AC,BC	CD,AB	AB,AC



Guidelines

- Understanding the characteristics of non-mixtures before interpreting mixtures is a prerequisite.



- Heterozygotes
- Stutters
- Artefacts
- Genetic phenomena
- Gill et al (1997) Development of guidelines...
Forensic Sci Int. 89:185-197



Parameters

heterozygous balance and stutter

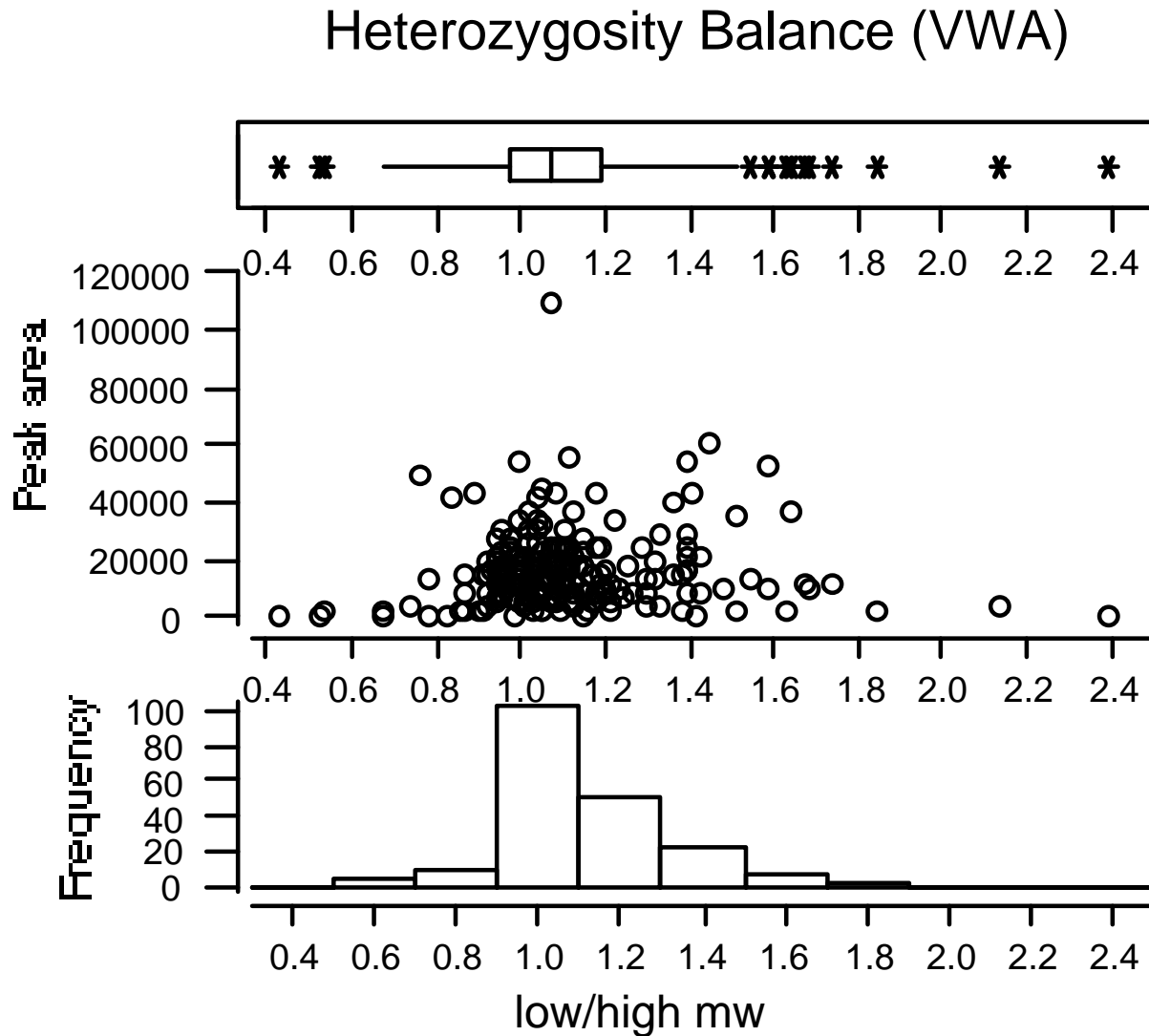
- Hb= low molecular weight allele (\emptyset) 'n' + 'n+1' band
high molecular weight allele (\emptyset) 'n' + 'n+1' band

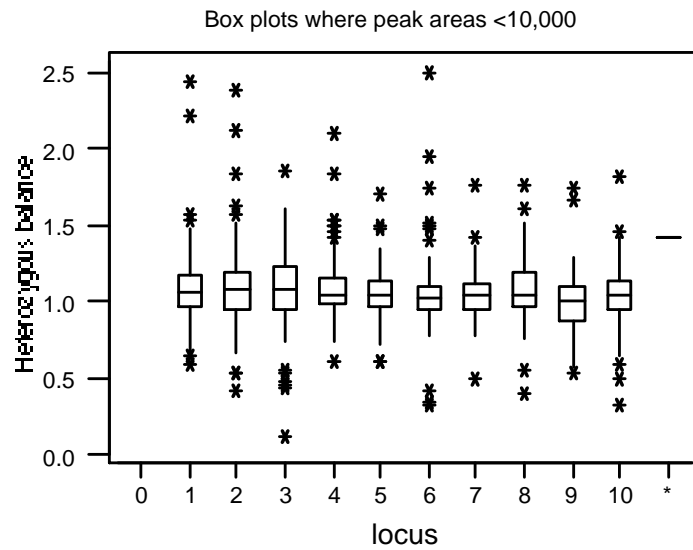
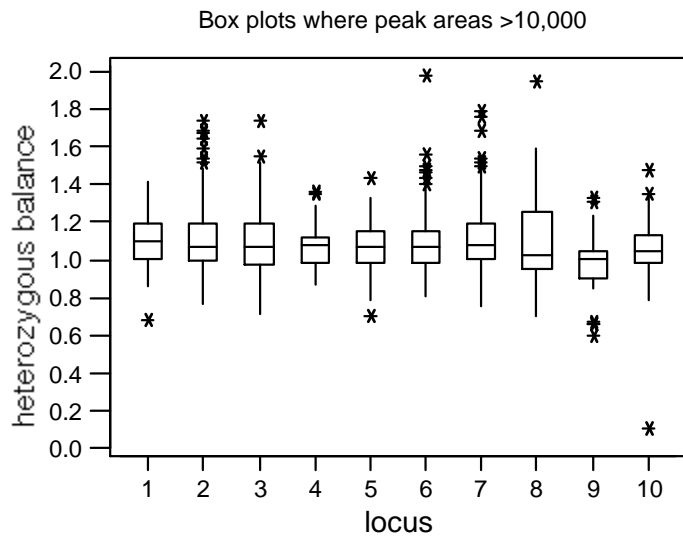
$$Sr = \frac{\emptyset S}{\emptyset('n') + \emptyset('n+1')}$$

- Also measure how often stutter occurs
- Also measure 'n' band prevalence/area
- Also measure interlocus variation

Forens. Sci. Int. 108 (2000) 1-29

Guidelines depend on characteristics of loci







Stutters or alleles?

- Following this strategy means that stutters will be scored as possible alleles
- Stutters have the following characteristics
 - 4bp less than a major allele
 - 15% the size of the major allele
 - They usually appear in pairs (check below threshold level if only 1 has been scored)



Why do we do this?

- To interpret mixtures
- To identify loci where potential drop-out has occurred (limited or degraded samples)
- To ensure that allelic artefacts and non-allelic artefacts are not confused with alleles.



Step 1 - Identify the presence of a mixture.

1) By the presence of extra bands

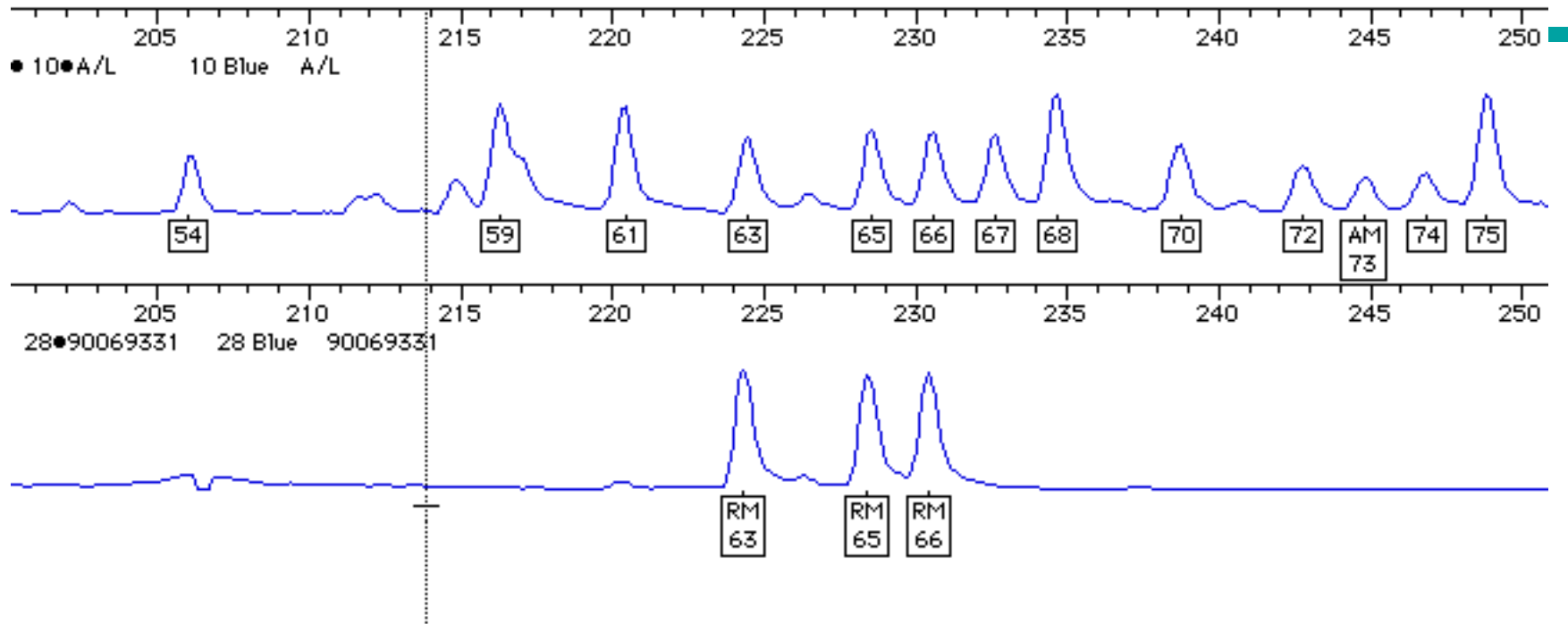


Fig1: *D21S11* trisomy or translocation in the lower pane. Note that the bands are equivalent in size. Allelic ladder in the upper pane.

