

Topics and Techniques for Forensic DNA Analysis
Continuing Education Seminar

Mixture Interpretation

NYC OCME
Dept of Forensic
Biology

New York City, NY
April 18, 2012



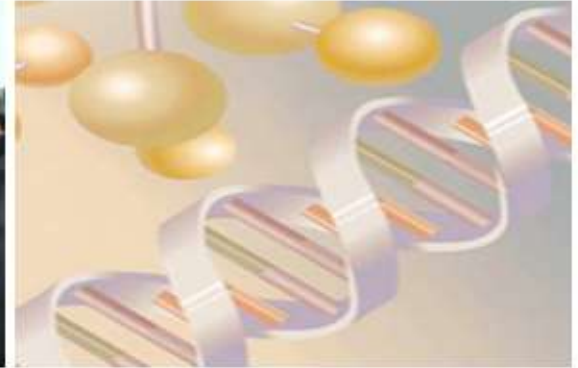
Dr. Michael D. Coble
National Institute of
Standards and Technology

michael.coble@nist.gov



SWGDM Website and Resources Available

<http://www.swgdam.org/resources.html>



Additional Resources

Beginning with the development or/and revision of its next draft guidance document(s), SWGDAM will make a "Draft for Comment" or other work product available for the purpose of receiving comments from the general public. This "Draft for Comment" solicitation will be open for a minimum of 60 days, usually through SWGDAM.org. SWGDAM will make all reasonable efforts to advise the forensic DNA community of the open comment period for a proposed guidance document or standard, guideline, best practice, study, or other recommendation and/or finding via as many avenues as possible to include posting notices through discipline-specific and related professional organizations. SWGDAM strongly encourages all interested parties to regularly monitor SWGDAM.org for the posting of such draft documents as well. All public comments received by SWGDAM will be forwarded to the appropriate SWGDAM Committee for review and consideration as a part of its formal business practice for the development of the guidance documents or other work product.

The following information resources have been produced and reviewed by members of the Mixture Committee of SWGDAM and are available at
www.cstl.nist.gov/biotech/strbase/mixture/SWGDAM-mixture-info.htm

- Home
- ByLaws
- Members
- Committees
- Meetings
- Publications

Link to <http://www.cstl.nist.gov/biotech/strbase/mixture/SWGDAM-mixture-info.htm>

Mixture Training Materials

Reviewed by SWGDAM Mixture Committee

SWGDAM Mixture Committee Resource Page

The following information resources have been produced and reviewed by [members](#) of the [Mixture Committee](#) of the Scientific Working Group on DNA Analysis Methods ([SWGDAM](#)) -- see <http://www.swgdam.org/resources.html> for additional information.

Mixture Training Examples

- Download "[Mixture 6" PowerPoint show](#) (56 Mb)
 - with voice-over by Bruce Heidebrecht (Maryland State Police); may work best if file is first saved to your computer
- Download "[Mixture IQAS2904" PowerPoint show](#) (35 Mb)
 - with voice-over by Bruce Heidebrecht (Maryland State Police); may work best if file is first saved to your computer

Recent Training Workshops



John Butler

Mike Coble



- AAFS (February 22, 2011)
 - **Mixture Interpretation (with 6 other speakers)**



- ISFG (August 30, 2011)
 - **CE Fundamentals and Troubleshooting**



- Int. Symp. Human Ident. (October 3, 2011)
 - **Mixture Interpretation (with Boston University)**



- Int. Symp. Human Ident. (October 6, 2011)
 - **Troubleshooting Laboratory Systems**

Slide handouts available at
<http://www.cstl.nist.gov/strbase/training.htm>

Mixture Workshop (Promega ISHI 2010)

<http://www.cstl.nist.gov/biotech/strbase/mixture.htm>

October 11, 2010

Handout >200 pages

Literature list of >100 articles

13 Modules Presented

- Introduction
- SWGID (John)
- Alleles (Catherine)
- Stochastic effects (Robin)
- Peak height ratios (Charlotte)
- Number of contributors (John)
- Mixture ratios (John)
- Mixture principles (Charlotte)
- Statistics (Mike)
- Case Example 1 (Robin)
- Case Example 2 (Charlotte)
- Case Example 3 (John)

Regional workshops presented in FL, TX, MI, and AZ (April – June 2011)
Updated mixture workshop presented at ISHI 2011 (October 3, 2011)

Catherine Grgicak
Boston U.

Mike Coble
NIST

Robin Cotton
Boston U.

John Butler
NIST

Charlotte Word
Consultant

NIJ Grant to Boston University funded ~150 state & local lab analysts to attend

Promega ISHI 2012 Mixture Workshop



- **John Butler**, Ph.D., NIST, Gaithersburg, MD
- **Michael Coble**, Ph.D., NIST, Gaithersburg, MD
- **Robin Cotton**, Ph.D., Boston University, Boston, MA
- **Catherine Grgicak**, Ph.D., Boston University, Boston, MA
- **Charlotte J. Word**, Ph.D., Gaithersburg, MD

This workshop is for analysts, technical reviewers and technical leaders performing and interpreting validation studies and/or interpreting and reviewing STR data, particularly more difficult mixtures. Various DNA profiles will be analyzed and interpreted using selected analytical thresholds and stochastic thresholds to demonstrate the impact of those values on the profiles amplified with low-template DNA vs. higher amounts of DNA. Different statistical approaches and conclusions suitable for the profiles will be presented.

Useful Articles on DNA Mixture Interpretation

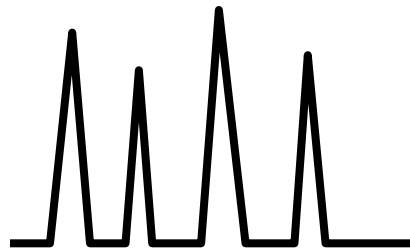
- Buckleton, J.S. and Curran, J.M. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2: 343-348.
- Budowle, B., *et al.* (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J. Forensic Sci.* 54: 810-821.
- Clayton, T.M., *et al.* (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 91: 55-70.
- Gill, P., *et al.* (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101.
- Gill, P., *et al.* (2008) National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *FSI Genetics* 2(1): 76–82.
- Schneider, P.M., *et al.* (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *Int. J. Legal Med.* 123: 1-5.

German Mixture Classification Scheme

Schneider *et al.* (2009) *Int. J. Legal Med.* 123: 1-5

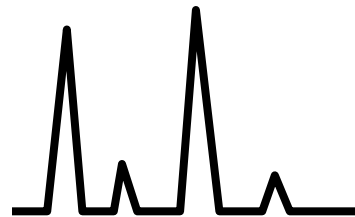
(German Stain Commission, 2006):

- **Type A:** no obvious major contributor, no evidence of stochastic effects
- **Type B:** clearly distinguishable major and minor contributors; consistent peak height ratios of **approximately 4:1** (major to minor component) for all heterozygous systems, no stochastic effects
- **Type C:** mixtures without major contributor(s), evidence for stochastic effects



Type A

“Indistinguishable”



Type B

“Distinguishable”



Type C

“Uninterpretable”

SWGDM

Available for download from the ISFG Website:
<http://www.isfg.org/Publication;Gill2006>



Available online at www.sciencedirect.com



Forensic Science International 160 (2006) 90–101



www.elsevier.com/locate/forsciint

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

P. Gill^{a,*}, C.H. Brenner^b, J.S. Buckleton^c, A. Carracedo^d, M. Krawczak^e, W.R. Mayr^f,
N. Morling^g, M. Prinz^h, P.M. Schneiderⁱ, B.S. Weir^j

^aForensic Science Service, Trident Court, 2960 Solihull Parkway, Birmingham, UK

^bForensic Science Group, School of Public Health, University of California, Berkeley, CA 510-339-1911, USA

^cESR, Private Bag 92021, Auckland, New Zealand

Our discussions have highlighted a significant need for continuing education and research into this area.

^jUniversity of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, USA

Received 4 April 2006; accepted 10 April 2006

Available online 5 June 2006

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101



ISFG Recommendations on Mixture Interpretation

<http://www.isfg.org/Publication;Gill2006>

1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
2. Scientists should be trained in and use LRs
3. Methods to calculate LRs of mixtures are cited
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
5. Prosecution determines H_p and defense determines H_d and multiple propositions may be evaluated
6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
7. Allele dropout to explain evidence can only be used with low signal data
8. No statistical interpretation should be performed on alleles below threshold
9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101



Forensic Science International
91 (1998) 55–70

**Forensic
Science
International**

Analysis and interpretation of mixed forensic stains using DNA STR profiling

T.M. Clayton^{a,*}, J.P. Whitaker^a, R. Sparkes^b, P. Gill^b

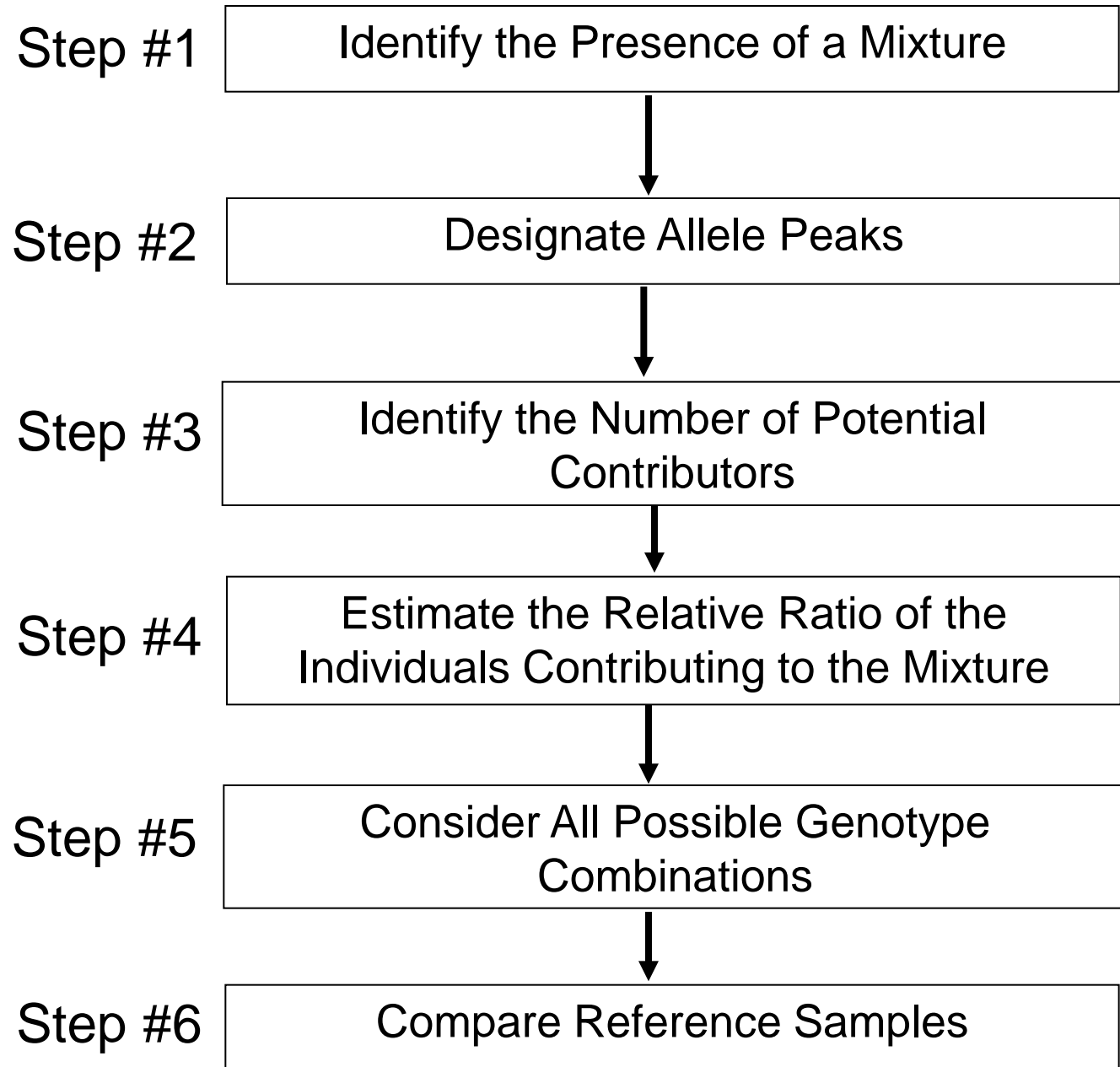
^a*Forensic Science Service, Wetherby Laboratory, Sandbeck Way, Audby Lane, Wetherby, West Yorkshire
LS22 4DN, UK*

^b*Forensic Science Service, Priory House, Gooch Street North, Birmingham B56QQ, UK*

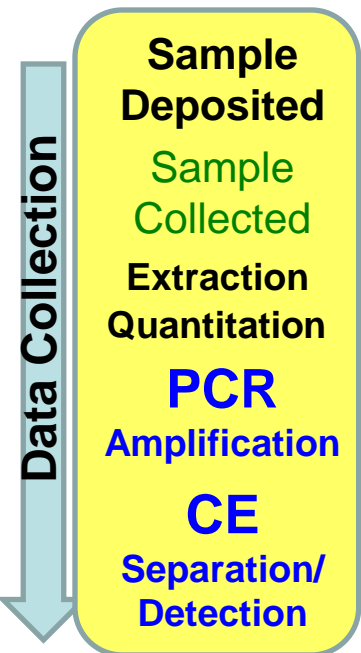
Received 13 May 1997; received in revised form 9 October 1997; accepted 27 October 1997

Steps in the interpretation of mixtures

(Clayton *et al.*
Forensic Sci. Int.
1998; 91:55-70)



Steps in DNA Interpretation



Signal observed

Data Interpretation

Analytical Threshold

Peak

Stutter Threshold

Allele

Stochastic Threshold

All Alleles Detected?

Peak Height Ratio

Genotype(s)

Mixture Ratio

Contributor profile(s)

Comparison to Known(s)
Weight of Evidence (Stats)

Overview of Two Thresholds

Example values
(empirically determined
based on own internal
validation)

200 RFUs

Called Peak

*(Greater confidence a sister
allele has not dropped out)*

MIT

Stochastic Threshold

The value above which it is
reasonable to assume that
allelic dropout of a sister
allele has not occurred