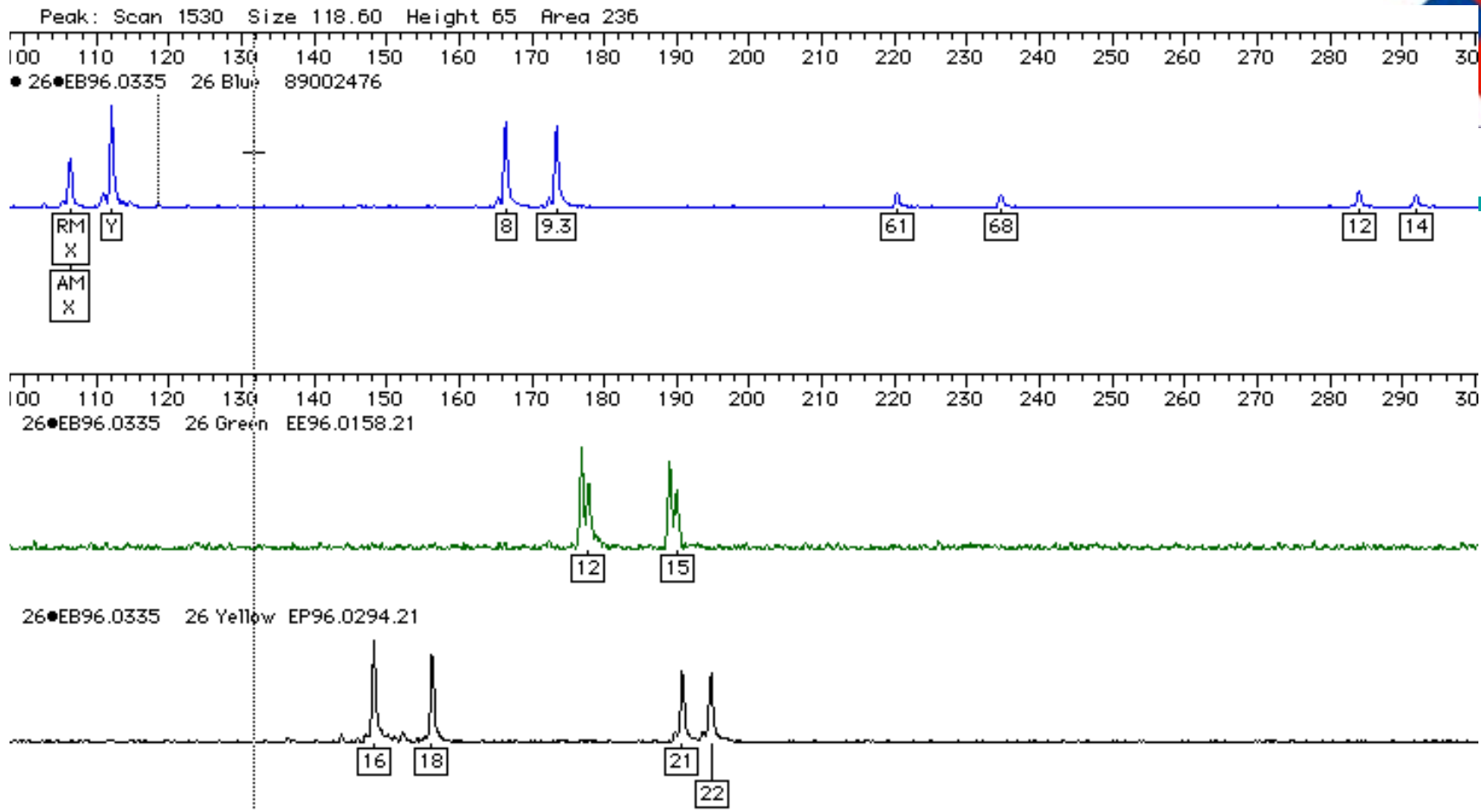
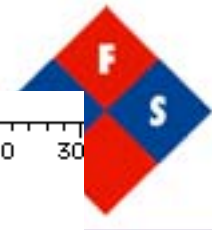


Fig 2: *XYY* individual, upper pane left, showing a *Y* peak twice the size of the *X* peak. The remaining loci of the multiplex are balanced.

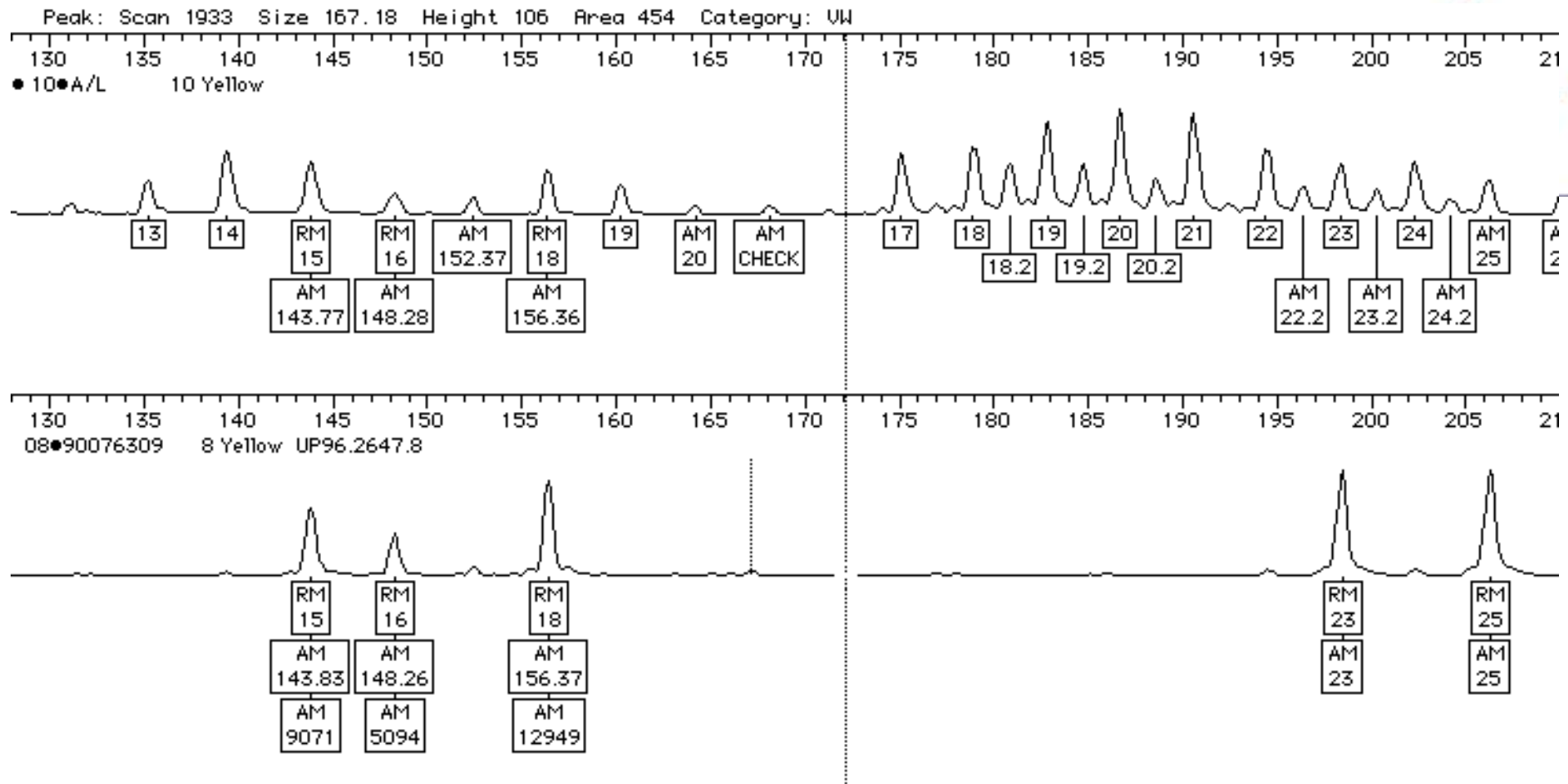


Fig 3: Somatic mutation of *HUMVWA*, lower left pane. Note three peaks are present of different sizes. *HUMFIBRA/FGA* peaks are shown on the right side. The upper pane shows *HUMVWA* and *HUMFIBRA* allelic ladders.



How often are genetic mutations observed?

Trisomy

Locus	Frequency (in c. 600,000 profiles)
Amelogenin	1191
D21S11	9
D18S51	7
D8S1179	24
FGA	12
vWA	8
TH0	1



Somatic mutations

Locus

Frequency

(in c. 600,000 profiles)

D21S11

58 (56 x 4bp; 2x > 4 bp)

D18S51

92 (77 x 4bp; 15 x > 4 bp)

D8S1179

23 (18 x 4bp ; 5 x > 4 bp)

VWA

48 (44 x 4bp ; 4 x > 4bp)

TH01

4 (2 x 4 bp; 2 x > 4 bp)

FGA

62 (49 x 4bp; 13 x > 4 bp)



Characteristics of somatics

- Somatics can be tissue specific
- Hairs have a high rate
- Some individuals might be particularly prone to SM - record is 7 mutated alleles in one persons mouth
- Many somatics missed as stutters
- Not inherited (unless gametic)

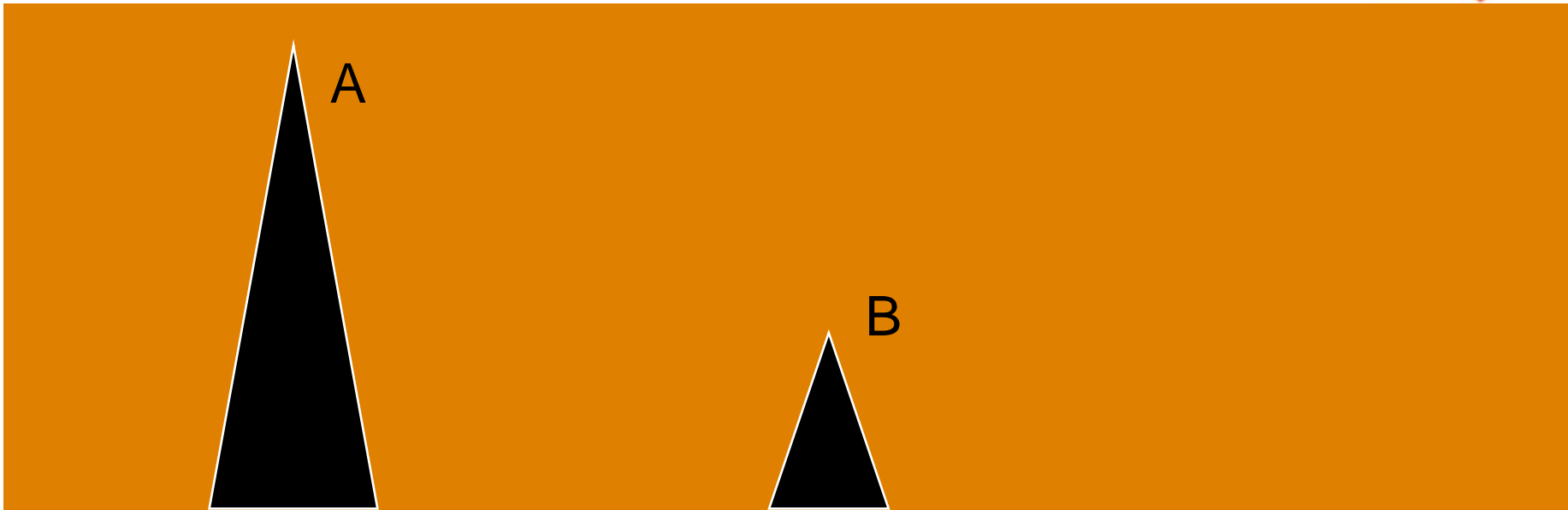


Step 1 - Identify the presence of a mixture.