Called Peak

*(Cannot be confident
dropout of a sister allele
did not occur)*

50 RFUs

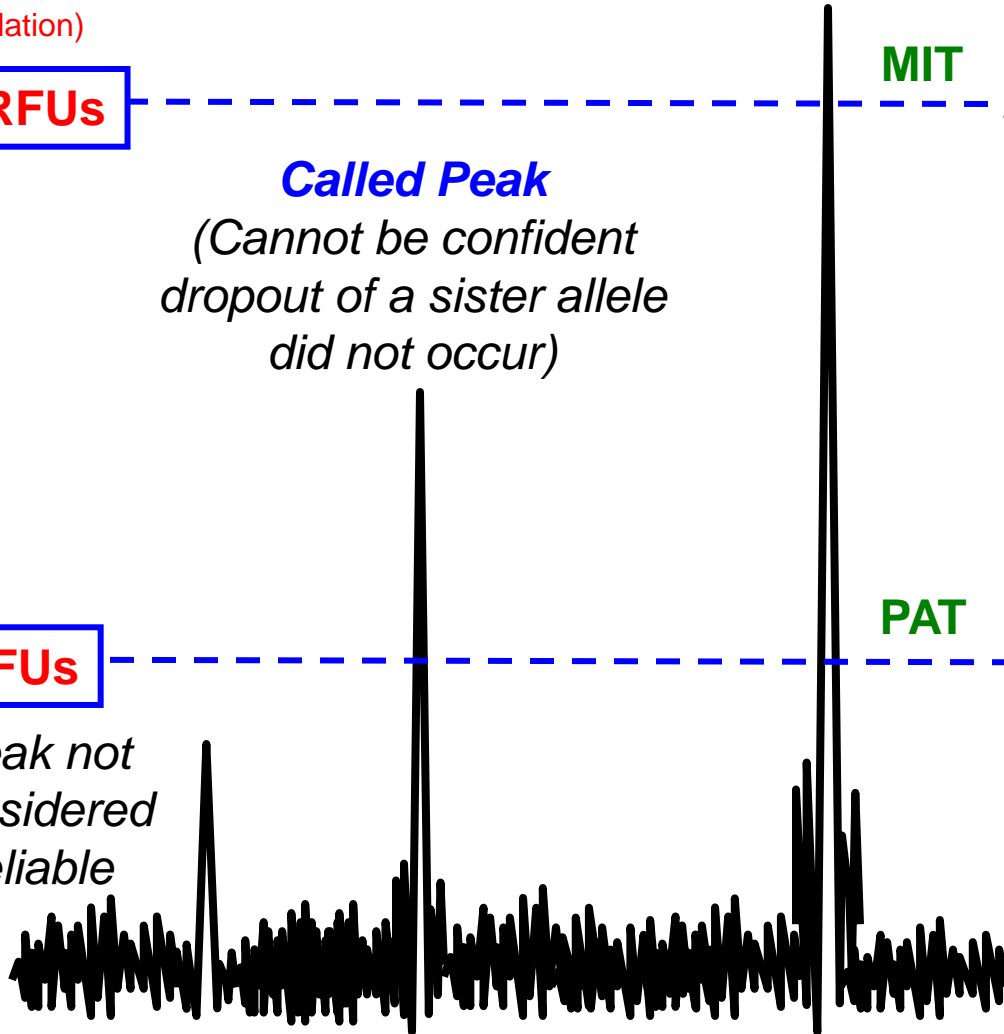
PAT

Analytical Threshold

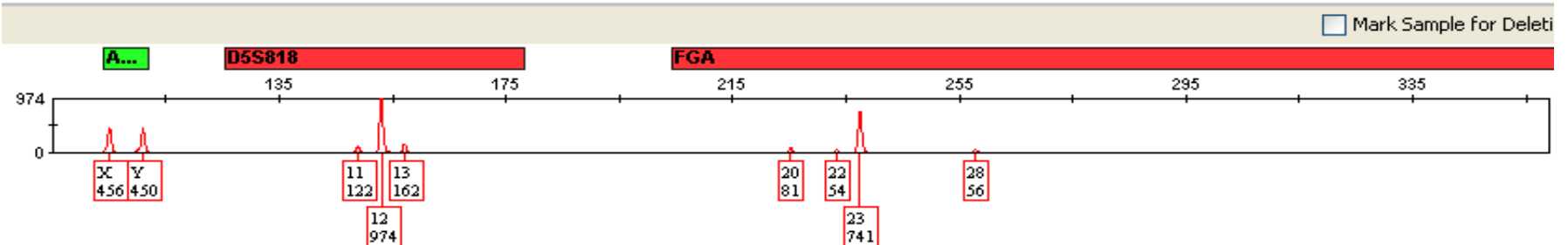
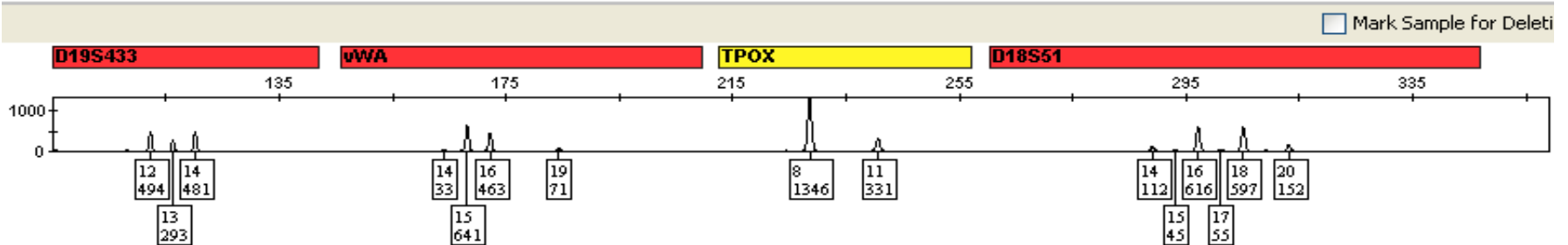
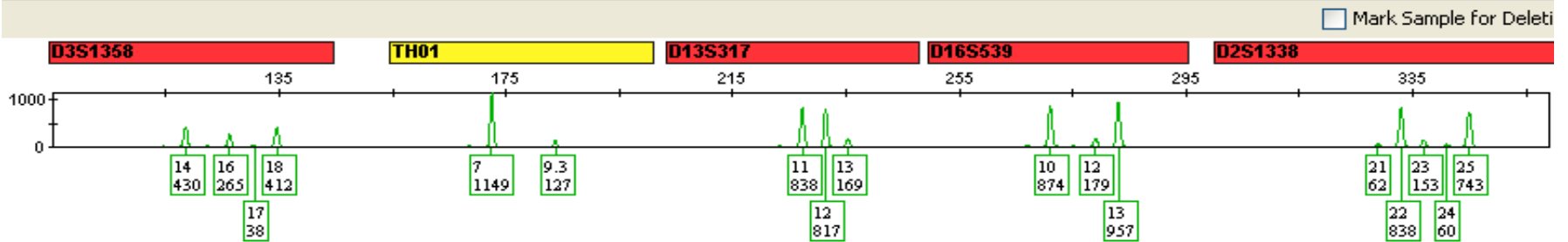
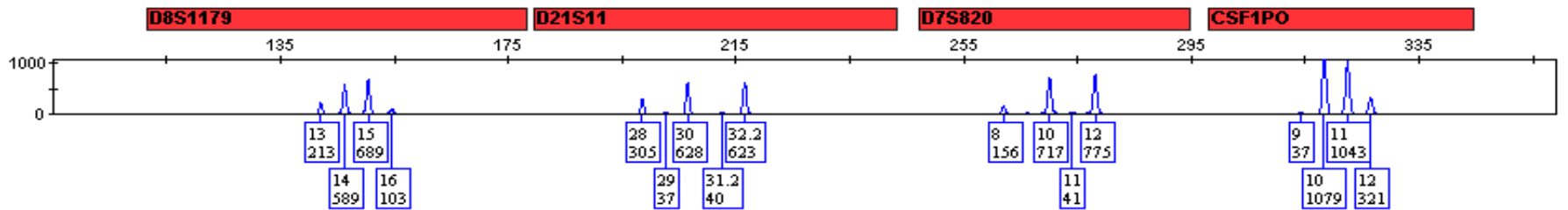
Minimum threshold for data
comparison and peak
detection in the DNA typing
process

*Peak not
considered
reliable*

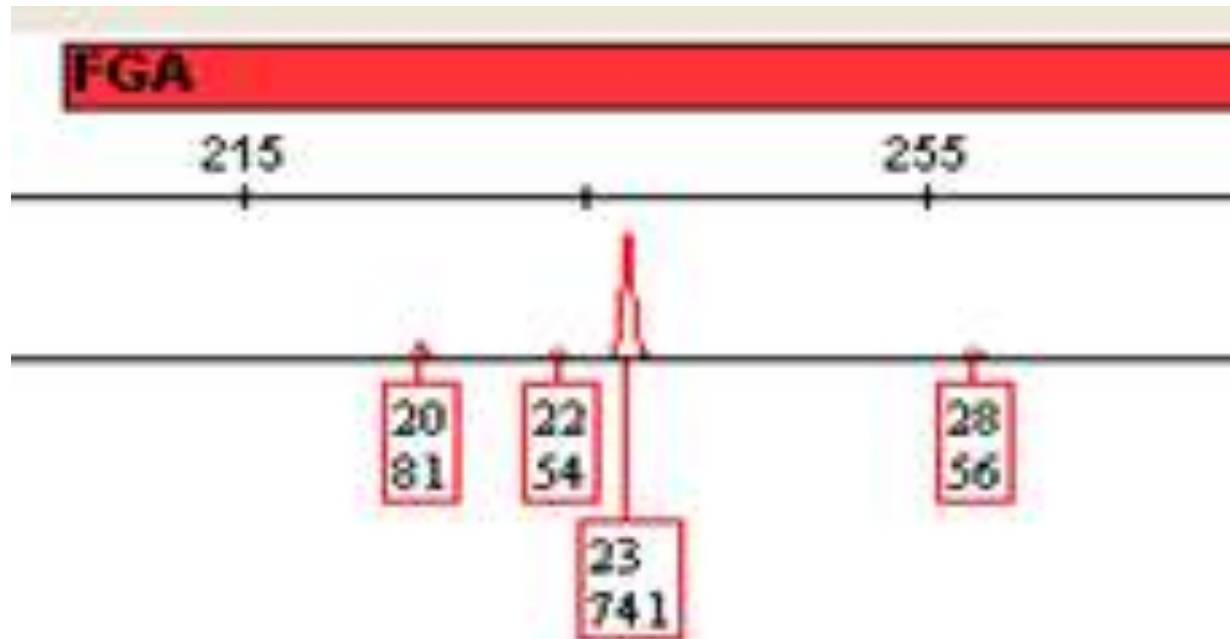
Noise



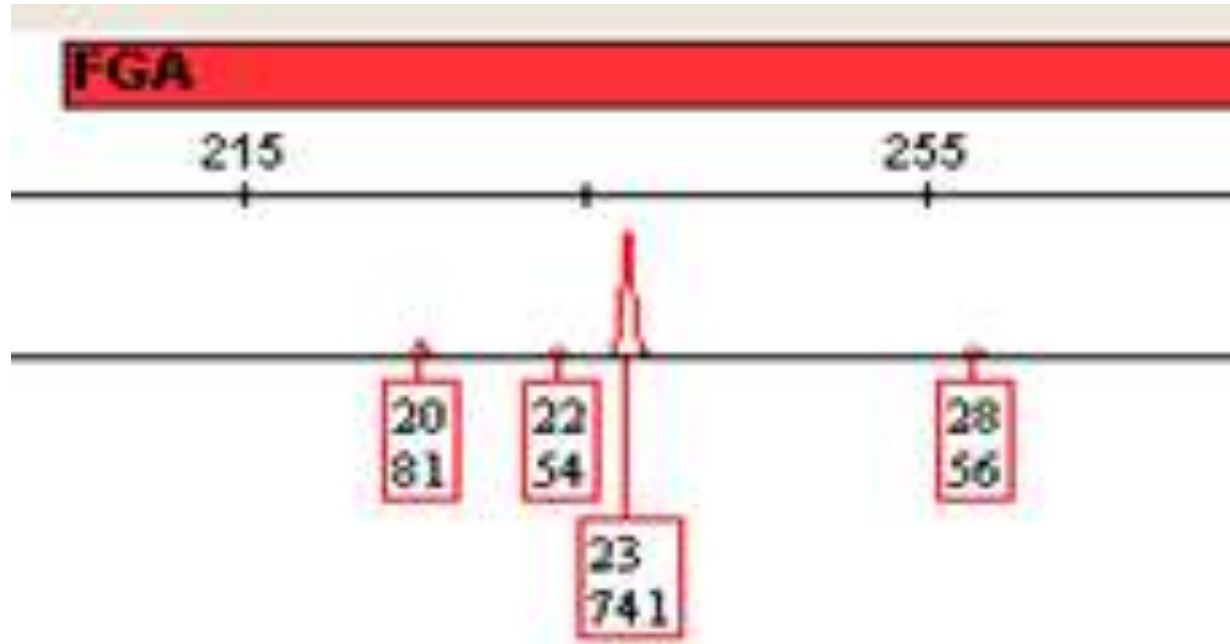
Profile 1 (stutter filter off)



Analytical Threshold (Peaks vs. Noise)

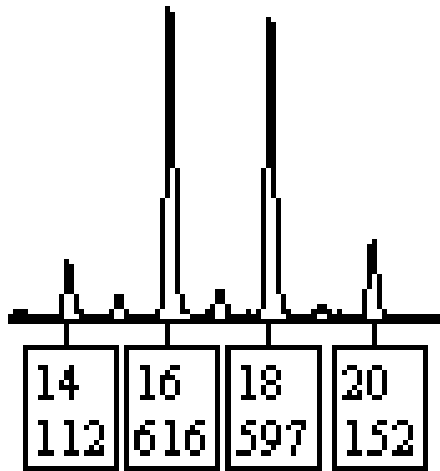


Stutter Threshold (Alleles vs. Artifacts)



Assumptions based upon # of contributors

Determination of Genotypes (PHR)



D18S51

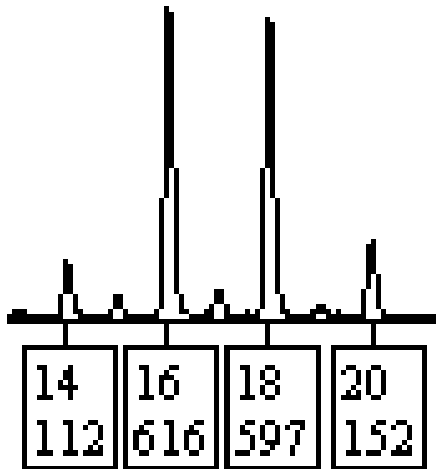
Possible Combinations

~~14, 16 and 18, 20
(18%) (25%)~~

~~14, 18 and 16, 20
(19%) (25%)~~

14, 20 and 16, 18
(74%) (97%)

Determination of Mixture Ratio



Major: 16,18

Minor: 14,20

D18S51

Total of all peak heights

$$= 112 + 616 + 597 + 152$$

$$= 1477 \text{ RFUs}$$

Minor component:

$$("14"+"20")/\text{total} = (112+152)/1477 = \mathbf{0.179}$$

Major component:

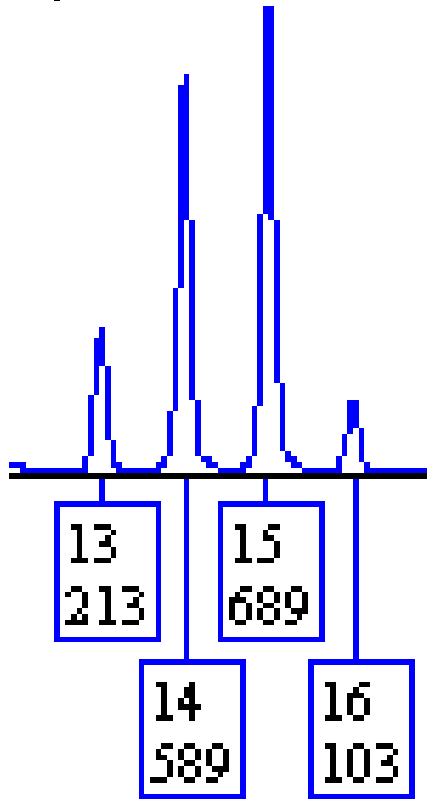
$$("16"+"18")/\text{total} = (616+597)/1477 = \mathbf{0.821}$$

≈ 4.6 : 1

Four Peaks (4 allele loci)

heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

Determination of Genotypes (PHR)



Possible Combinations

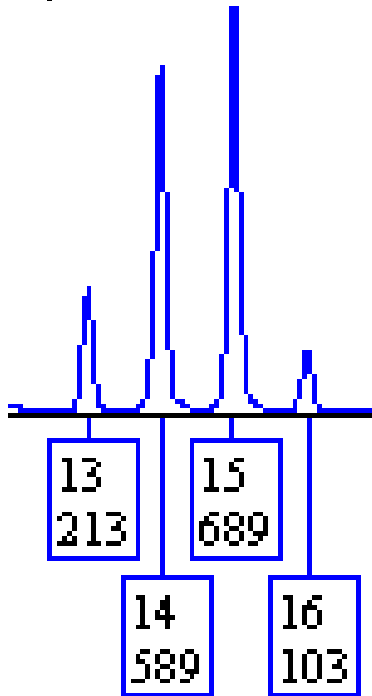
13, 14 and 15, 16
(36%) (15%)

13, 15 and 14, 16
(31%) (17%)

13, 16 and 14, 15
(48%) (85%)

Includes "stutter"
from the 14 allele

Determination of Mixture Ratio



Major: 14,15

Minor: 13,16

D8S1179

Total of all peak heights

$$= 213 + 589 + 689 + 103$$

$$= \mathbf{1594 \text{ RFUs}}$$

Minor component:

$$("13" + "16") / \text{total} = (213 + 103) / 1594 = \mathbf{0.198}$$

Major component:

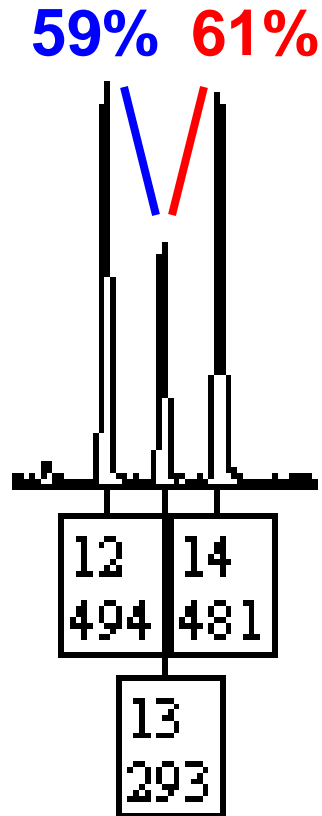
$$("14" + "15") / \text{total} = (589 + 689) / 1594 = \mathbf{0.802}$$

≈ 4 : 1

Four Peaks (4 allele loci)

heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

Application of the Mixture Ratio



D19S433

Using peak height ratio,
all genotypes possible:

12,12

12,13

13,13

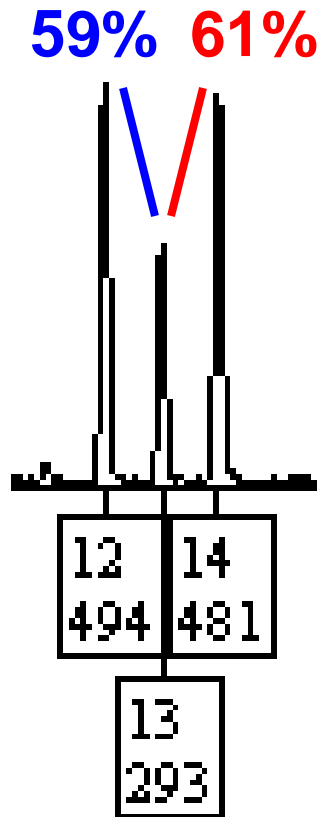
12,14

14,14

13,14

Is there a major:minor here?

Application of the Mixture Ratio



All possible genotype combinations:

12,12 + 13,14 1:1.6

13,13 + 12,14 1:3.3

14,14 + 12,13 1:1.6

12,13 + 12,14 1:1.4

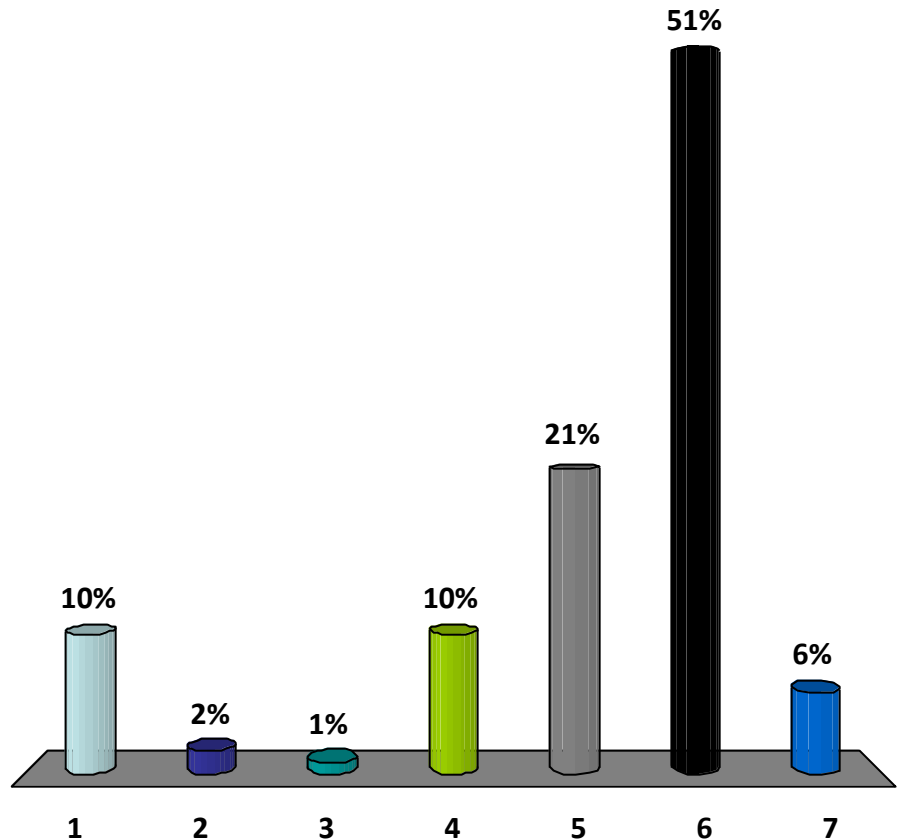
~~12,13 + 13,14 1:1~~

12,14 + 13,14 1:1.4

Using MIXTURE RATIO calculations, can eliminate genotype pairs

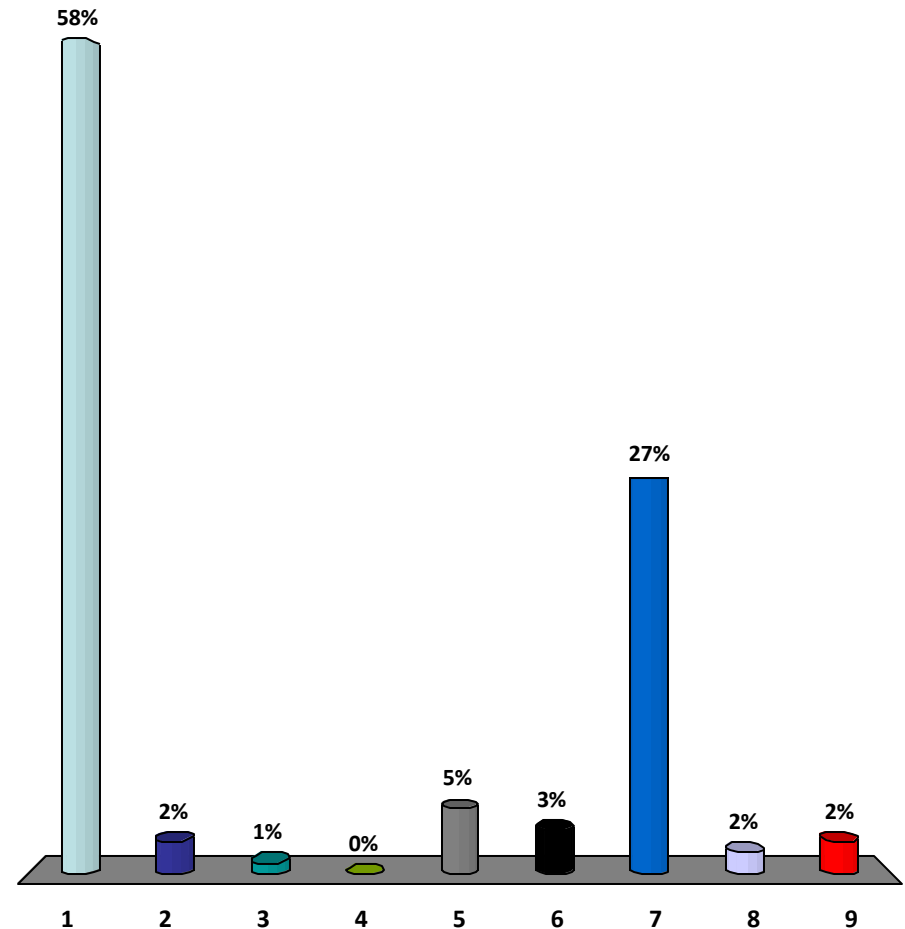
Does your lab use any software to help calculate mixture parameters (PHR, Mx, etc...)?

1. GMID-X
2. FSS I³
3. GeneMarker HID
4. True Allele
5. In-house Excel program
6. On a calculator (painfully)
7. Other



Does your lab use any software to help calculate mixture stats?

1. PopStats
2. GMID-X
3. FSS I³
4. GeneMarker HID
5. True Allele
6. DNA-View
7. In-house Excel program
8. On a calculator (painfully)
9. Other



Statistical Approaches with Mixtures

See Ladd *et al.* (2001) *Croat Med J.* 42:244-246

“Exclusionary” Approach

Random Man Not Excluded
(RMNE)

Combined Prob. of Inclusion
(CPI)

Combined Prob. of Exclusion
(CPE)

“Inferred Genotype” Approach

Random Match Probability
(RMP)

Likelihood Ratio
(LR)

A discussion of the merits of random man not excluded and likelihood ratios

John Buckleton^{a,*}, James Curran^b

^a*ESR, PB 92021, Auckland, New Zealand*

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

Received 15 January 2008; received in revised form 29 April 2008; accepted 1 May 2008

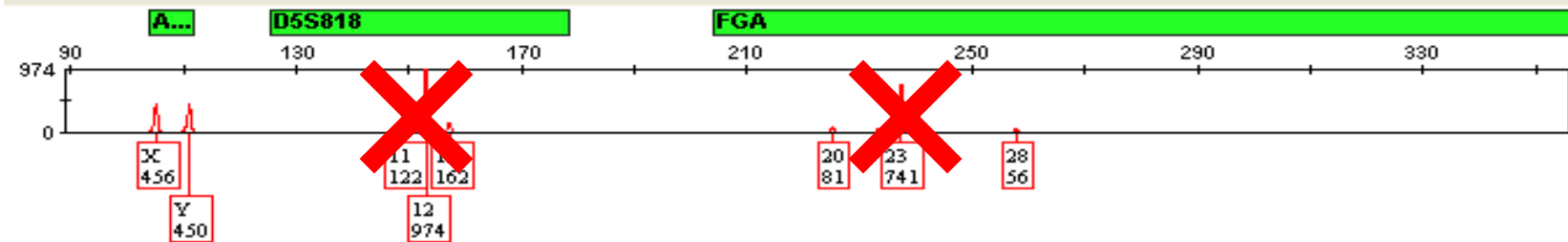
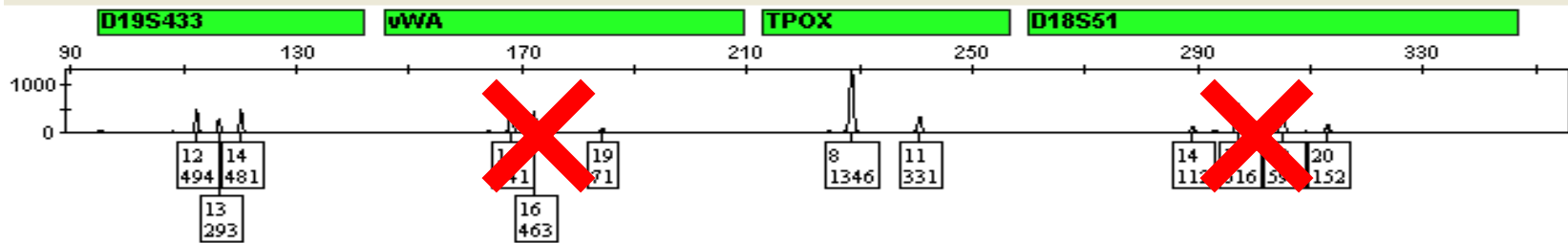
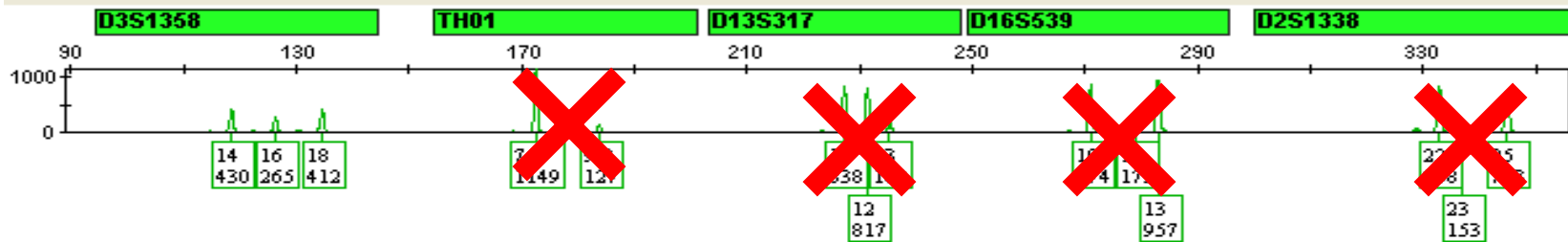
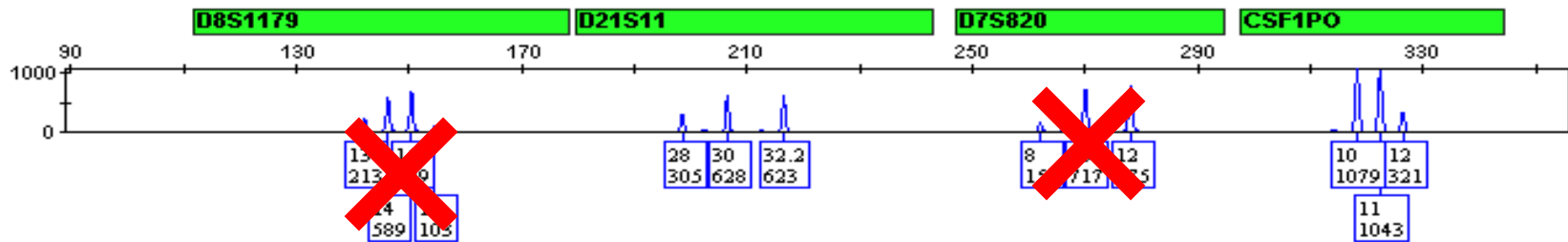
We conclude that the two matters that appear to have real force are:

- (1) LR_s are more difficult to present in court and
- (2) the RMNE statistic wastes information that should be utilised.

If CPI/CPE Stats are Used

Since exclusionary statistics cannot adjust for the possibility of dropout, and does not take the number of contributors into account, any loci where alleles are below stochastic levels cannot be used in the CPI statistic.

If CPI/CPE Stats are Used



If CPI/CPE Stats are Used

Can use

D21

CSF

D3

D19

TPOX

Cannot use

D8

D7

TH01

D13

D16

D2

vWA

D18

D5

FGA

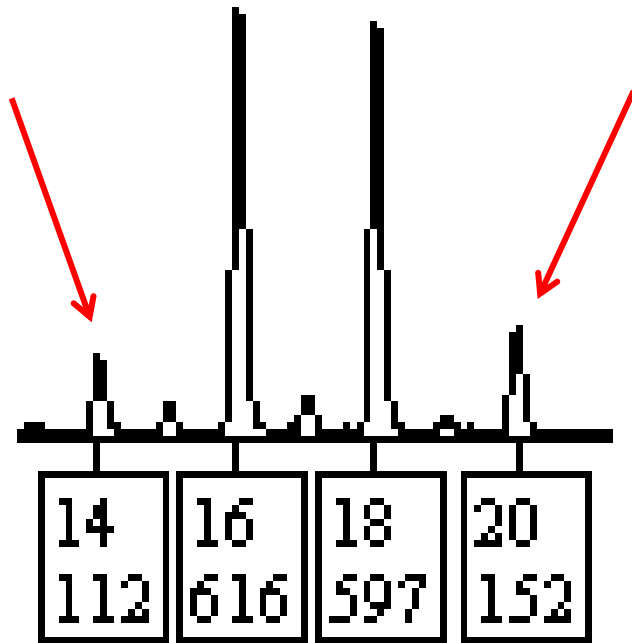
If CPI/CPE Stats are Used

- CPI statistics using FBI Caucasian Frequencies
- 1 in 71 Caucasians included
- 98.59% Caucasians excluded

If RMP/LR Stats are Used

- Since there is an assumption to the number of contributors, it is possible to use data that falls below the ST.

RMP - D18S51



If Assume 2 Contributors....

Major

16,18

Minor

14,20

$$\text{RMP}_{\text{minor}} = 2pq$$

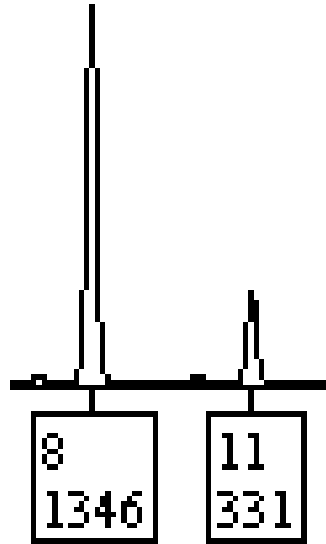
$$= 2 \times f(14) \times f(20)$$

$$= 2 \times (0.1735) \times (0.0255)$$

$$= 0.00884 \quad \text{or 1 in 113}$$

(LR = 113)

RMP - TPOX



If Assume 2 Contributors....

Major

8,8

Minor

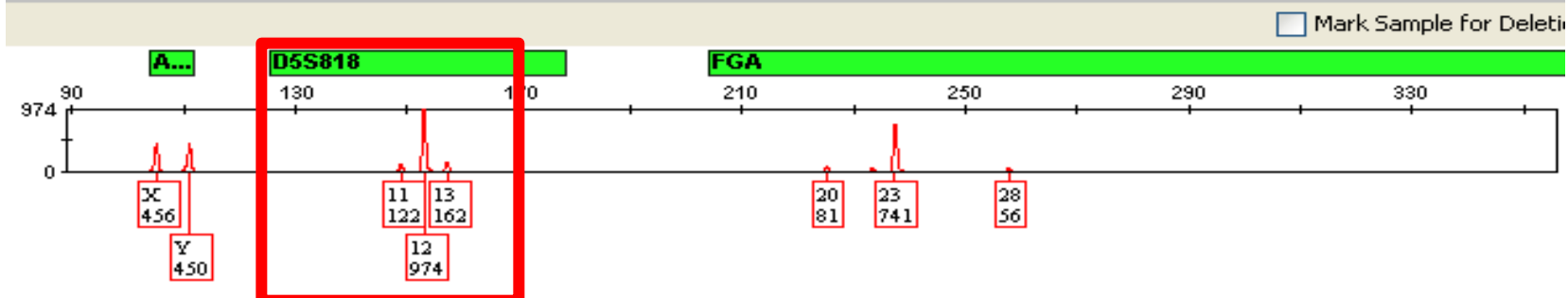
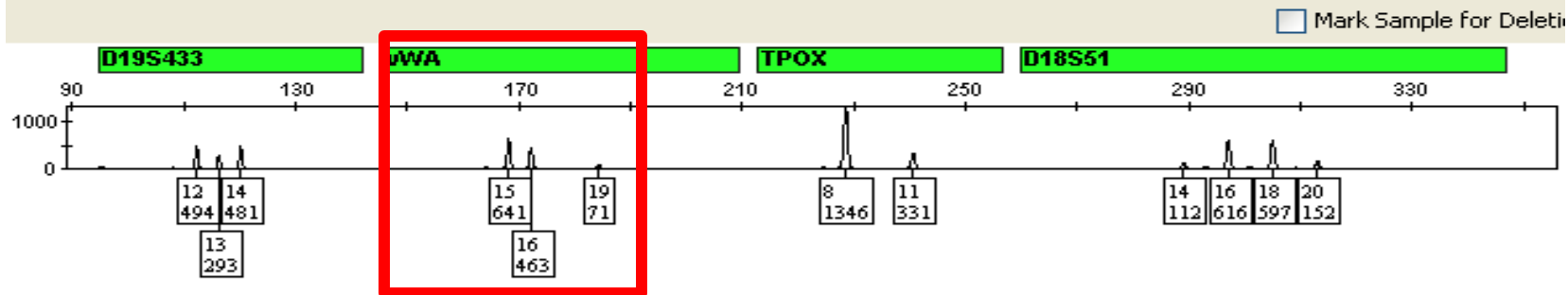
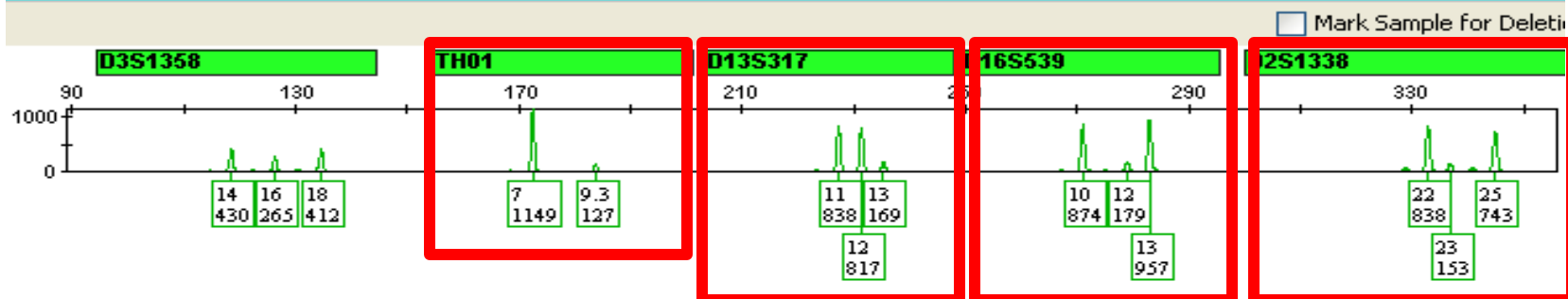
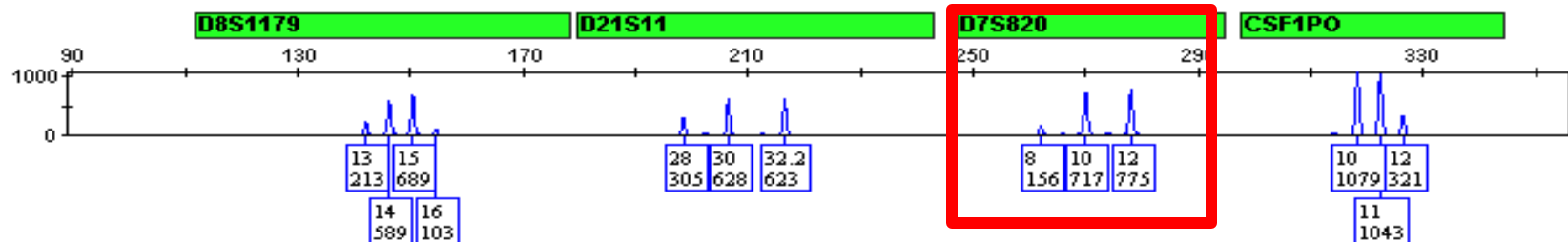
11,8 *OR* 11,11

$$\text{RMP} = 8,11 + 11,11$$

$$\text{RMP} = 2pq + (q^2 + q(1-q)\theta)$$

$$\begin{aligned} \text{RMP} &= 2(0.5443)(0.2537) + \\ & (0.2537)^2 + (0.2537)(0.7463)(0.01) \\ &= 0.3424 \quad \text{or 1 in 2.9} \end{aligned}$$