2) By peak imbalance



Peak imbalance

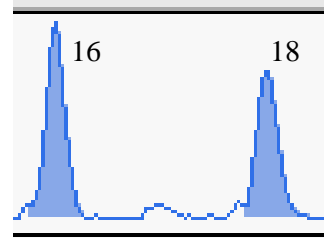


- *Consider one locus - if the mixture is 1:1*
- *The phenotype is AA, AB (**Masking has occurred**)*
- *There are 3 parts A to 1 part B hence the peak area ratios are 3:1*

Natural variation in heterozygote peak areas



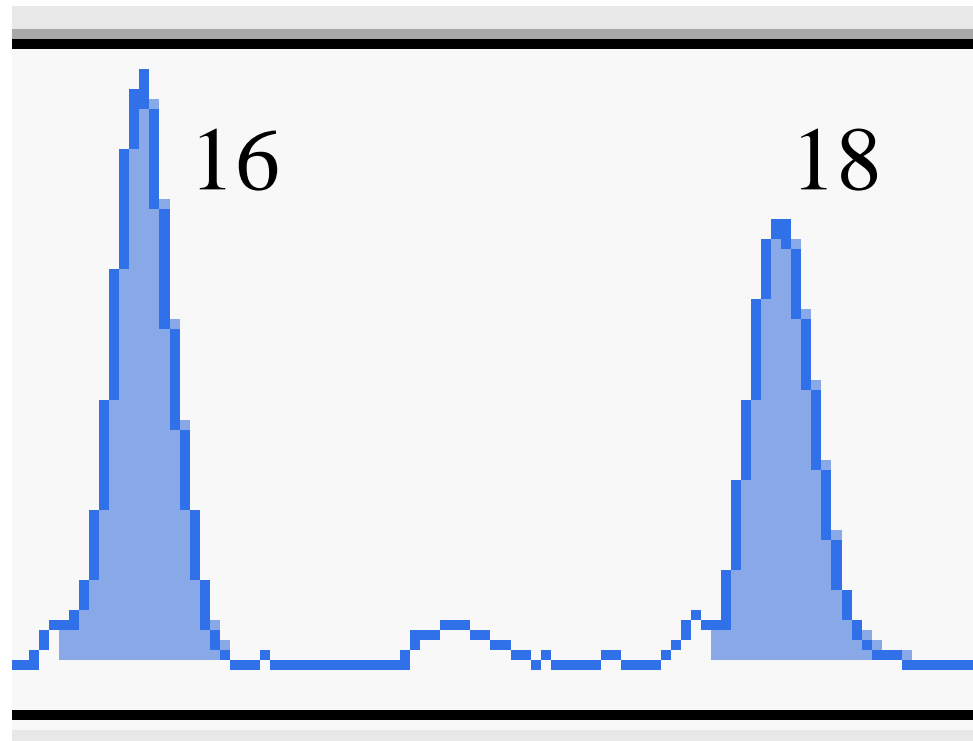
- Some variation between heterozygote peaks due to amplification efficiency
- Tendency for low molecular weight peak to be bigger (not always)
- Variation between loci
- General guideline - smaller peak area (A) is within 60% area of the larger peak (B)
- ie $(A/B)\% > 60\%$





Typical heterozygote imbalance

- General guideline - smaller peak > 60% size larger peak or should be $< 1/0.6 = 1.66$
- PCR less efficient for hmw allele





Genetic causes of peak imbalance

- Mutation at the primer binding site
- If at the 3' end of the primer then amplification is inhibited completely resulting in a null allele
- Elsewhere, amplification will be suppressed



Genetic phenomena are rare but:

- Check reference samples
- May not help with somatic mutation since body fluid/tissues may be different
- Only 1 locus will be affected

Is it possible for a non-mixture to be confused with a mixture?



- A mixture may be identified by presence at 3 or 4 bands at each locus
- Masking will occur - this happens when two individuals share alleles
- Therefore it is possible for a mixture to have just one or two alleles at a locus
- is it possible for only 1 or 2 alleles to be seen at every locus in the multiplex?

Is it possible for a non-mixture to be confused with a mixture?

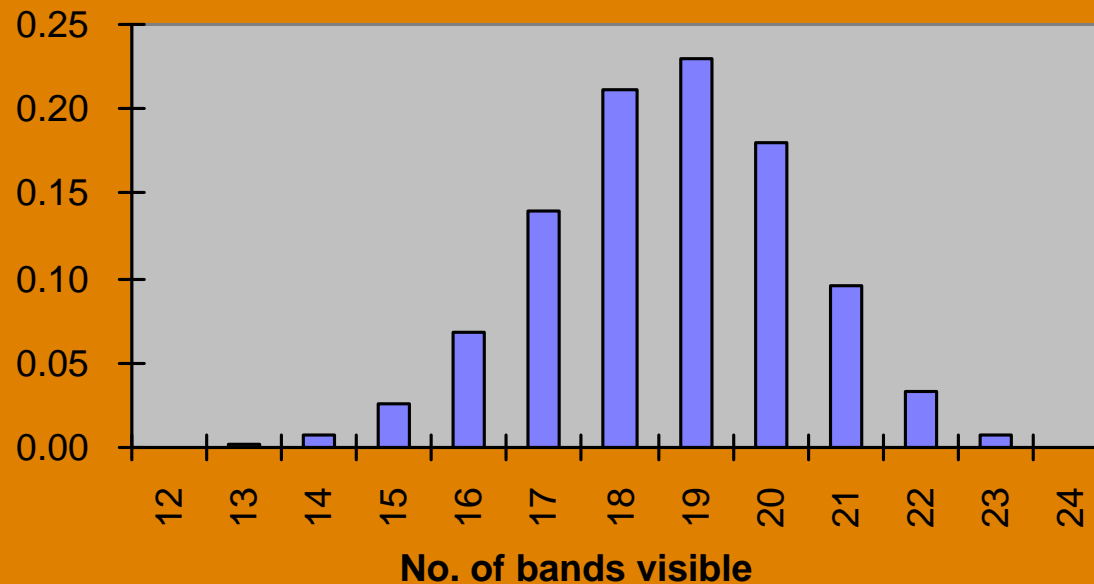


- To answer this question we carried out 200,000 pairwise comparisons of our frequency database - effectively simulating 200,000 cases where simple mixtures were observed from random members of the population
- SGM system – 6 loci



No of bands visible when a simple mixture is present

No. of bands visible when a mixture is present



No of bands visible when a simple mixture is present



- Only 4 samples out of 200,000 showed 1 or 2 alleles per locus
- e.g.

Locu	D 1 8 S 5 1	D 1 8 S 5 1	D 2 1 S 1 1	D 2 1 S 1 1	HUMTH01	HUMTH01
Allele	1	2	1	2	1	2
Allele designations (1)	14	14	61	63	8	9.3
Allele designations (2)	14	17	63	63	8	9.3

- Note imbalance. If mixture is 1:1 then peaks for 2 loci will show 3:1 peak area imbalance. Only THO is balanced



Forensic Bioinformatics Article

J Forensic Sci, Nov. 2005, Vol. 50, No. 6
Paper ID JFS2004475
Available online at: www.astm.org

David R. Paoletti,¹ M.S.; Travis E. Doom,^{1,2} Ph.D.; Carissa M. Krane,³ Ph.D.;
Michael L. Raymer,^{1,2} Ph.D.; and Dan E. Krane,⁴ Ph.D.

Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures

TABLE 2—Count and percent of three-person mixtures in which a particular number of unique alleles was the maximum observed across all loci, both for the original and randomized individuals[†].

Unique Alleles	Count	Percent (%)
2	0	0.00%
3	78	0.00%
4	4,967,034	3.39%
5	93,037,010	63.49%
6	48,532,037	33.12%

with 3-person mixtures



Recent Article by Buckleton et al.



Available online at www.sciencedirect.com



Forensic Science International: Genetics 1 (2007) 20–28



www.elsevier.com/locate/fsig

Towards understanding the effect of uncertainty in the number of contributors to DNA stains

John S. Buckleton^a, James M. Curran^{b,*}, Peter Gill^c

^a*The Institute of Environmental Science and Research Ltd., Private Bag 92021, Auckland, New Zealand*

^b*Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand*

^c*The Forensic Science Service, Trident Court, Solihull Parkway, Birmingham Business Park, Solihull B37 7YN, UK*

Received 31 May 2006; received in revised form 12 September 2006; accepted 13 September 2006

Abstract

DNA evidence recovered from a scene or collected in relation to a case is generally declared as a mixture when more than two alleles are observed at several loci. However, in principle, all DNA profiles may be considered to be potentially mixtures, even those that show not more than two alleles at any locus. When using a likelihood ratio approach to the interpretation of mixed DNA profiles it is necessary to postulate the number of potential contributors. However, this number is never known with certainty. The possibility of a, say three-person mixture, presenting four or fewer peaks at each locus of the CODIS set was explored by Paoletti et al. [D.R. Paoletti, T.E. Doom, C.M. Krane, M.L. Raymer, D.E. Krane, Empirical analysis of the STR profiles resulting from conceptual mixtures, *J. Forensic Sci.* 50 (2005) 1361–1366]. In this work we extend this analysis to consider the profiler plus and SGM plus multiplices. We begin the assessment of the risk associated with current practice in the calculation of LR's. We open the discussion of possible ways to surmount this ambiguity.

© 2006 Elsevier Ireland Ltd. All rights reserved.

Two-Person Mixtures for Simulated Profiles:

Probability by Locus of A Particular Number of Alleles Being Observed



Table 1

The probability of observing a given number of alleles in a two-person mixtures for simulated profiles at the SGMTM loci

Loci	No. of alleles			
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441
D8	0.011	0.194	0.521	0.274
D21	0.007	0.147	0.505	0.341
D18	0.003	0.095	0.472	0.430
D19	0.020	0.261	0.516	0.203
THO	0.016	0.271	0.547	0.166
FGA	0.003	0.116	0.500	0.381

Three-Person Mixtures for Simulated

Profiles: Probability by Locus of A Particular Number of Alleles Being Observed



Table 2

The probability of observing a given number of alleles in a three-person mixtures for simulated profiles at the SGM⁺™ loci

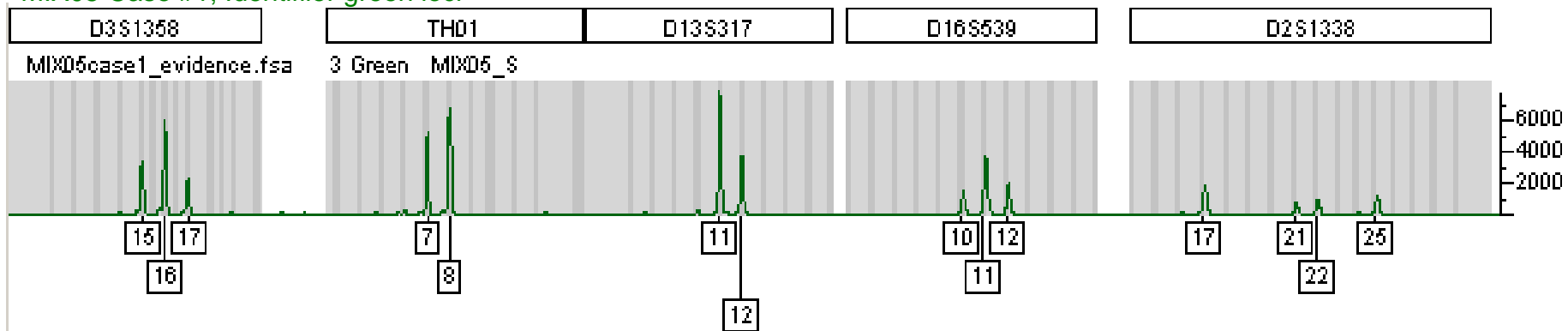
Loci	No. of alleles showing					
	1	2	3	4	5	6
D3	0.000	0.053	0.366	0.463	0.115	0.002
vWA	0.000	0.037	0.285	0.468	0.194	0.016
D16	0.001	0.086	0.397	0.411	0.100	0.005
D2	0.000	0.008	0.104	0.385	0.393	0.110
D8	0.001	0.041	0.258	0.436	0.236	0.029
D21	0.000	0.023	0.192	0.428	0.302	0.055
D18	0.000	0.007	0.109	0.392	0.396	0.096
D19	0.003	0.078	0.352	0.401	0.152	0.014
THO	0.001	0.074	0.395	0.439	0.088	0.002
FGA	0.000	0.012	0.144	0.424	0.346	0.074

Levels of Locus Heterozygosity Impact Number of Alleles Observed in Mixtures



Loci	No. of alleles			
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441

MIX05 Case #1; Identifiler green loci



**3 peaks more
common for D3**

**4 peaks more
common for D2**

*Number of Alleles Observed with Simulated **Four-Person Mixtures***

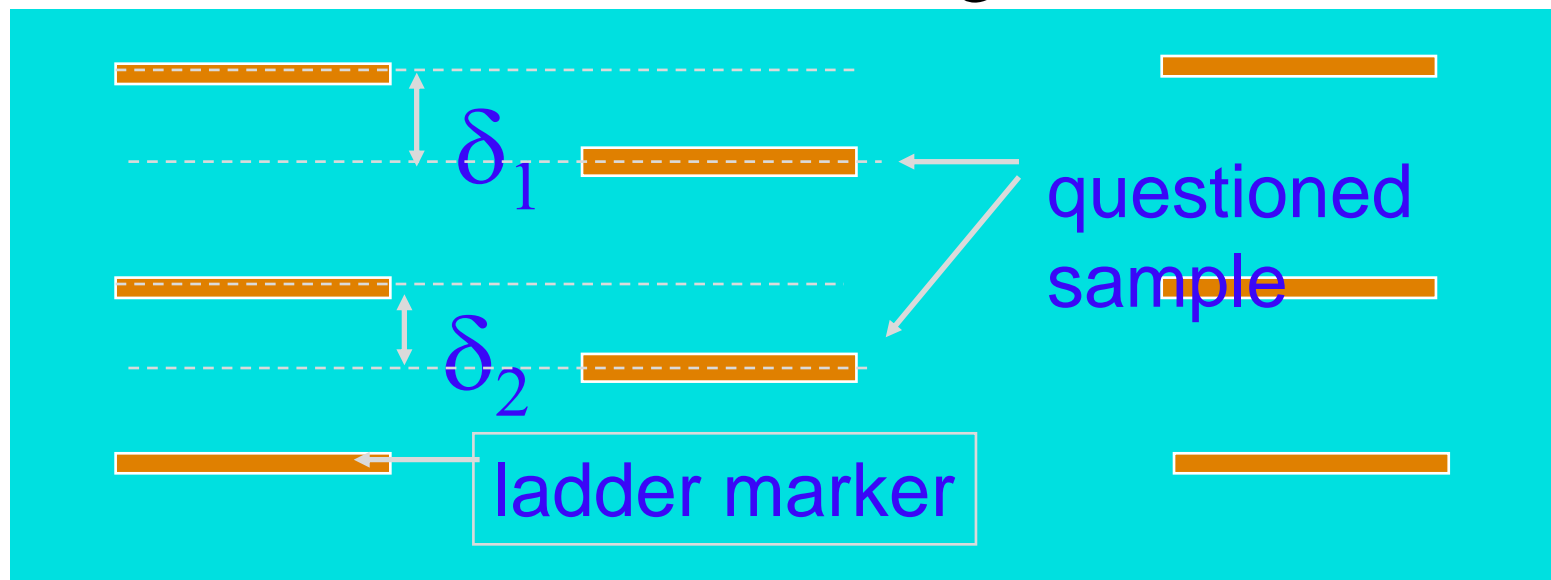


- The simulation of four person mixtures suggests that 0.014% of four person mixtures would show four or fewer alleles and that 66% would show six or fewer alleles for the SGM Plus loci.
- The results for the Profiler Plus loci were 0.6% and 75%.
- The equivalent values for the CODIS set from Paoletti et al. were 0.02% showing four or fewer and 76.35% showing six or fewer.



Step 2: Designation of allelic peaks

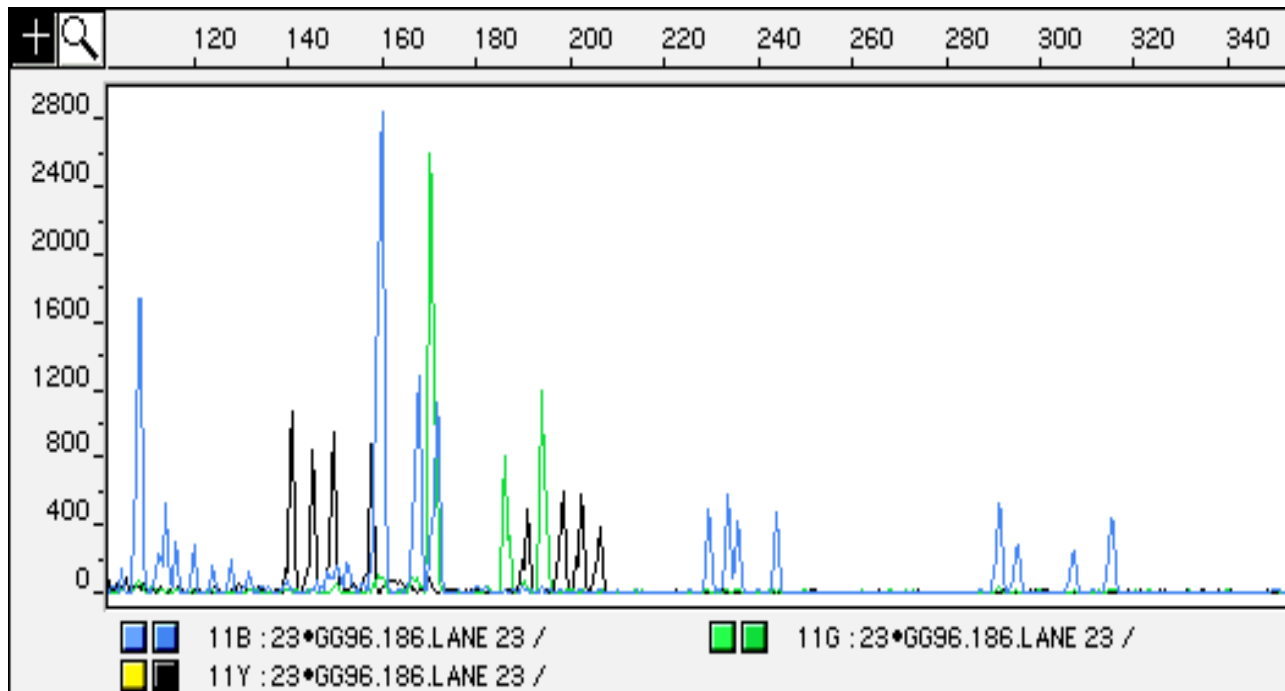
- Check within range of allelic ladders (± 0.5 bp)
- Check band shift is consistent within heterozygotes.
- Gill et al (1996) Int. J. Leg Med 109:14-22



Step 3: Identify potential number of contributors

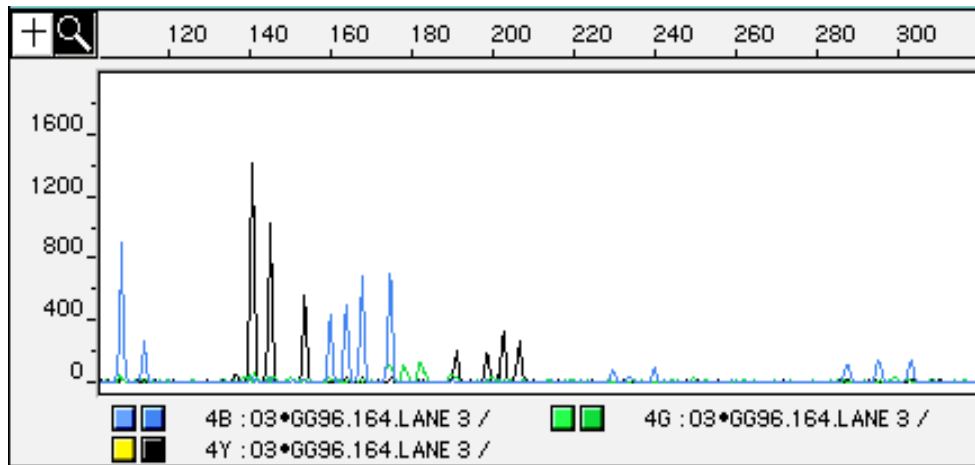


- No. of bands per locus
- Peak imbalance

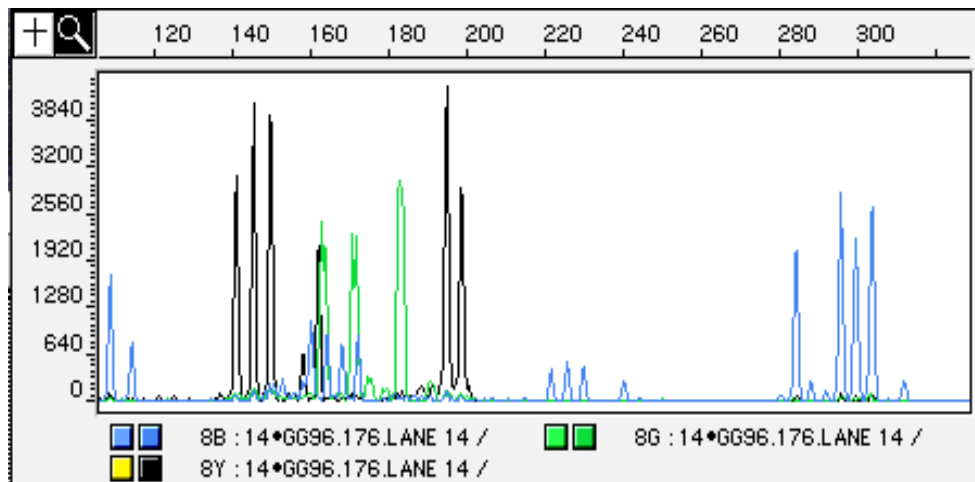




Potential number of contributors



**No more than four alleles at a given Locus
Probable 2 person mixture**



**6 Alleles present at D18
and 5 at VWA
More than two individuals contributing**

Estimation of the mixture ratio /proportion of the contributors



- Mixtures can range from the contributors being in equal proportion to one being in great excess
- It is useful in the interpretation to establish what type of mixture is present
 - **Type A** No defined major
 - **Type B** Clearly defined major and minor
 - **Type C** Low level minor
 - **Type D** All components are low level

Step 4 - Determine the approximate 'ratio' of the components in the mixture.