Profile 1: ID_2_SCD_NG0.5_R4,1_A1_V1.2



If RMP/LR Stats are Used

Can use

D8

D21

D18

D3

D19

TPOX

FGA

CSF

Loci with potential D-out

D7

D2

TH01

vWA

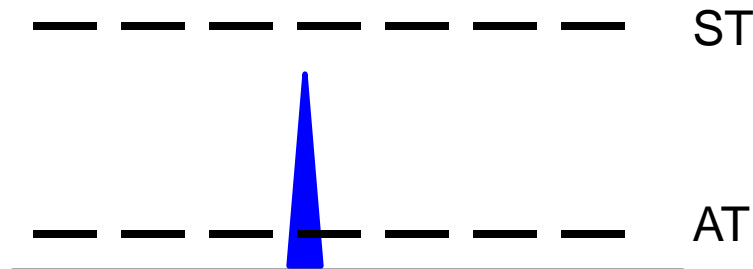
D13

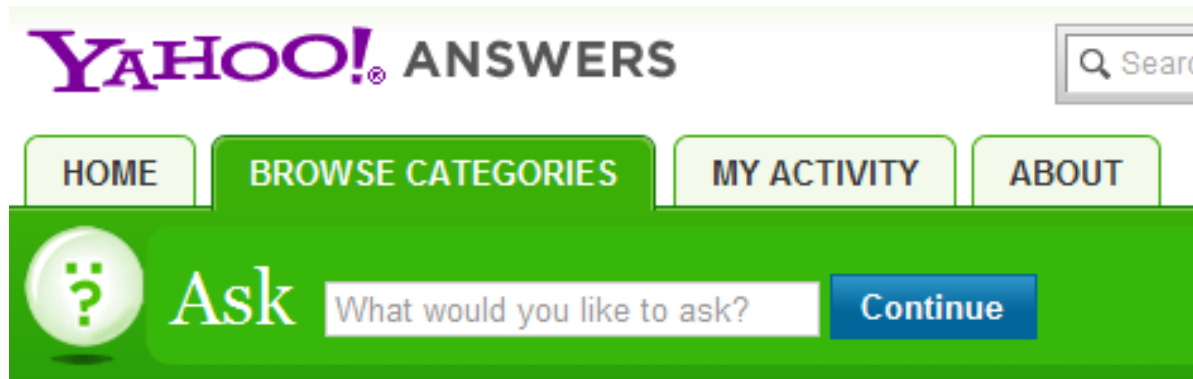
D5

D16

The “2p” Rule

- The “2p” rule can be used to statistically account for zygosity ambiguity – i.e. is this single peak below the stochastic threshold the result of a homozygous genotype or the result of a heterozygous genotype with allele drop-out of the sister allele?





Resolved Question

[Show me another »](#)

To pee or not to pee? That is the question...?

“Drink sir, is a great provoker of three things....
nose painting, sleep and urine.”

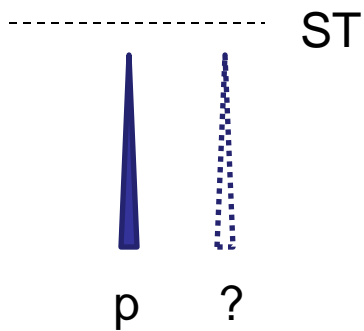
Macbeth: Act 2, Scene 3

2p – SWGDAM Guidelines

- 5.2.1.3.1. The formula $2p$, as described in recommendation 4.1 of NRCII, may be applied to this result.
- 5.2.1.3.2. Instead of using $2p$, the algebraically identical formulae $2p - p^2$ and $p^2 + 2p(1-p)$ may be used to address this situation without double-counting the proportion of homozygotes in the population.

$$2p - p^2 \text{ and } p^2 + 2p(1-p)$$

Suppose 5 allele system – P, Q, R, S & T



The possible genotype could be anything...

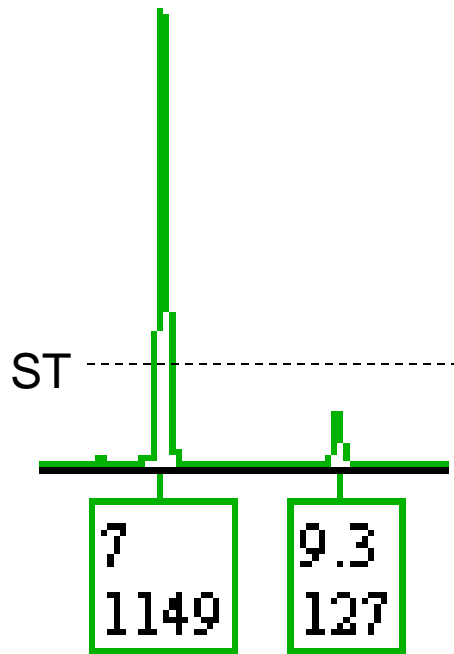
$$= PP + PQ + PR + PS + PT$$

$$= p^2 + 2pq + 2pr + 2ps + 2pt$$

$$= p^2 + 2p (q + r + s + t) \longrightarrow = (1-p)$$

$$= p^2 + 2p(1-p) = p^2 + 2p - 2p^2 = 2p - p^2$$

Profile 1 - TH01



Major – 7, 7

Possible Minor Contributors

7, 9.3 $(2pq)$

9.3, 9.3 p^2

9.3, ? $2p$ (or $p^2 + 2p(1 - p)$)

Profile 1 - TH01 (LR)

$$\frac{P(E | H_1)}{P(E | H_2)} = \frac{V \& S}{V \& U} = \frac{f_7^2 + f_7(1-f_7)\theta \& 1}{f_7^2 + f_7(1-f_7)\theta \&}$$

$$p^2 + 2p(1-p)$$

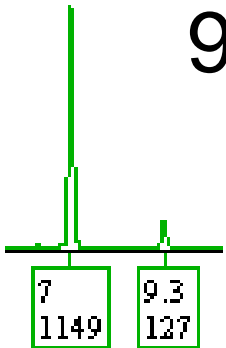
$$V = 7, 7$$

$$U = 7, 9.3$$

$$9.3, 9.3$$

$$9.3, ?$$

$$= \frac{1}{f_{9.3}^2 + 2f_{9.3}(1-f_{9.3})}$$



$$f_{9.3} = 0.3054$$

$$= 1 / 0.5175 = 1.93$$

Profile 1 - TH01 (LR)

$$\frac{P(E | H_1)}{P(E | H_2)} = \frac{V \ \& \ S}{V \ \& \ U} = \frac{1}{p^2 + p(1-p)\theta + 2pq}$$

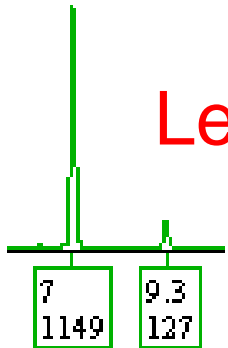
$$V = 7, 7$$

$$U = 7, 9.3$$

$$9.3, 9.3$$

$$= \frac{1}{f_{9.3}^2 + f_{9.3}(1-f_{9.3})\theta + 2f_{9.3}f_7}$$

Let $ST = 125$ RFU



$$f_{9.3} = 0.3054$$

$$f_7 = 0.1724$$

$$= 1 / 0.2007 = 4.98$$

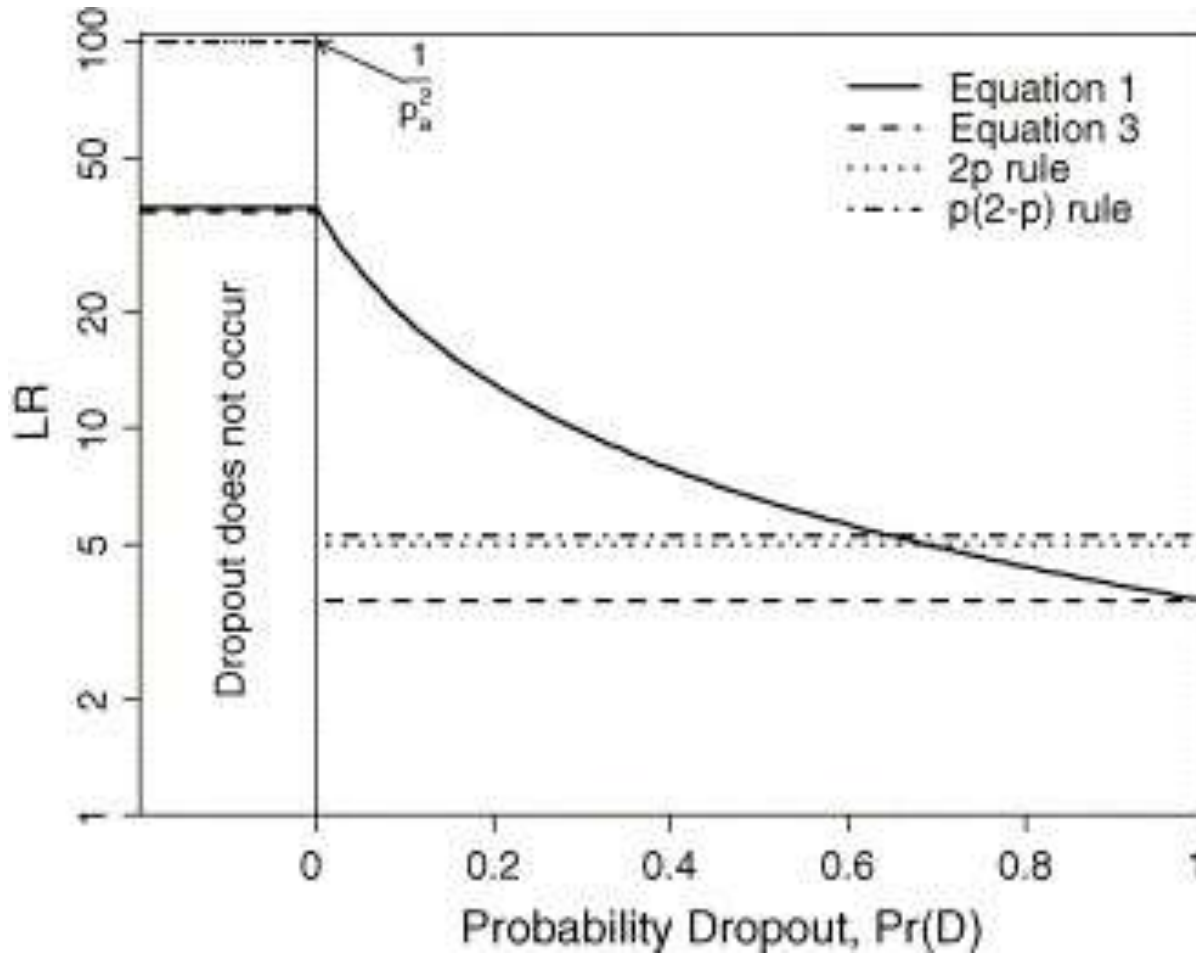
The “2p” Rule

- “This rule arose during the VNTR era. At that time many smaller alleles “ran off the end of the gel” and were not visualised.”

- Buckleton and Triggs (2006)

“Is the $2p$ rule always conservative?”