- if two DNA templates are mixed 2:1, then this approximate ratio will be maintained when the peak areas of the different component alleles within a locus are compared

Estimates of mixture proportion are similar across all loci



Sum of MIXTURE RATIO		ESTIMATED MIX						
NANOGRAMS DNA	LOCUS	1-1	1-10	1-2	1-5	10-1	2-1	5-1
1	D18	0.438900487	0.056382055	0.190444399	0.071656051	0.936987463	0.609566772	0.787082007
	D21	0.400271739	0	0.199728656	0.107755311	0	0.574754193	0.774910514
	D6	0.383542256	0	0.200319167	0.11651851	0	0.640705882	0.785885886
	FGA	0.450078247	0.132647288	0.252823683	0.146642468	0.868232944	0.561014263	0
	TH01	0.357883817	0	0.191841584	0.030945402	0.809622563	0.463647359	0.829716064
	VWA	0.413990826	0	0.24541797	0.166040181	0	0.582480362	0.837837838
5	D18	0.428681826	0.066388558	0.209330714	0.114621731	0.908953136	0.593318907	0.808556032
	D21	0.394698773	0.069672281	0.205883497	0.108104659	0.869415808	0.563085043	0.761076815
	D6	0.456590739	0.098497613	0.416630763	0.176317038	0.92350297	0.595953368	0.812808201
	FGA	0.484560652	0.148960803	0.281113901	0.144153897	0.844436006	0.641676122	0.786210785
	TH01	0.338320528	0.043332368	0.36197373	0.072725125	0.876234017	0.521763191	0.754897826
	VWA	0.441059122	0.129909366	0.301467863	0.176556874	0.893550228	0.611547085	0.808318137

Mixture proportions are similar across loci within a given mixture

Estimating the Mixture Proportion - M_x



- where minor peaks are easily distinguishable from the major component.
- $$M_x = \frac{(A+B)}{(A+B+C+D)}$$

Estimating the Mixture Proportion - M_x



- This is best used when we are dealing with more complex mixture and the scientist may wish to consider observed verses expected calculations

Estimation of the mixture ratio /proportion of the contributors



- Note these calculations only hold true when comparing peak area within a locus
 - (Intraloci ratio/proportion)
- But do not when observing alleles between different loci
 - (Interloci ratio/proportion) which can vary greatly between different amplifications.

Importance of the mixture proportion

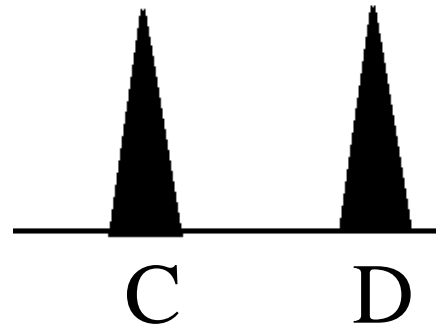


- Given the mixture proportion, and the observed peak areas, it is possible to calculate the expected peak areas for all possible combinations of genotypes that can be conditioned on.

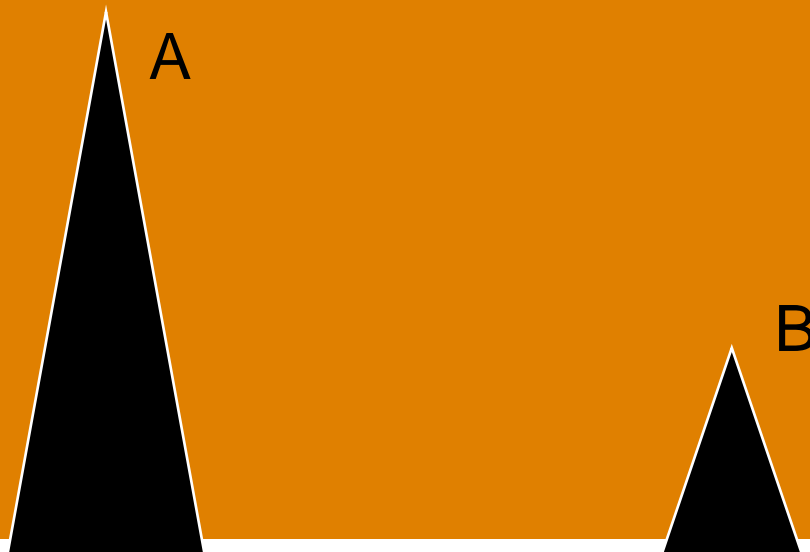


Heterozygote peaks

- Check to make sure minor peak is greater than 60% size (peak area) of the major peak



- Check $C/D > 0.6$ (C is smallest peak)
- If $C/D < 0.6$ then reanalyse (consider possible amplification enhancement or suppression)



• *Consider one locus - if the mixture is 1:1*

If there are 2 bands ratio 3:1

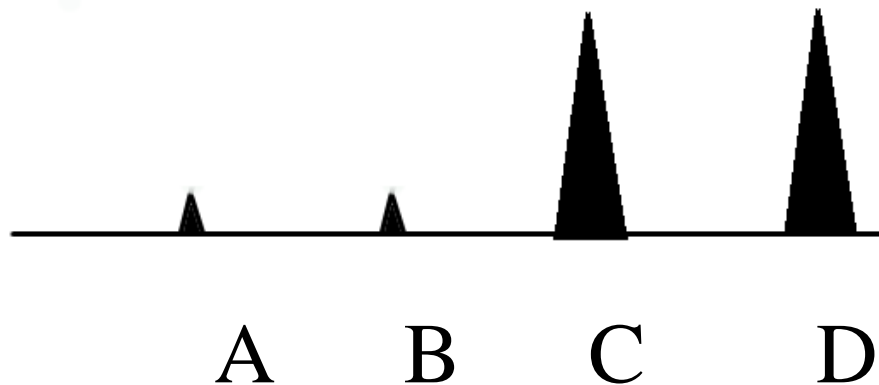
The evidence suggests the most supported genotype is AA,AB

AB,AB is less supported

AB,BB is even less supported

ie we can rank the most supported genotypes

Determine the mixture proportion *usually the mixture is not 1:1*



- Mixture proportion (M_x) = $(A+B)/(A+B+C+D)$
- Easiest when the mixture is 4-banded
- M_x is considered across loci (guidelines to be refined)

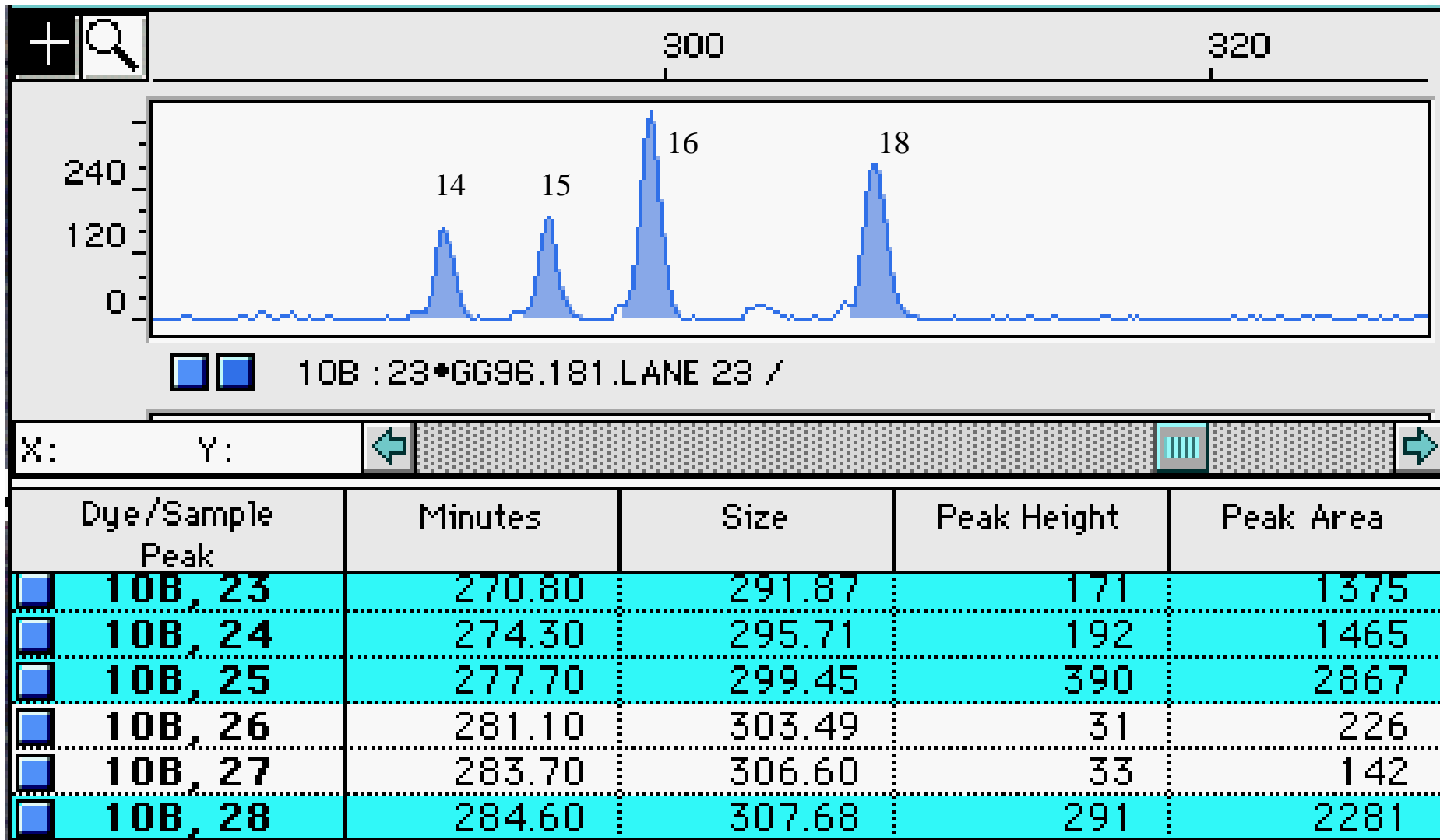


Fig 5: ABD GS Analysis of D18S51 mixture showing heterozygotes from two different individuals.