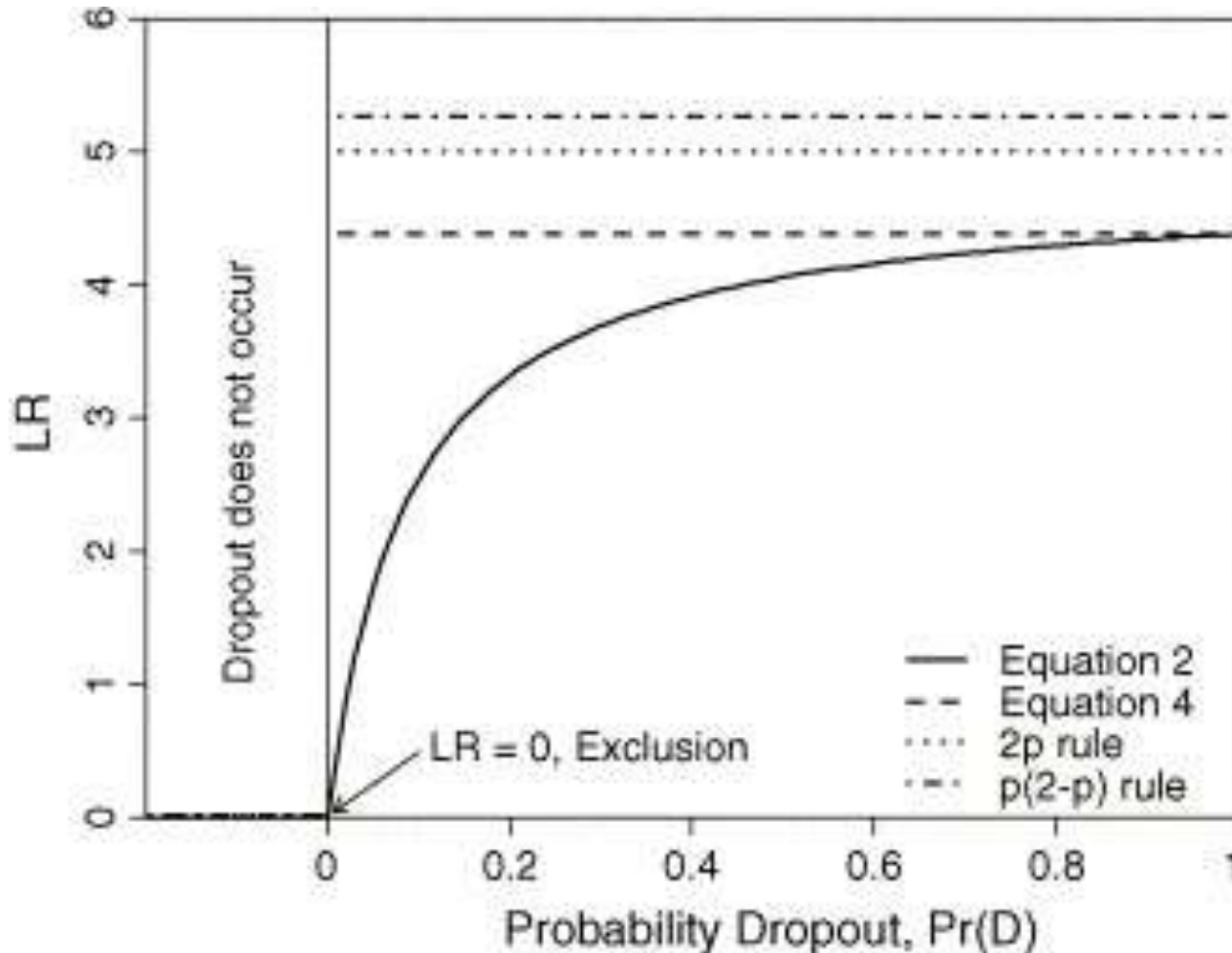
The “2p” Rule



Stain = aa

Suspect = aa

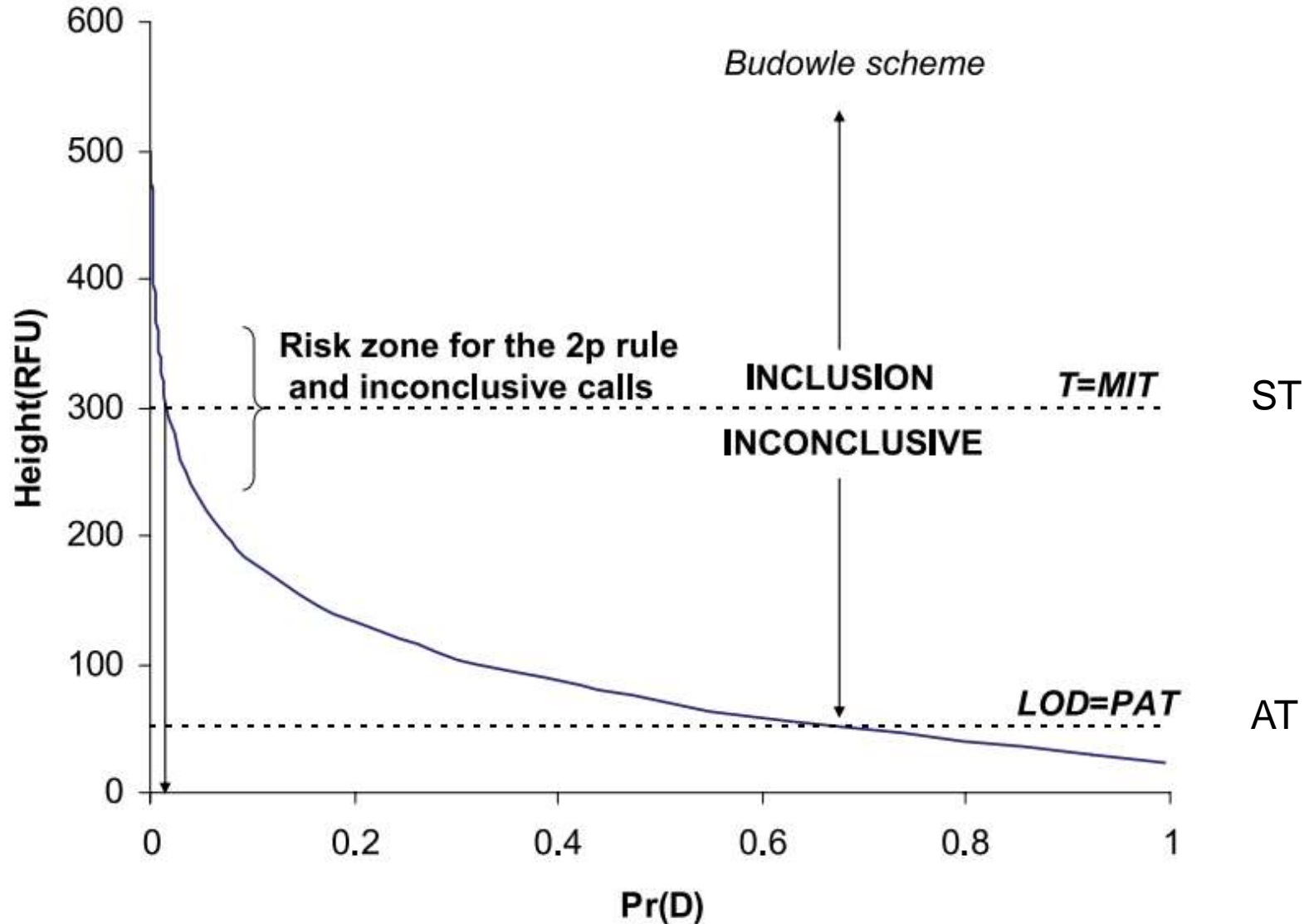
The “2p” Rule



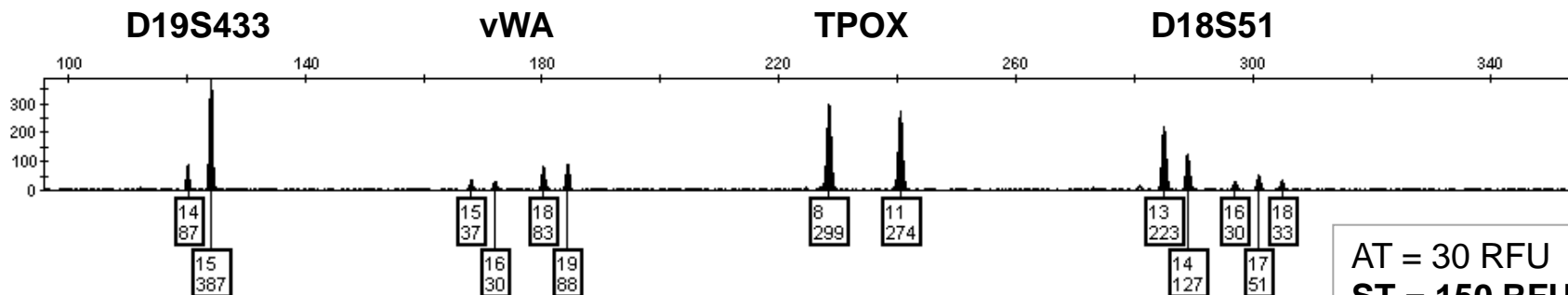
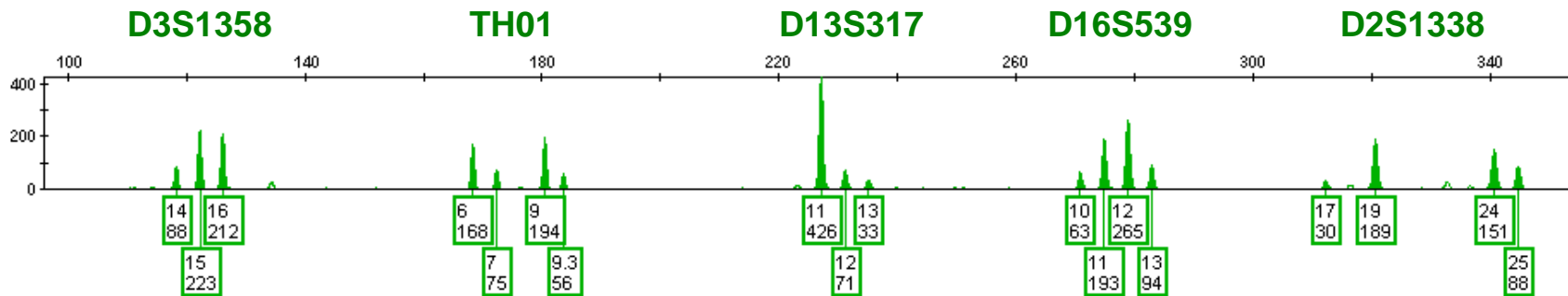
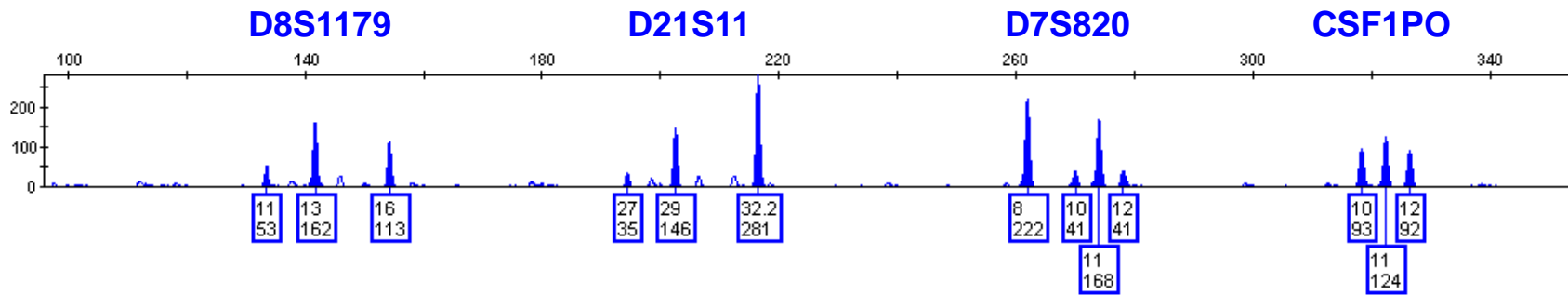
Stain = aa

Suspect = ab

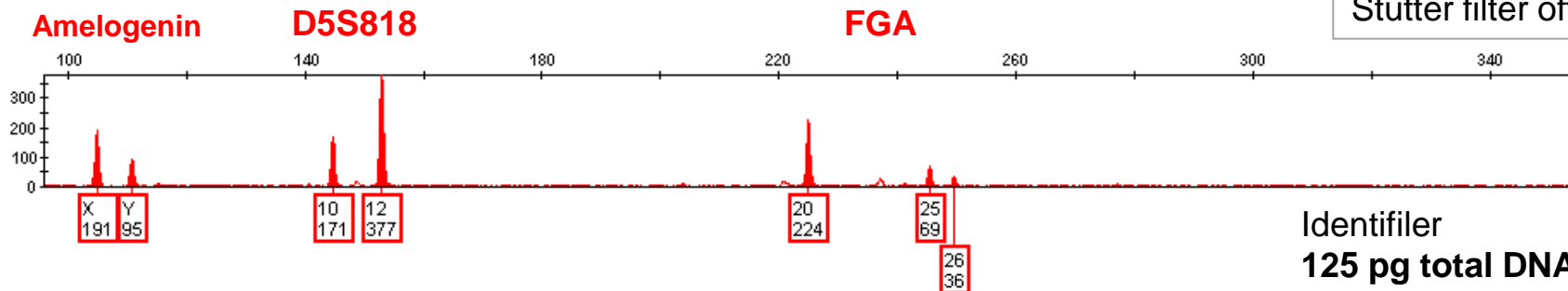
Gill and Buckleton (2010)



Challenges with low level,
complex mixtures



AT = 30 RFU
 ST = 150 RFU
 Stutter filter off



Identifiler
 125 pg total DNA

Clayton et al. (1998)
ISFG (2006) Rec. #4

Step #1

Identify the Presence of a Mixture

Step #2

Designate Allele Peaks

Step #3

Identify the Number of Potential Contributors

Step #4

Estimate the Relative Ratio of Contributors

Step #5

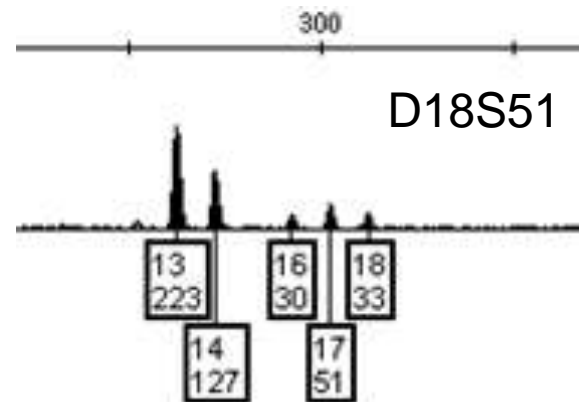
Consider All Possible Genotype Combinations

Step #6

Compare Reference Samples

Impact of Results with Low Level DNA

When amplifying low amounts of DNA (e.g., 125 pg), allele dropout is a likely possibility leading to **higher uncertainty** in the potential number of contributors and in the possible genotype combinations

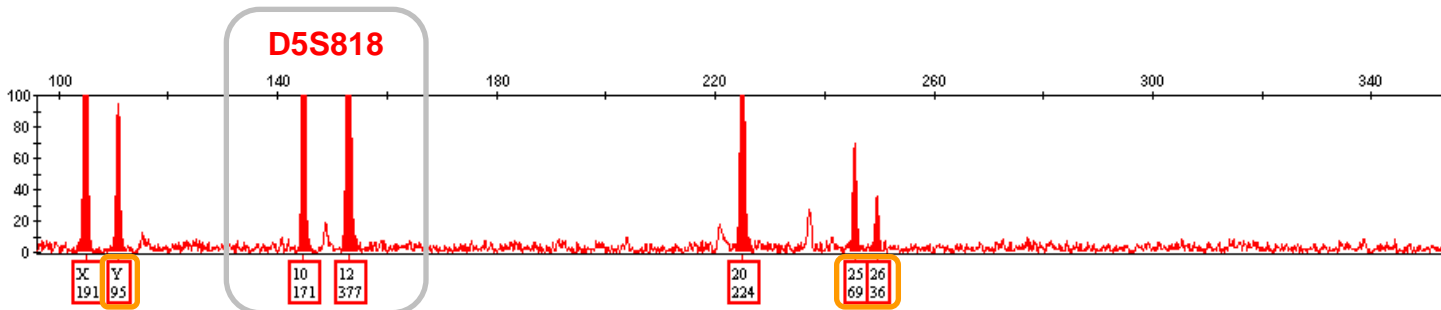
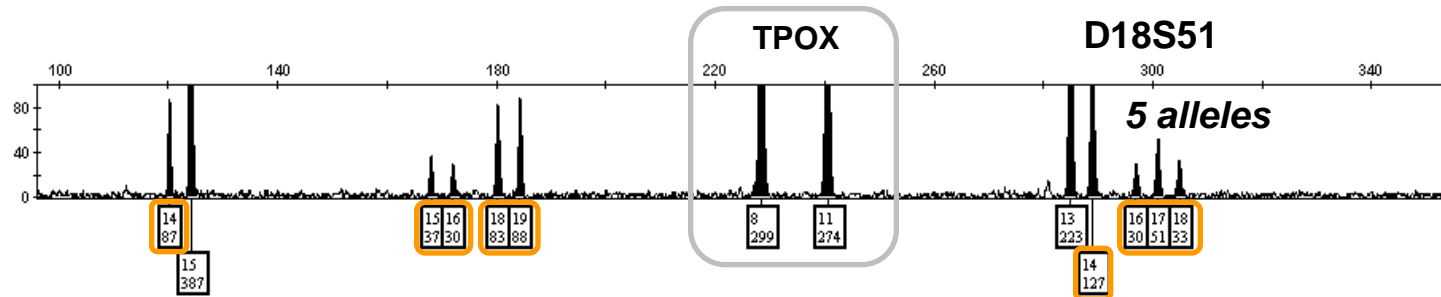
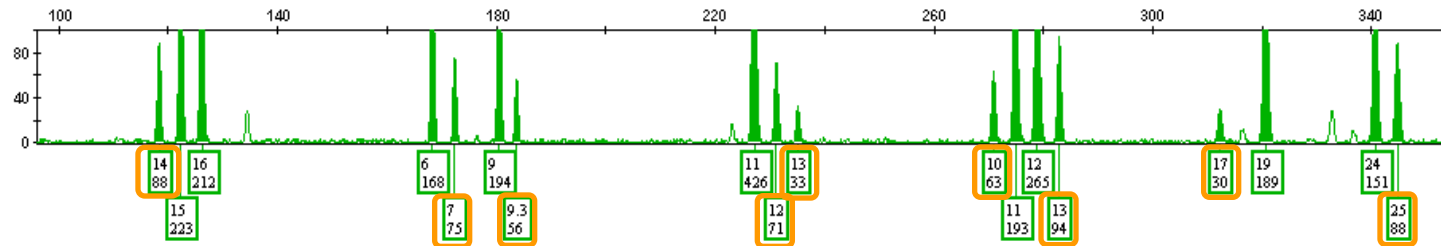
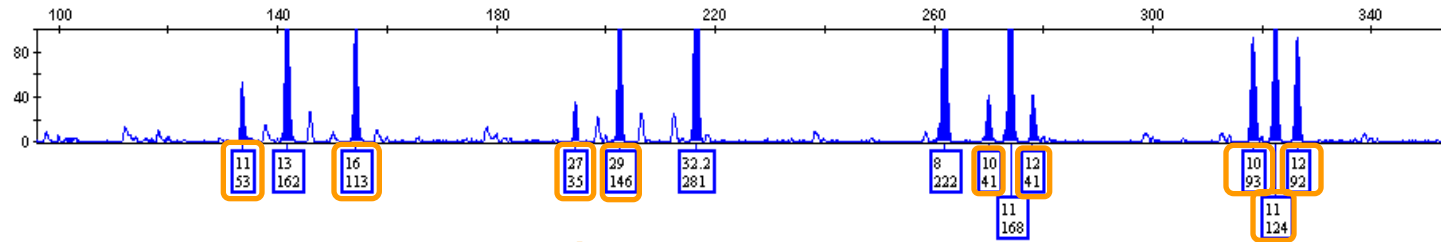


Identifiler
125 pg total DNA

Complex Mixture

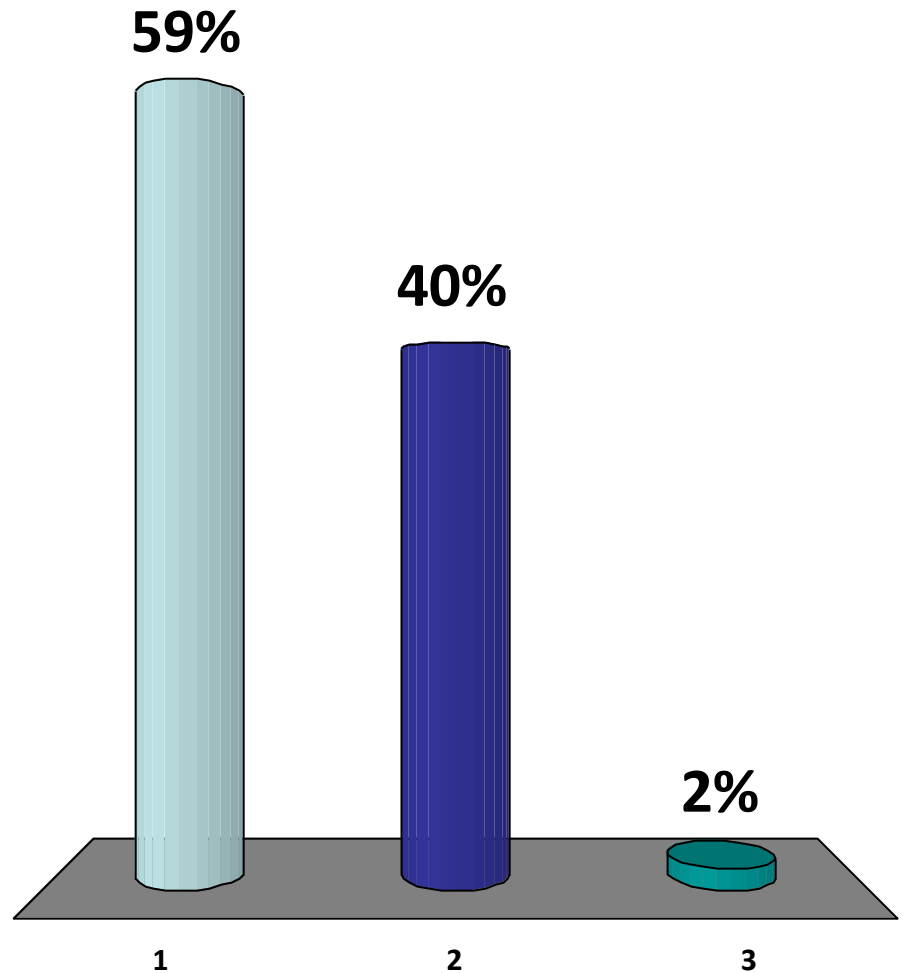
AT = 30 RFU
ST = 150 RFU
Stutter filter off

y-axis zoom to 100 RFU



Would you do a CPE/CPI statistic on TPOX and D5S818 because all alleles are above the stochastic threshold?

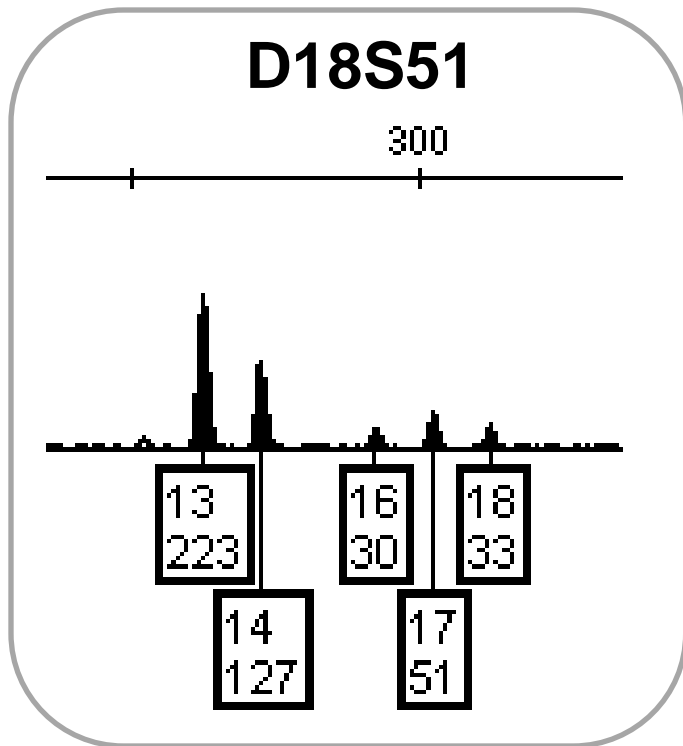
1. Yes
2. No
3. I don't work in a lab



What Can We Say about this Result?

- Low level DNA (only amplified 125 pg total DNA)
 - likely to exhibit stochastic effects and have allele dropout
- Mixture of at least 3 contributors
 - Based on detection of 5 alleles at D18S51
 - If at equal amounts, ~40 pg of each contributor (if not equal, then less for the minor contributors); **we expect allele dropout**
- At least one of the contributors is male
 - Based on presence of Y allele at amelogenin
- Statistics if using CPI/CPE
 - Would appear that we can only use TPOX and D5S818 results with a stochastic threshold of 150 RFU (*will explore this further*)
- **Due to potential of excessive allele dropout, we are unable to perform any meaningful Q-K comparisons**

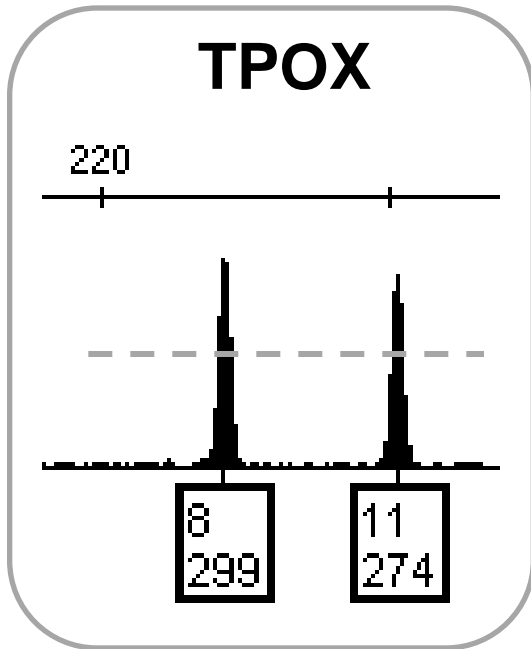
Uncertainty in the Potential Number of Contributors with this Result



5 alleles observed

- Several of the peaks are barely above the analytical threshold of 30 RFU
In fact, with an analytical threshold of 50 RFU or even 35 RFU, there would only be three detected alleles at D18S51
- Stochastic effects could result in a high degree of stutter off of the 17 allele making alleles 16 and 18 potential stutter products
- No other loci have >4 alleles detected

All Detected Alleles Are Above the Stochastic Threshold – **Or Are They?**



Stochastic
threshold =
150 RFU

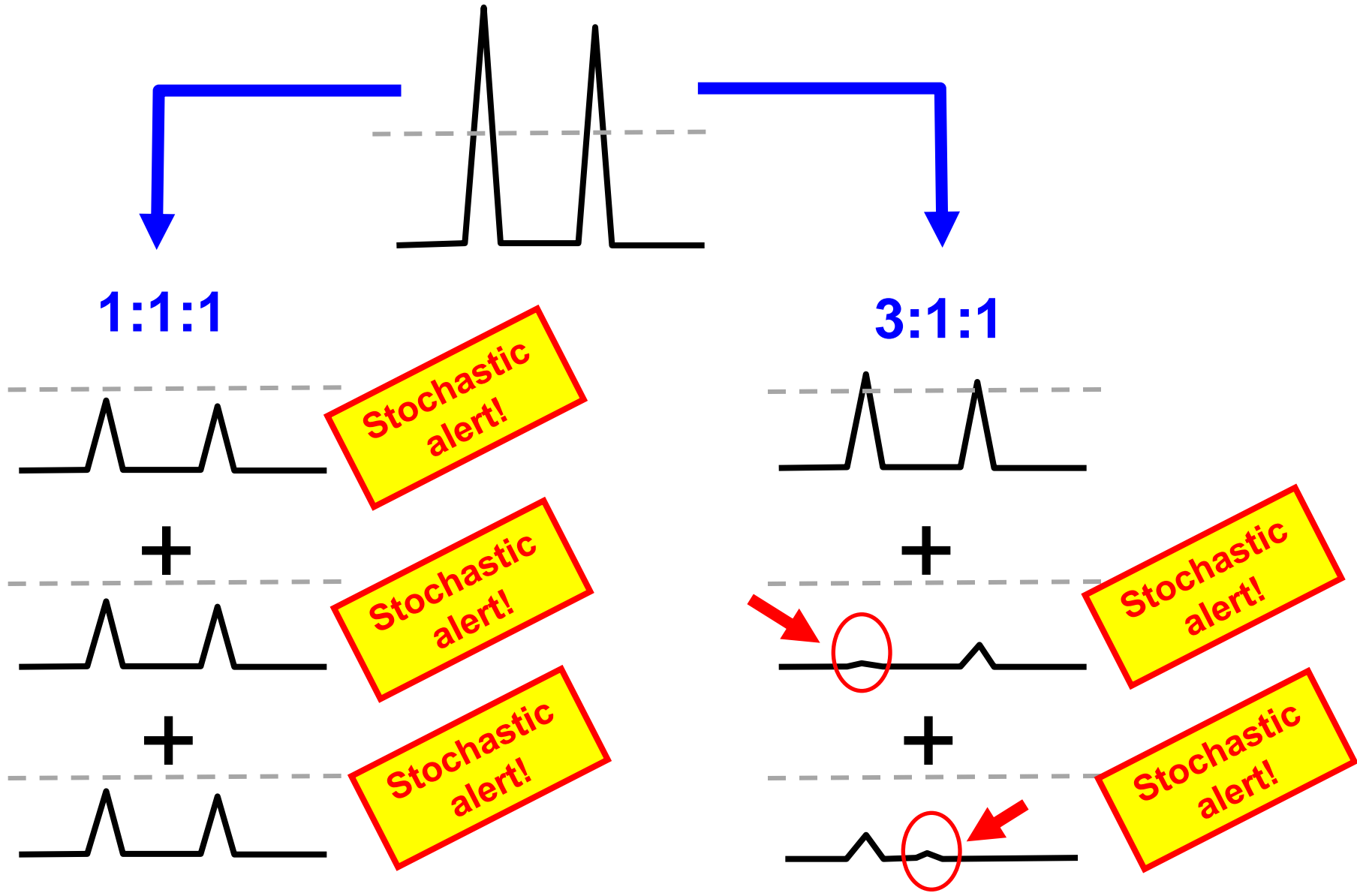
Does this result guarantee no allele drop-out?

We have assumed three contributors. If result is from an equal contribution of 3 individuals...

Then some alleles from individual contributors would be below the stochastic threshold and we could not assume that all alleles are being observed!

Assuming Three Contributors...

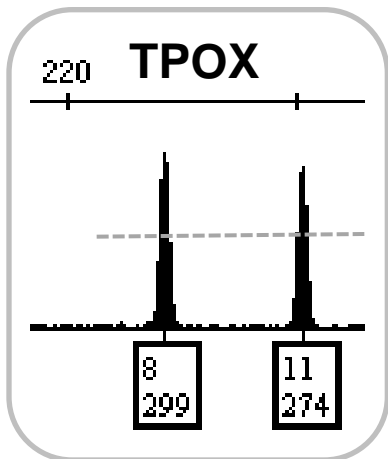
Some Possible Contributions to This Result



All Loci Are Not Created Equal when it comes to mixture interpretation

- In the case of less polymorphic loci, such as TPOX, there are fewer alleles and these occur at higher frequency. Thus, there is a greater chance of allele sharing (peak height stacking) in mixtures.
- **Higher locus heterozygosity is advantageous for mixture interpretation** – we would expect to see more alleles (within and between contributors) and thus have a better chance of estimating the true number of contributors to the mixture

Even if you did attempt to calculate a CPI/CPE statistic using loci with all observed alleles above the stochastic threshold on this result...



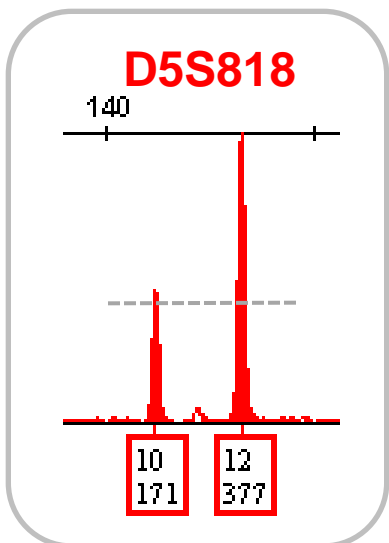
TPOX Allele Frequencies (NIST Caucasian, Butler et al. 2003)

8 = 0.53

11 = 0.24

CPI = $(0.53 + 0.24)^2 = 0.59$ or **59%**

Combine loci = $0.59 \times 0.18 = 0.11$ or **11%**
Approximately 1 in every 9 Caucasians could be included in this mixture



D5S818 Allele Frequencies (NIST Caucasian, Butler et al. 2003)

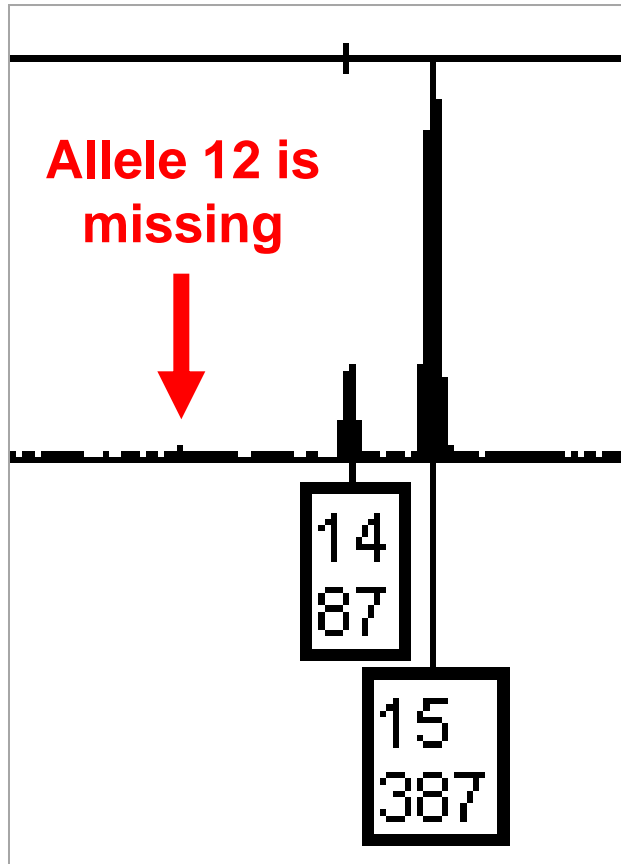
10 = 0.05

12 = 0.38

CPI = $(0.05 + 0.38)^2 = 0.18$ or **18%**

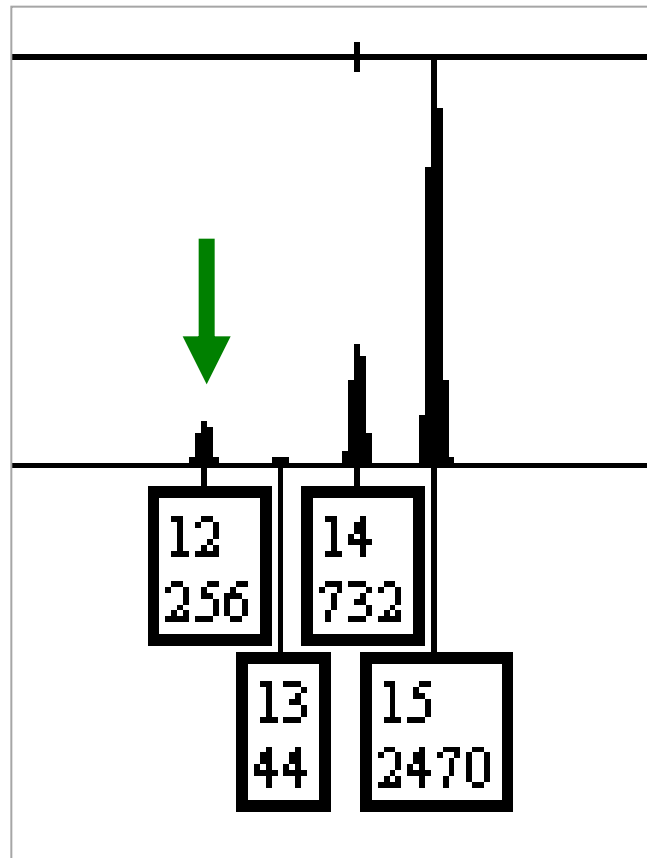
Impact of Amplifying More DNA

D19S433



125 pg total DNA amplified

D19S433



500 pg total DNA amplified

True Contributors
3 contributors
with a 2:1:1 mixture

15,15 (2x)
14,15 (1x)
12,14 (1x)

How should you handle the suspect comparison(s) with this case result?

- **No suspect comparisons should be made** as the mixture result has too much uncertainty with stochastic effects that may not account for all alleles being detected
- **Declare the result “inconclusive”**

How not to handle this result

- “To heck with the analytical and stochastic thresholds”, **I am just going to see if the suspect profile(s) can fit into the mixture allele pattern observed** – and then if an allele is not present in the evidentiary sample try to explain it with possible allele dropout due to stochastic effects
- This is what Bill Thompson calls “painting the target around the arrow (matching profile)...”

What to do with low level DNA mixtures?