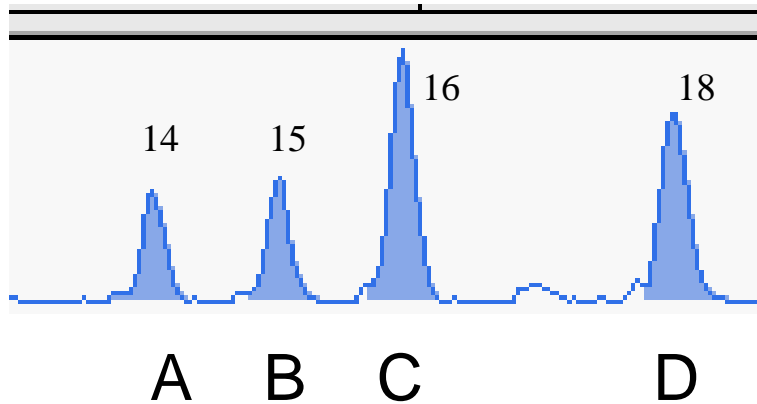
$$Mx = 0.35$$



Step 4 - interpretation

- Step 4 - Determine the possible pairwise combinations for the components of the mixture.
 - interpretation of the mixed profile is conducted independently of knowledge of the results of reference samples.
 - a) The interpretation cannot be influenced by the reference sample results, and is therefore demonstrably objective.
 - b) since the scientist is unhindered by prior knowledge of the results from the reference samples, the various alternative interpretations can be more easily considered.

With and without considering peak areas the possibilities are:



WITHOUT PEAK AREA CONSIDERATION:

AB,CD
AC,BD
AD,BC
CD,AB
BD,AC
BC,AD

WITH PEAK AREA CONSIDERATION:

AB,CD

CD,AB

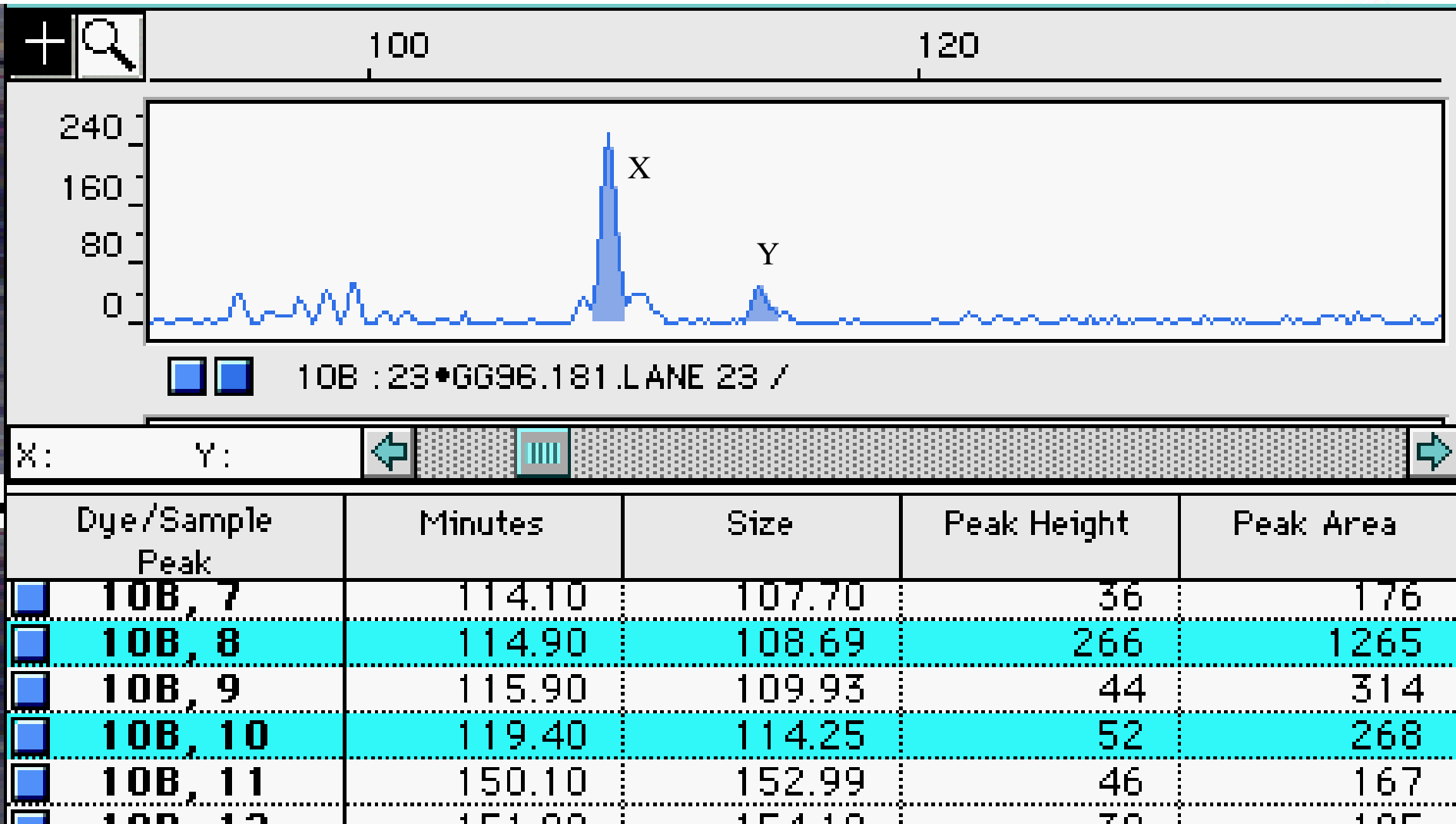
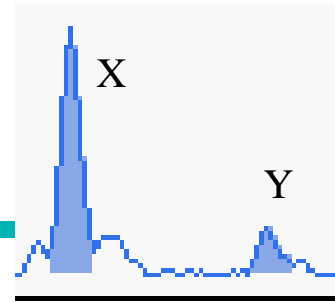


Fig 6: Amelogenin, showing imbalanced X:Y peaks, typical of a male/female mixture.

Amelogenin



- Can be used to independently assess the mixture proportion - assuming male/female mixture XY, XX
- We can use the information from amelogenin to decide the origin of major and minor contributors.
- In the example the mixture proportion is 0.36 male - ie minor contributor and agrees with 4-banded result

Estimate mixing proportion/ratio (M_x) Using Amelogenin



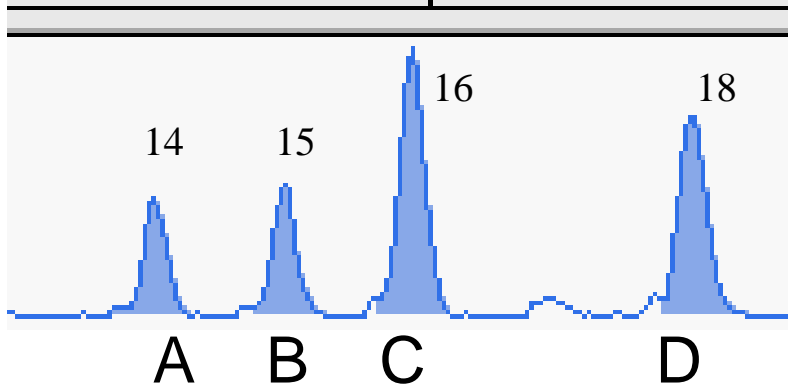
- If a male/female mixture is present, then the Amelogenin locus can provide another estimate of M_x
- However, this has shown itself to be the least robust estimator
- Of no use when both individuals are same sex or where the male predominates over the female



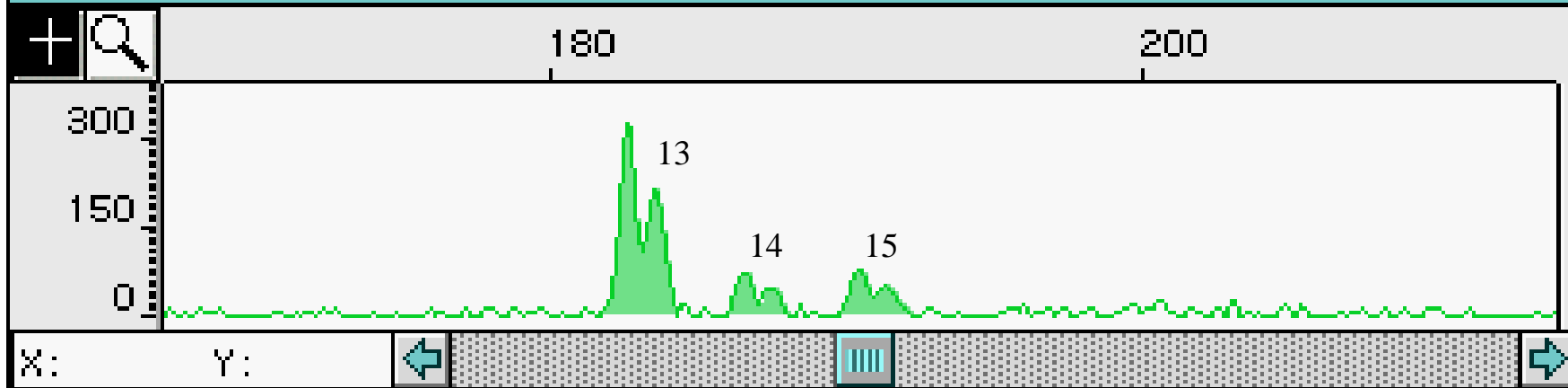
How dosage and ratios of components are related

Ratio of components		Dosage of products observed	
<u>Male</u> XY	<u>Female</u> XX	X	Y
10	1	12	10
5	1	7	5
4	1	6	4
3	1	5	3
2	1	4	2
1	1	3	1
1	2	5	1
1	3	7	1
1	4	9	1
1	5	11	1
1	10	21	1

Now consider the minor contributor is male. Condition on suspect and unknown female in LR denominator



Most supported genotype is Male 14,15
Female is 16,18



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

A three banded profile

Mixture proportion = 0.35 - this is used to compare observed and expected peak areas (female:male)