- **German Stain Commission “Category C”**

(Schneider et al. 2006, 2009)

- Cannot perform stats because stochastic effects make it uncertain that all alleles are accounted for

- **ISFG Recommendations #8 & #9** (Gill et al. 2006)

- Stochastic effects limit usefulness

- *Fundamentals of Forensic DNA Typing* (2010)

Butler 3rd edition (volume 1), chapter 18

- Don't go “outside the box” without supporting validation



ISFG Recommendations on Mixture Interpretation

<http://www.isfg.org/Publication;Gill2006>

1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
2. Scientists should be trained in and use LRs
3. Methods to calculate LRs of mixtures are cited
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
5. Prosecution determines H_p and defense determines H_d and multiple propositions may be evaluated
6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
7. Allele dropout to explain evidence can only be used with low signal data
8. No statistical interpretation should be performed on alleles below threshold
9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA



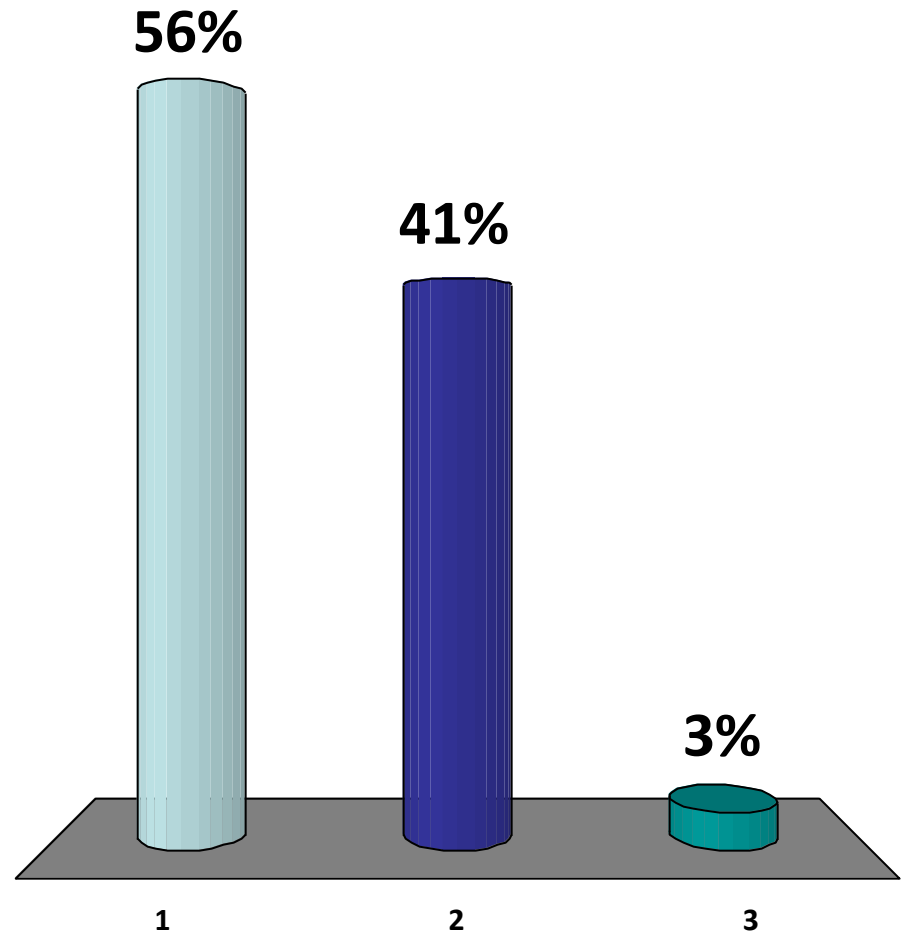
A Complexity/Uncertainty Threshold

New Scientist article (August 2010)

- **How DNA evidence creates victims of chance**
 - 18 August 2010 by Linda Geddes
- From the last paragraph:
 - **In really complex cases, analysts need to be able to draw a line** and say "This is just too complex, I can't make the call on it," says Butler. "Part of the challenge now, is that every lab has that line set at a different place. But the honest thing to do as a scientist is to say: **I'm not going to try to get something that won't be reliable.**"

Has your laboratory implemented a “stop testing” approach with complex and/or low-level mixture?

1. Yes
2. No
3. I don't work in a lab



Is there a way forward?

“On the Threshold of a Dilemma”

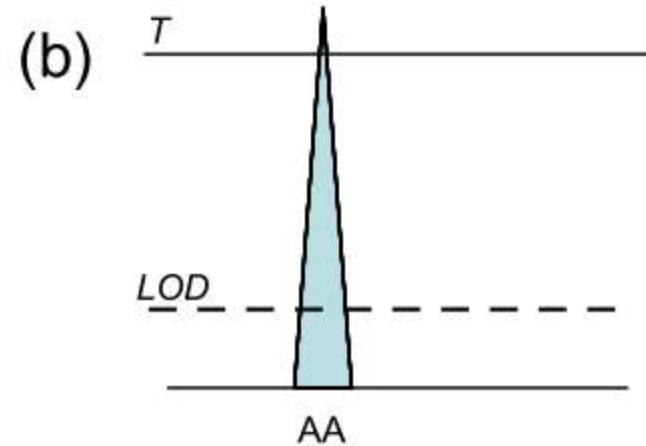
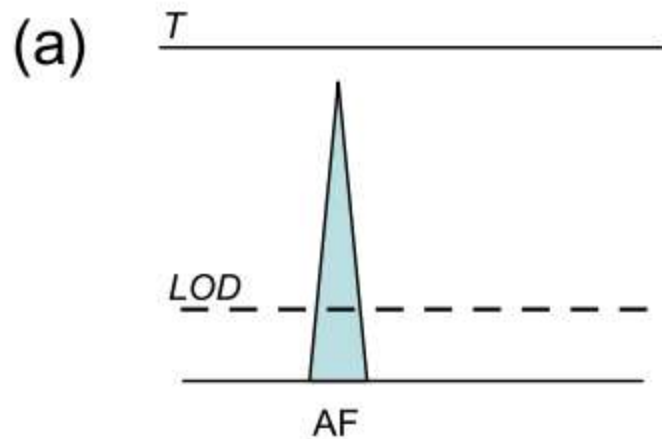
- Gill and Buckleton (2010)
- Although most labs use thresholds of some description, this philosophy has always been problematic because there is an inherent illogicality which we call the falling off the cliff effect.



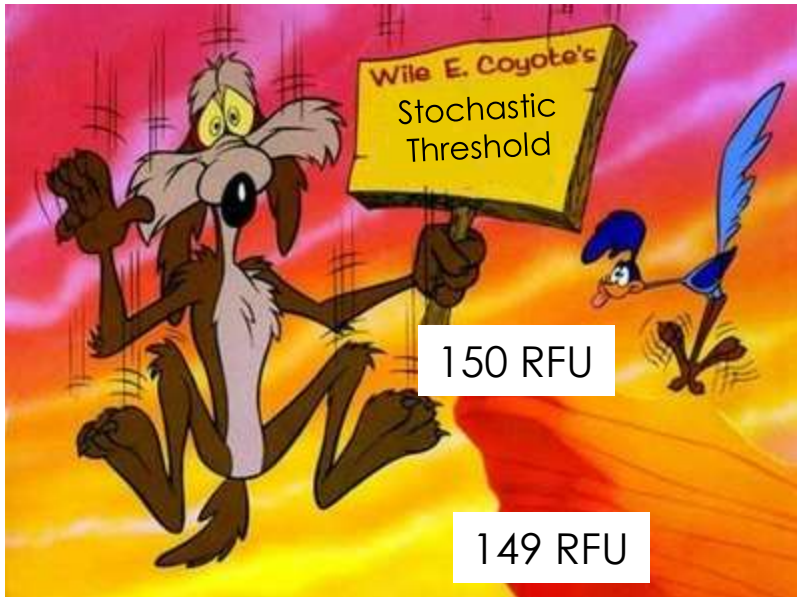
Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttmann JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J Forensic Sci* 2009;54(4):810–21.

“Falling off the Cliff Effect”

- If T = an arbitrary level (e.g., 150 rfu), an allele of 149 rfu is subject to a different set of guidelines compared with one that is 150 rfu even though they differ by just 1 rfu (Fig. 1).



Falling off the Cliff vs. Gradual Decline



<http://blog.sironaconsulting.com/.a/6a00d8341c761a53ef011168cc5ff3970c-pi>



<http://ultimateescapesdc.files.wordpress.com/2010/08/mountainbiking2.jpg>

Gill and Buckleton *JFS* **55: 265-268 (2010)**

- “The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of ***probabilistic models to circumvent the requirement for a threshold*** and to safeguard the legitimate interests of defendants.”

PAPER

J Forensic Sci, 2011
doi: 10.1111/j.1556-4029.2011.01859.x
Available online at: onlinelibrary.wiley.com

CRIMINALISTICS

Mark W. Perlin,¹ M.D., Ph.D.; Matthew M. Legler,¹ B.S.; Cara E. Spencer,¹ M.S.; Jessica L. Smith,¹ M.S.; William P. Allan,¹ M.S.; Jamie L. Belrose,² M.S.; and Barry W. Ducean,³ Ph.D.

Validating TrueAllele[®] DNA Mixture Interpretation^{*,†}

- Quantitative computer interpretation using Markov Chain Monte Carlo testing
- Models peak uncertainty and infers possible genotypes
- Results are presented as the Combined LR



True Allele Software (Cybergenetics)

- We purchased the software in September 2010.
- Three day training at Cybergenetics (Pittsburgh, PA) in October.
- Software runs on a Linux Server with a Mac interface.



True Allele Casework Workflow

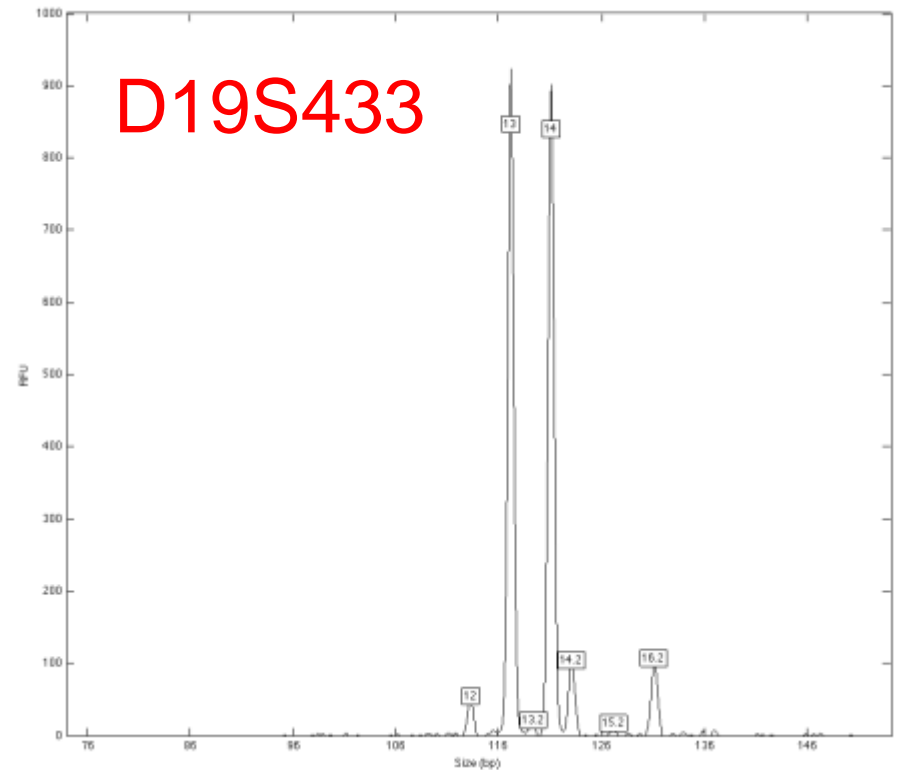
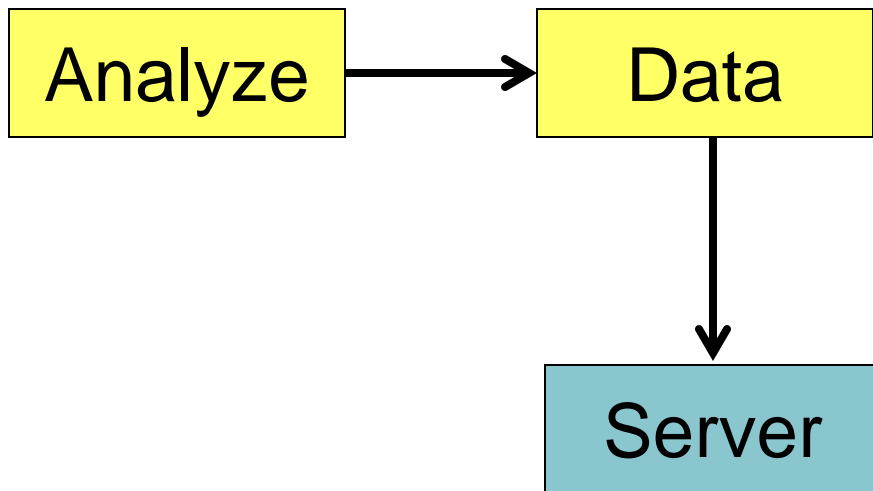
5 Modules

Analyze

.fsa files imported
Size Standard check
Allelic Ladder check
Alleles are called

True Allele Casework Workflow

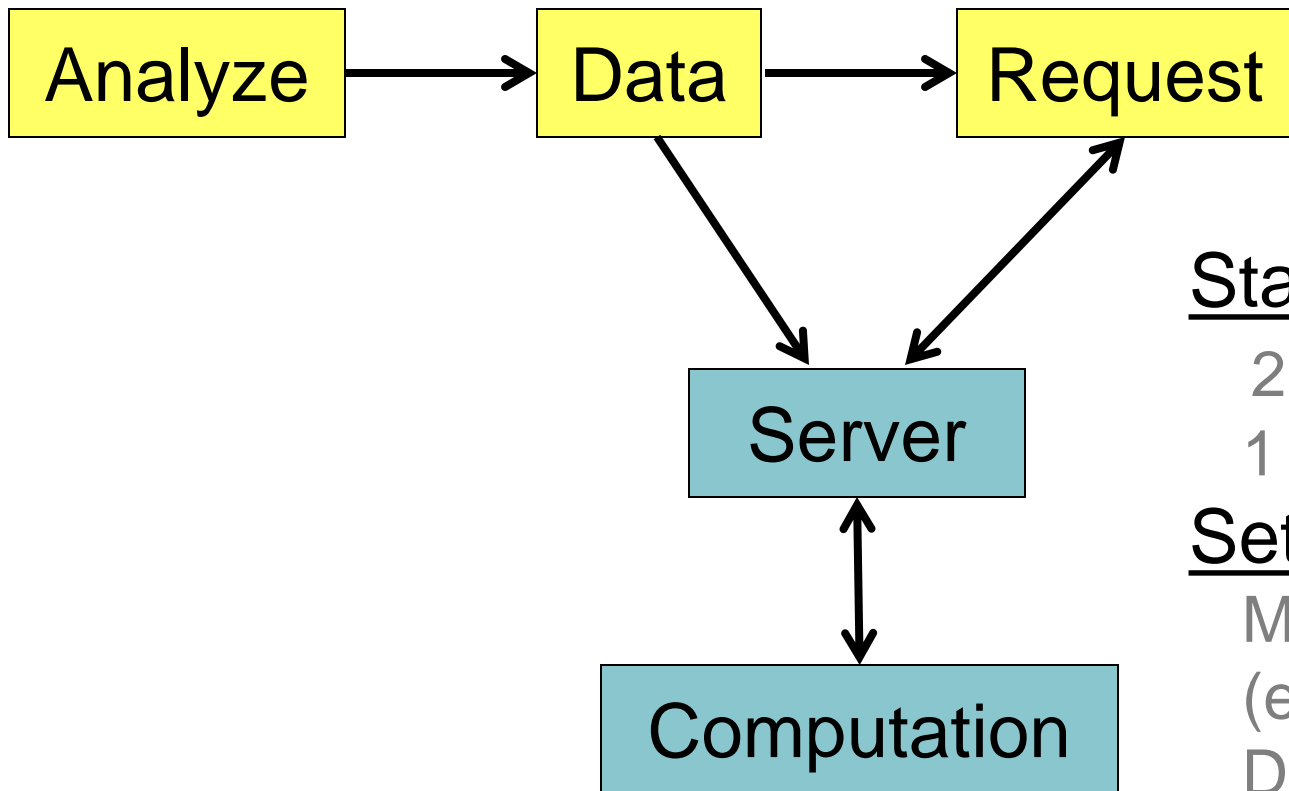
5 Modules



All Peaks above 10 RFU are considered

True Allele Casework Workflow

5 Modules



State Assumptions

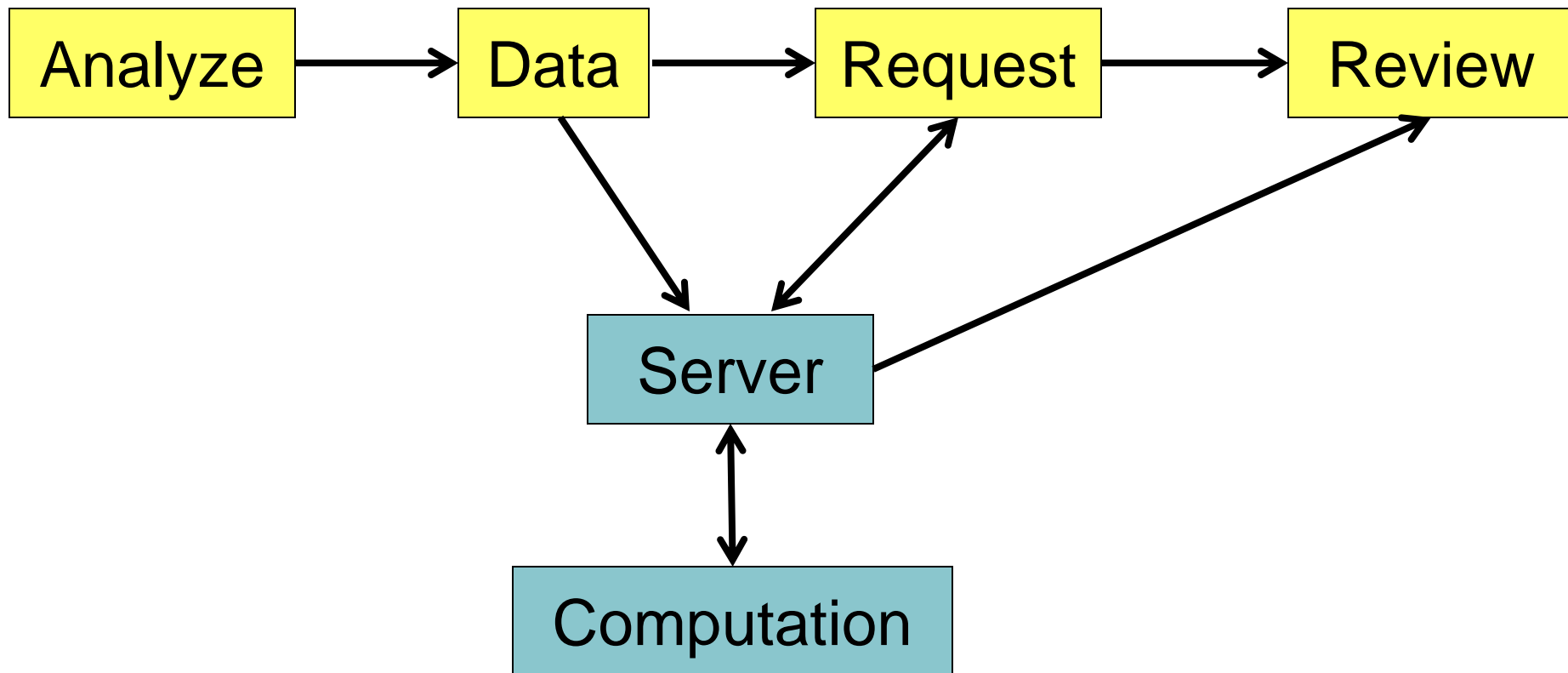
2, 3, 4 unknowns
1 Unk with Victim?