Possibilities are: 13,13 14,15 or 13,15 13,14
 or: 13,14 13,15
 not: 13,14 15,15



The possibilities

	Female	Male
Possibilities are:	13,13	14,15
or:	13,14	13,15
or:	13,15	13,14

Male is minor component from amelogenin

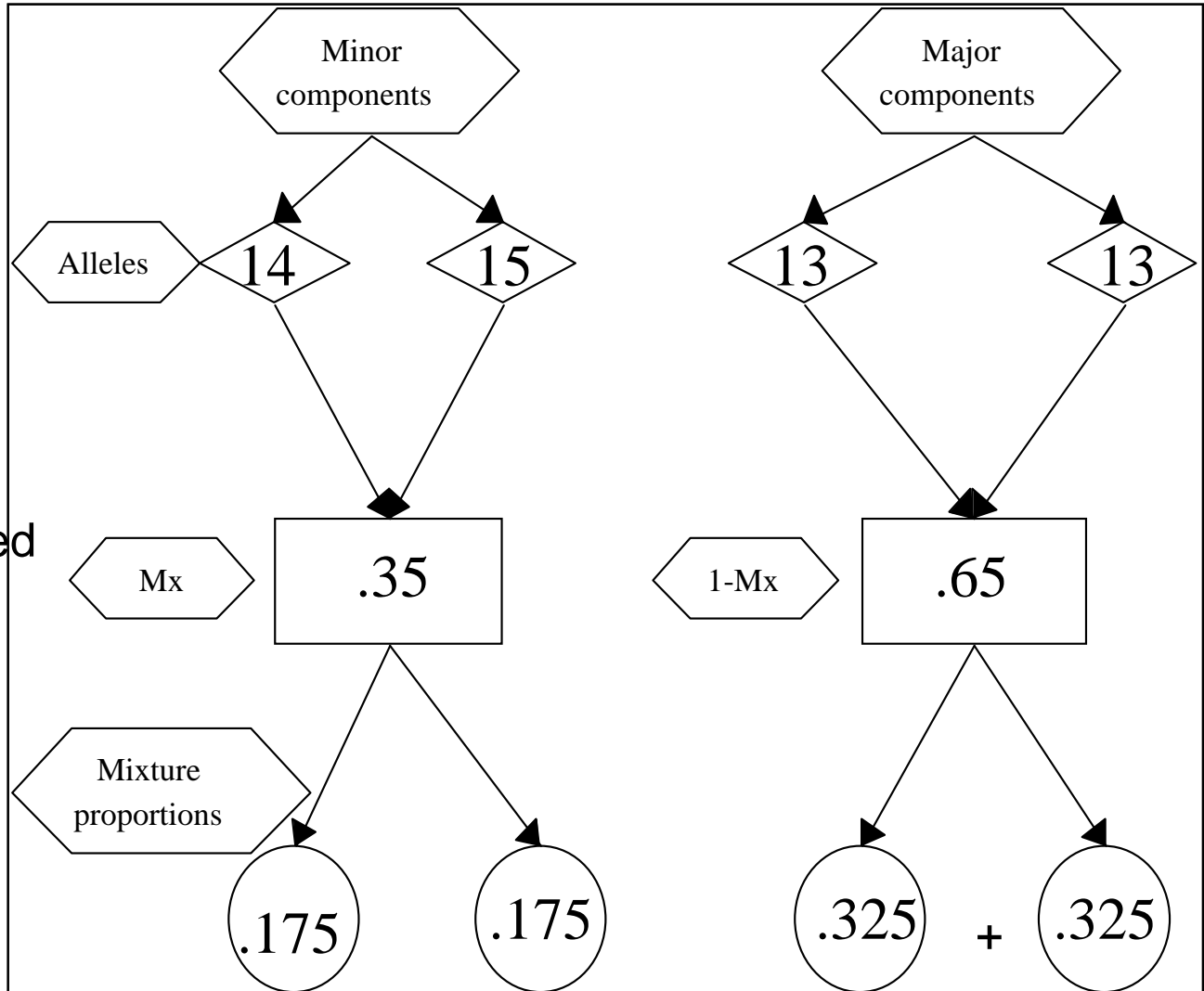


Which possibility is best explanation?

We can condition on the various genotypes
“If the minor components are 14,15 and the Major components are 13,13 then what are the Expected peak areas”

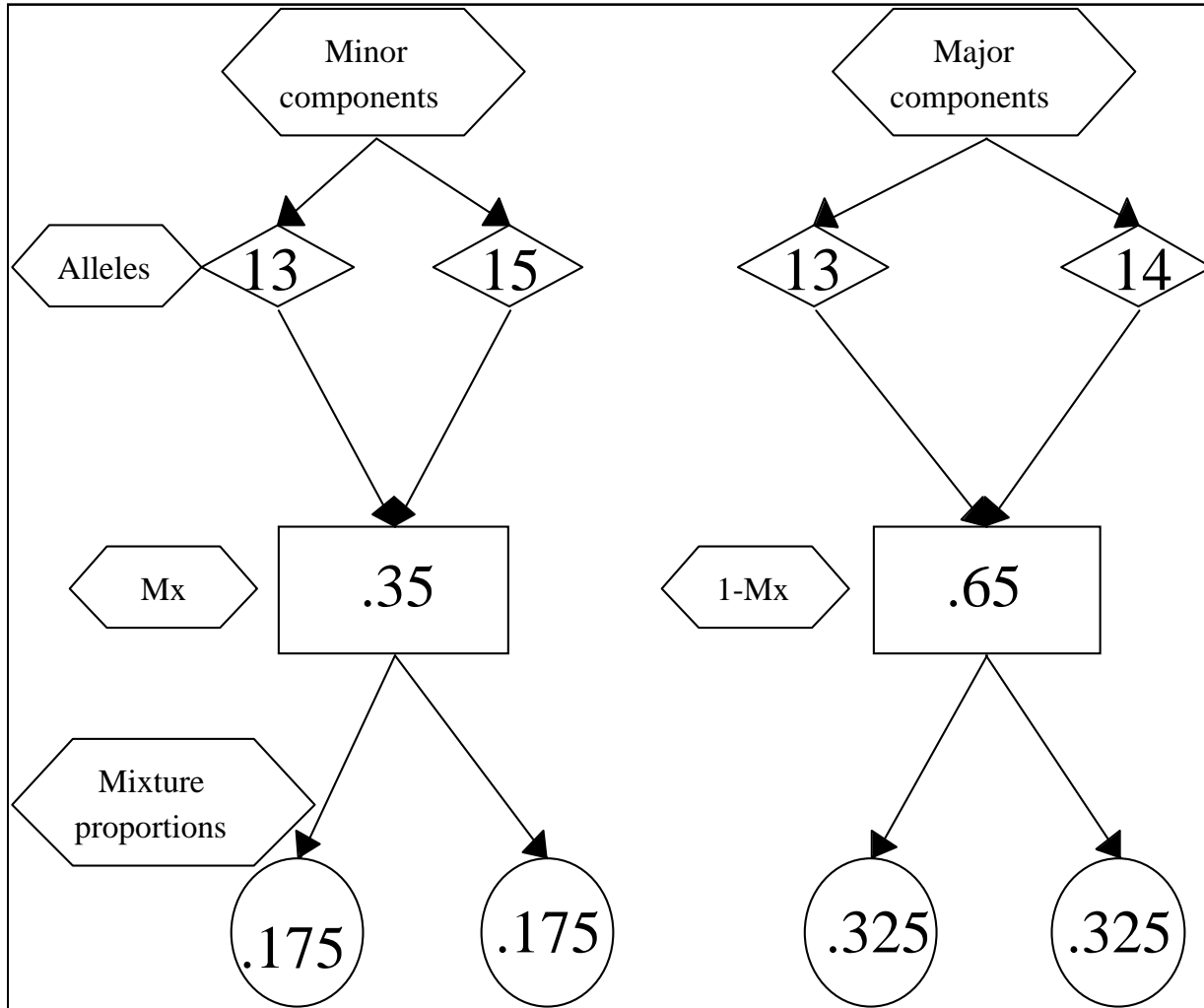
Allele	13	14	15
peak area	3299	738	927
	<u>Total</u>		4964
proportion	0.66	0.15	0.19

Calculation of expected peak areas from the mixture proportion



We worked
This out
earlier

We can compare any scenario we like



Allele 13
 $0.175 + 0.325 = 0.5$



To work out the expected peak area

- Multiply the total observed by the proportion expected
- Hence in the previous slide we expect that allele 13 would be half the total
- This is $4964 \times 0.5 = 2482$.
- Then we compare this with the actual observed peak area.



Now compare the various scenarios

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

This fits best because sum of differences is lowest

We can use a computer program to assess all different scenarios - best fit is least squares



Conditioning on the victim

- If suspect is 14,15 + victim is 13,13 then in the denominator of the likelihood ratio there is only one possibility if we can condition on the victim
- 13,13 14,15



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



-
- Make a list of possible genotypes
 - Refer to case circumstances
 - Possibility of conditioning - elimination of some genotypes - eg victim profile
 - Interpret with Evett model



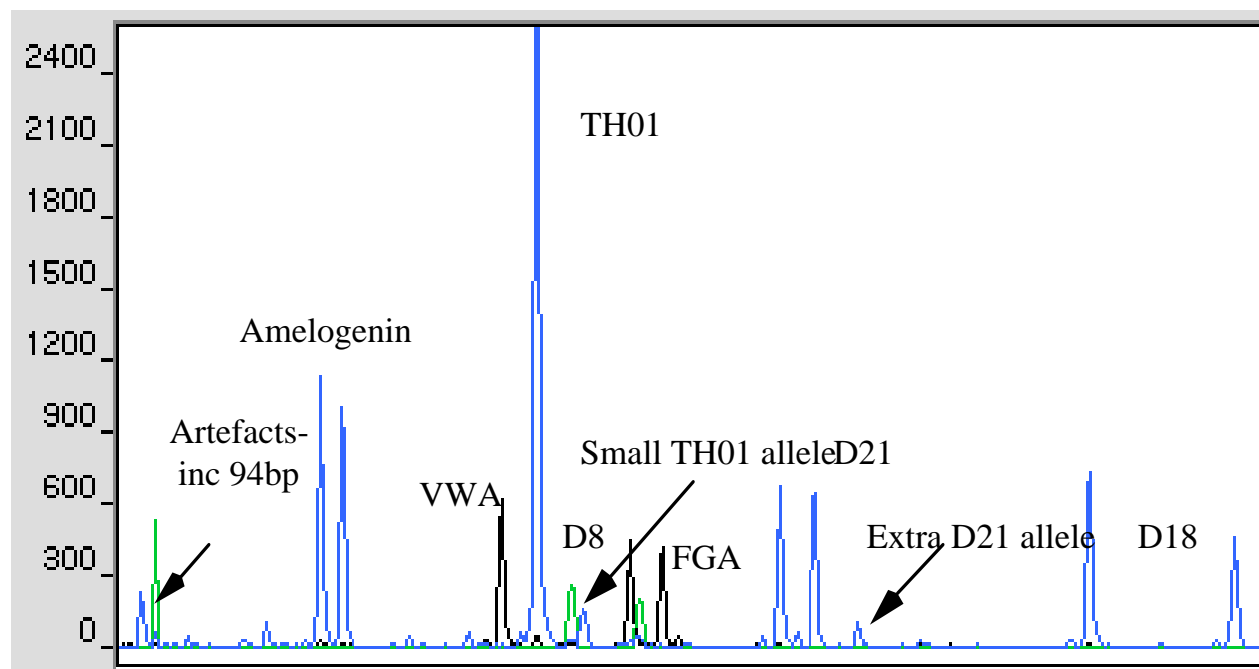
Step 5 Compare reference samples

- Watch out for unusual genetic phenomena in the reference sample
- **THIS SHOULD BE THE FIRST TIME THAT YOU HAVE LOOKED AT THE REFERENCE SAMPLE**



Assessing possible stutters

If a mixed profile has a minor peak in a stutter position it can never be certain if it is an allele or a stutter.

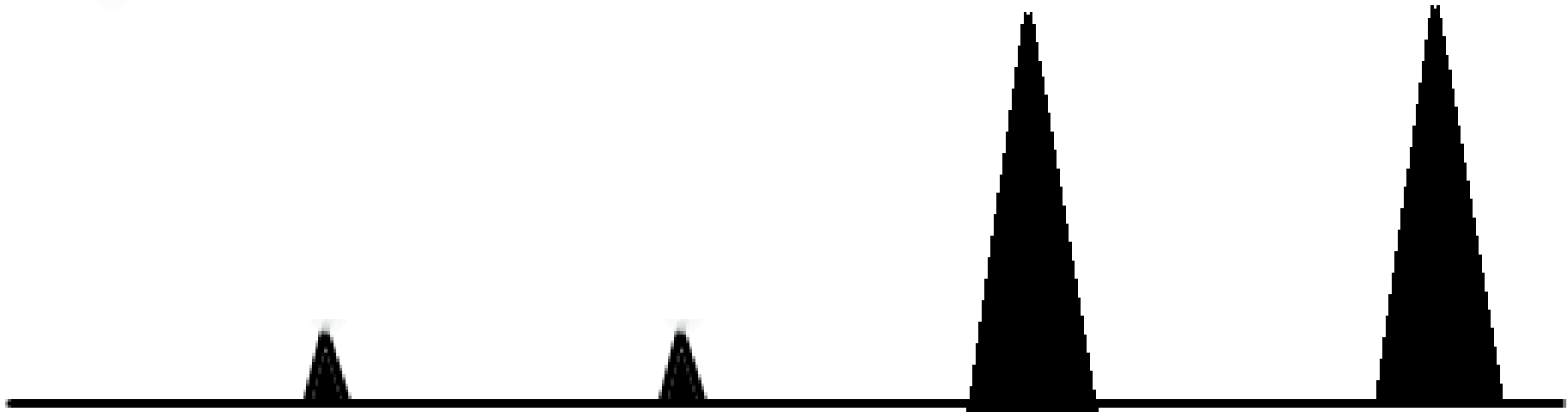


Interpretation of a mixed profile

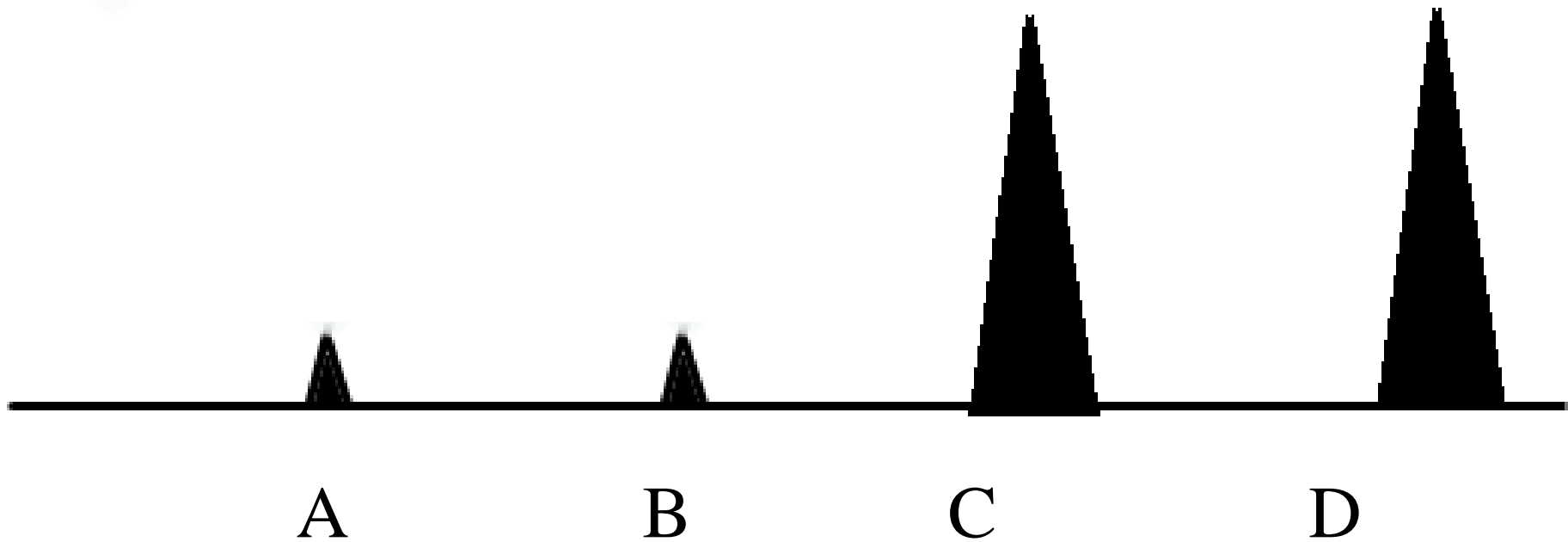
suspect is minor component



- Distance between B and C $> 4\text{bp}$
- B is unambiguous allele
- Suspect AB, Victim CD
- $LR=1/2AB$



Interpretation of a mixed profile

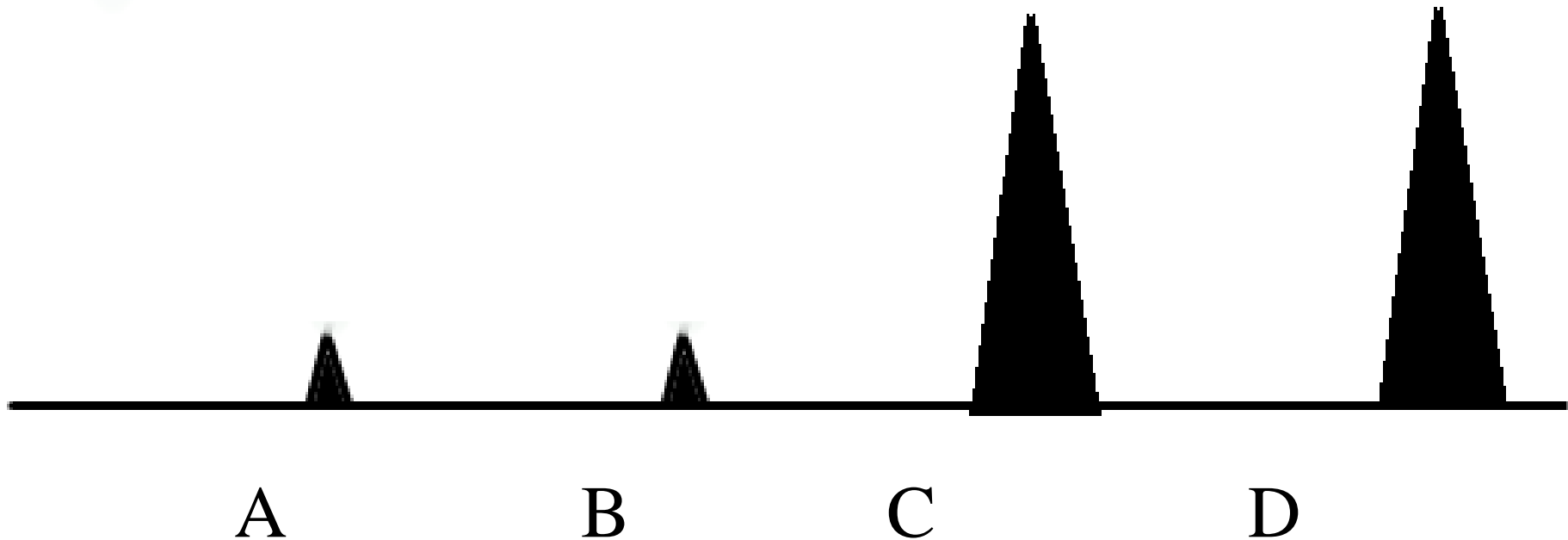


If $B=C-4\text{bp}$ then B could be a stutter

if suspect is AB and victim is CD (as before)

Denominator possibilities are AA, AB or AC or AD

$$LR \geq 1 / (A^2 + 2AB + 2AC + 2AD)$$



If $C-B > 4\text{bp}$ then the scenario changes

- There is no stutter associated with C
- The only possibility for the minor profile is AB
 $LR = 1/2AB$



Stutters

- Ignoring stutters is wrong
- Generally only significant if the minor profile is of evidential significance



Thresholds (different philosophy)

- All peaks are reported provided that they are distinguishable from background
- If a homozygote is observed that is below threshold of peak height 150 then it is treated as though drop-out may have occurred and the $P_m = 1/2p$
- Note that this is conservative provided that the homozygote peak area is low
- Reference samples must be complete profiles however



Reporting

How to report using likelihood ratio
philosophy



INFORMATION

A STATEMENT OF UNDERSTANDING

- From information received from my colleague Mr. Thomas of the Birmingham Laboratory, I understand that bloodstaining has been found on the sleeve of a jacket from Mr. Smith and that this blood may have come from Mr. Jones.

CONDITIONING THE STATEMENT



- My interpretation and conclusions are based on the information available at the time of this examination. Should this information change, I will need to reappraise the propositions considered. This reappraisal is more effective if carried out in advance of any trial.



PURPOSE

- To determine whether or not there is any support for the proposition that the bloodstain on the jacket came from Mr. Jones.
- In particular, to interpret the results of the DNA analysis undertaken in this case.



Statement of findings

- STR profiles have been obtained from the bloodstain from the jacket and from the two blood samples.
- The DNA profile from the bloodstaining on the jacket has the same DNA profile as that from the blood sample of Mr. Jones; it is different from that of the blood sample of Mr. Smith.



State the alternatives

- Propositions
- In order to assess the significance of the above findings I have considered two propositions:
 - the bloodstain came from Mr.Jones;
 - the bloodstain came from some unknown person unrelated to Mr.Jones.



Consider the alternatives

- If the bloodstain had in fact come from Mr. Jones, then I would expect to find that he had the same profile as the stain.
- On the other hand, the profiling technique that has been used in this case is so powerful that the chance of two unrelated people sharing the same profile is of the order of one in a billion.

We use a threshold of 1 billion



Publications

(without considering peak area)

- Evett et al (1991) A guide to interpreting single locus profiles of DNA mixtures.... J Forensic Sci Soc 31:41-47
- Weir et al (1997) Interpreting DNA mixtures J Forensic Sci Soc 42:213-222



Publications

considering peak area and artefacts

- Evett et al (1998) Taking account of peak areas.... J. Forensic Sci Soc 43:62-69
- Gill et al (1998) Interpreting simple STR mixtures....Forensic Sci Int 91:41-53
- Clayton et al (1998) Analysis and interpretation Forensic Sci Int 91:55-70



Publications

- Gill et al (1998) Interpretation of simple mixtures when stutters are present.. *Forens. Sci. Int.* 95:213-224
- Gill et al (1998) Interpretation of mixtures based on Peak area - identification of genetic anomalies, stutters and other artefacts. *Proceedings from the second European Symposium on human identification.* pp. 61-72



Publications

- Gill et al (2000) Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes. *Forens. Sci Int.* 108:1-29.
- Gill et al (2000) Interpretation of STRs when less than 100pg of DNA is present. *Forensic Sci Int* (Gill et al (2006))
- Gill et al (2006) DNA commission recommendations.... *FSI* 160, 90-101.