Set Parameters

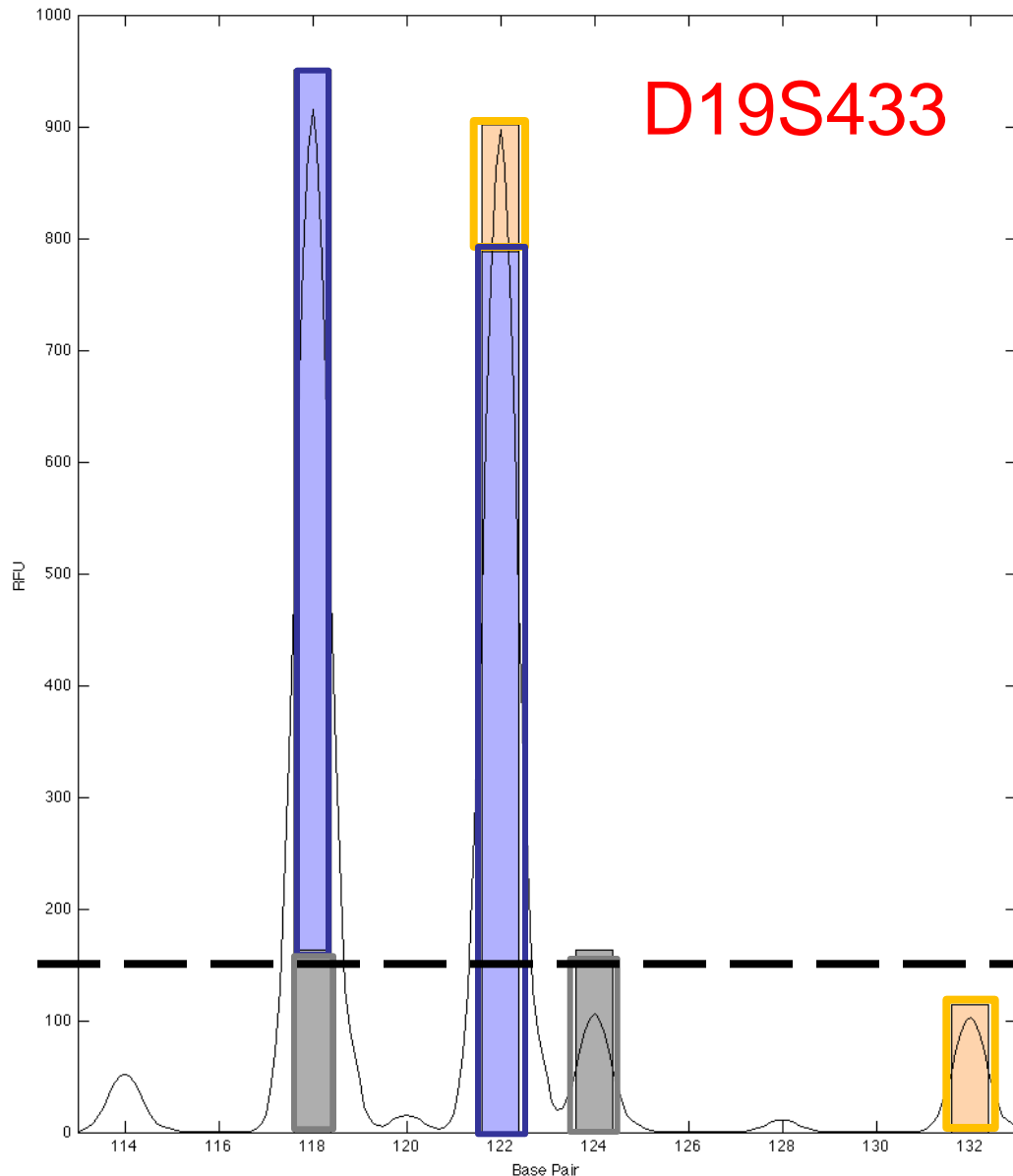
MCMC modeling
(e.g. 50K)
Degradation?

True Allele Casework Workflow

5 Modules



Review of One Replicate (of 50K)



3P mixture,
2 Unknowns,

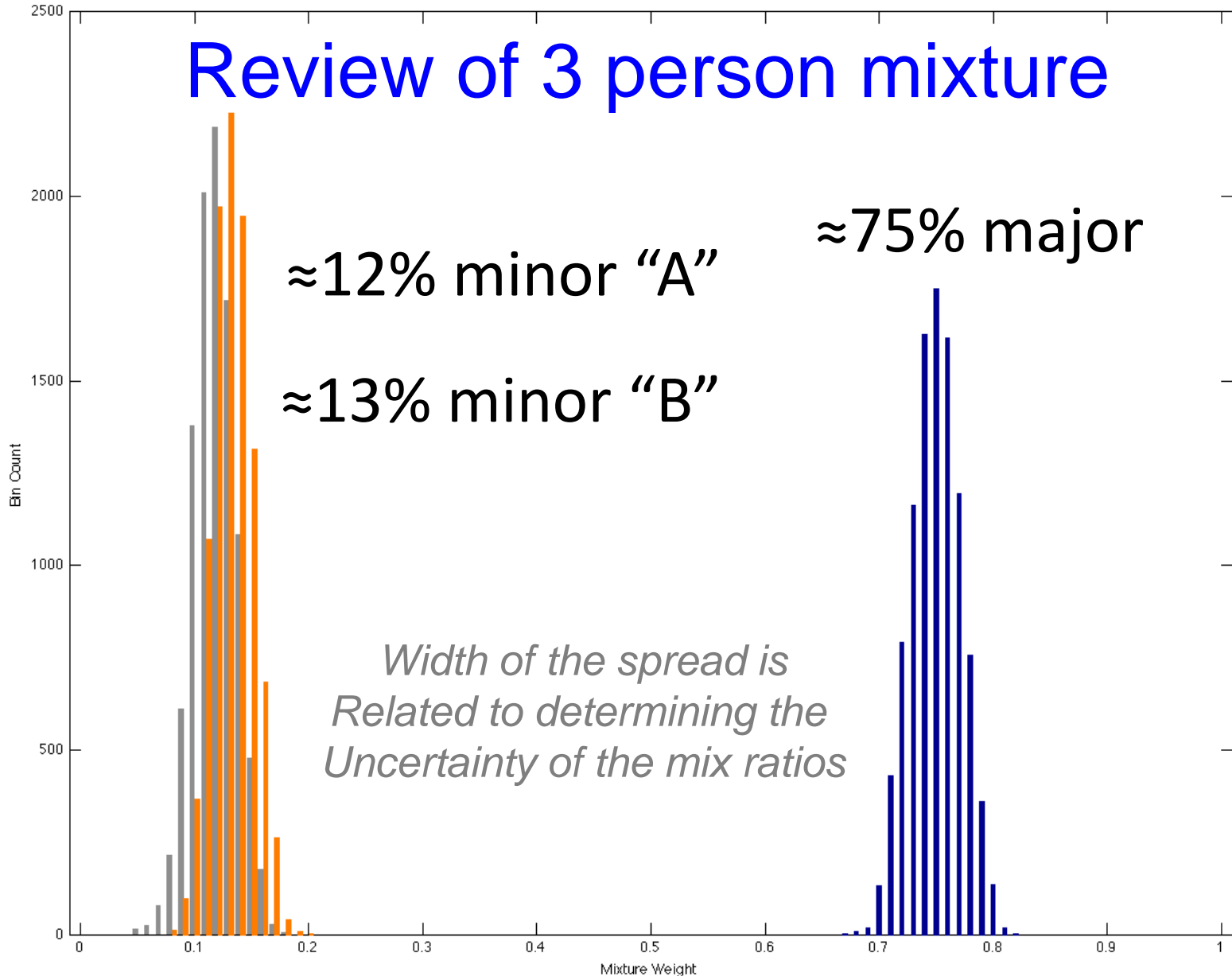
Conditioned
on the Victim
(major)

Good fit of the
data to the model

— 150 RFU

Review of 3 person mixture

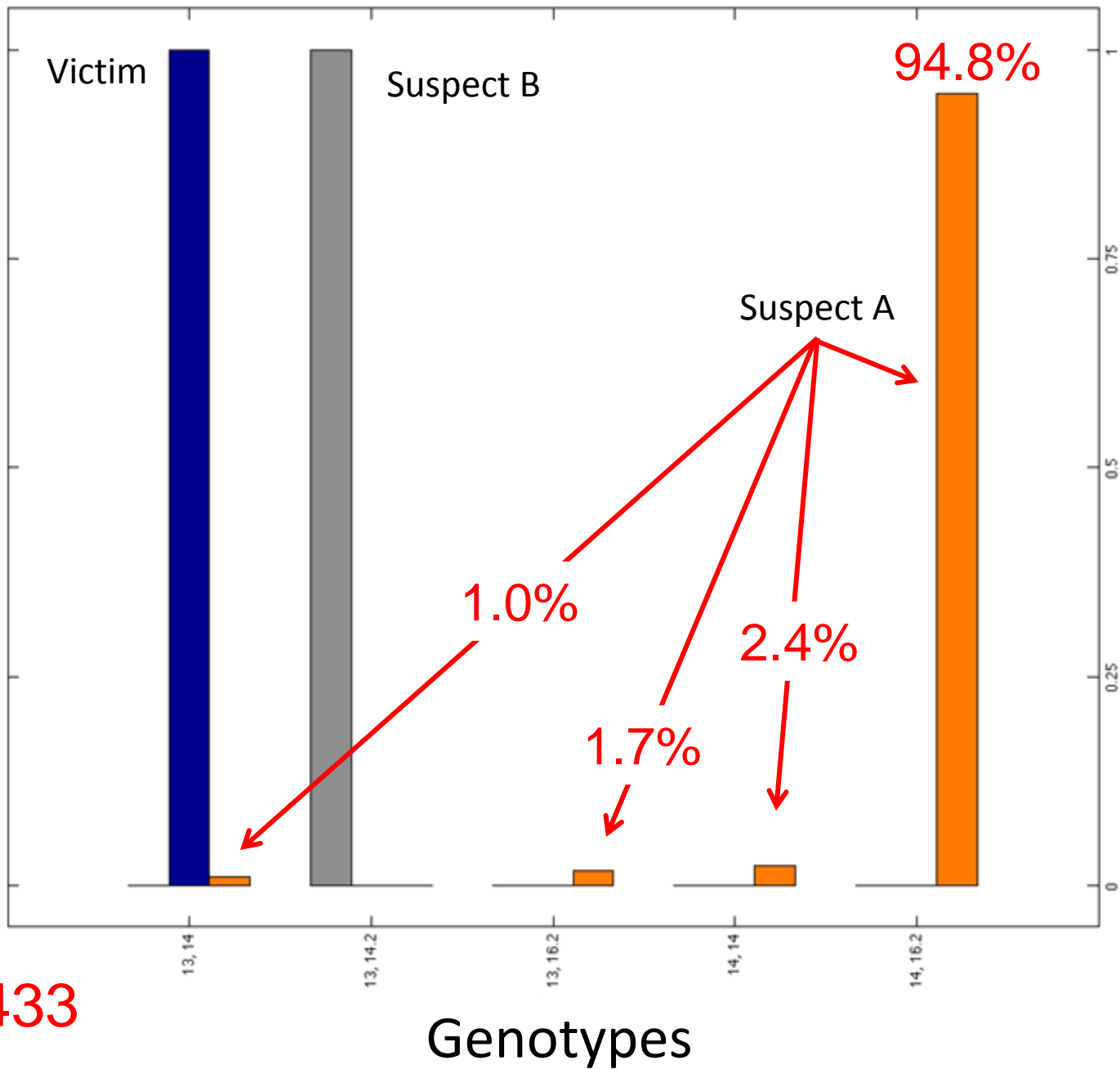
Bin Count



Mixture Weight

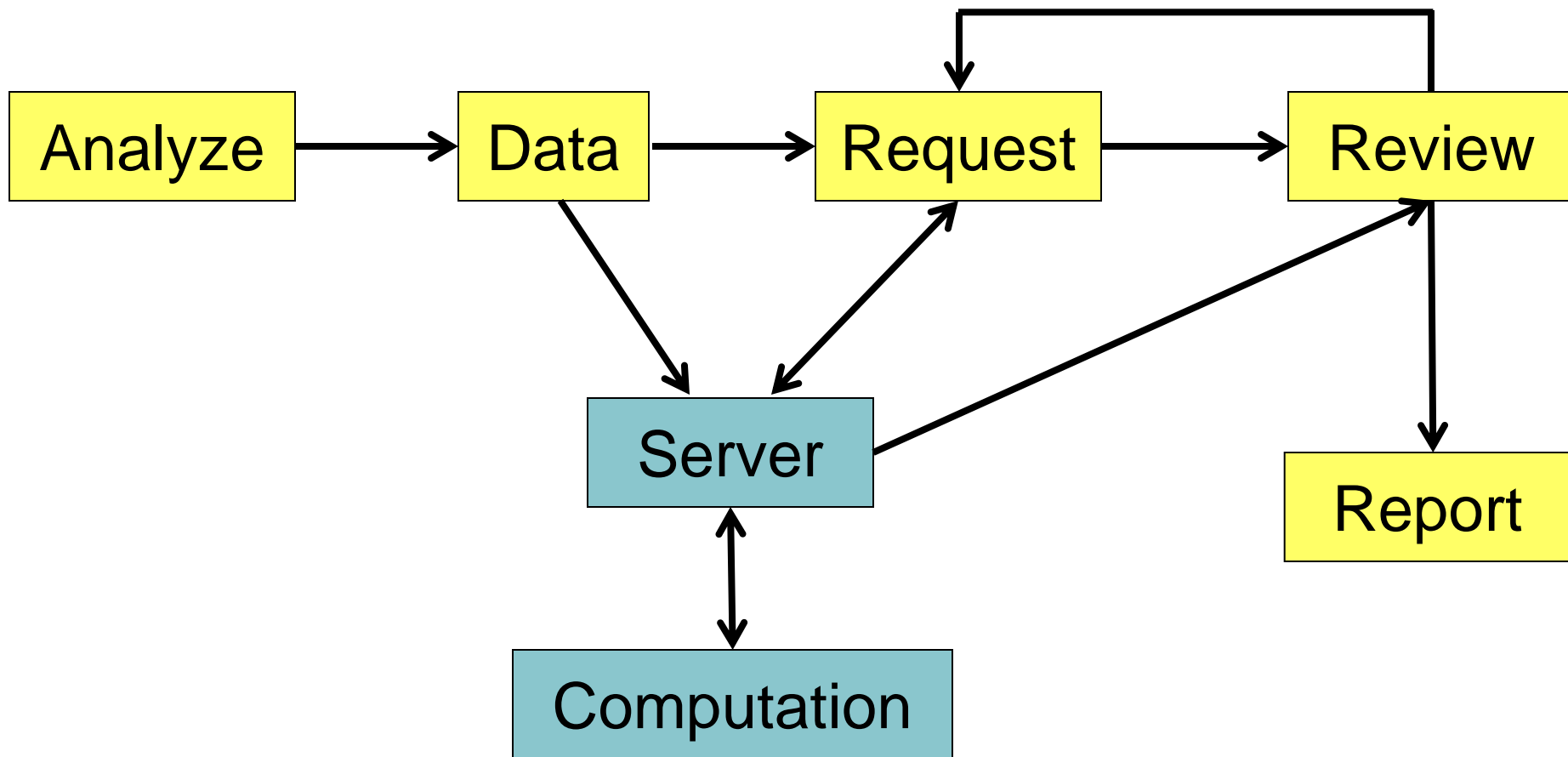
D19S433

Genotype Probability



True Allele Casework Workflow

5 Modules



Determining the LR for D19S433

Suspect A = 14, 16.2

$$H_p = 0.967$$

	Allele Pair	Probability Before Conditioning
→	14, 16.2	0.967
	14, 14	0.003
	13, 16.2	0.026
	13, 14	0.001

$$LR = \frac{0.967}{\quad}$$

Determining the LR for D19S433

Suspect A = 14, 16.2

$H_P = 0.967$

Allele Pair	Probability Before Conditioning	Genotype Frequency	Probability * Genotype Freq
14, 16.2	0.967	0.0120	0.01164
14, 14	0.003	0.0498	0.00013
13, 16.2	0.026	0.0131	0.00034
13, 14	0.001	0.1082	0.00009
		sum	0.0122

$$LR = \frac{0.967}{0.0122} = 79.26 \quad H_D$$

Combined LR = 5.6 Quintillion

locus	allele pair x	Likelihood l(x)	Genotype Probability Distribution			Weighted Likelihood		Likelihood Ratio	
			Questioned q(x)	Reference r(x)	Suspect s(x)	Numerator l(x)*s(x)	Denominator l(x)*r(x)	LR	log(LR)
CSF1PO	11, 12	0.686	0.778	0.1448	1	0.68615	0.1292	5.31	0.725
D13S317	9, 12	1	1	0.0291	1	0.99952	0.02913	34.301	1.535
D16S539	9, 11	0.985	0.995	0.1238	1	0.98451	0.12188	8.036	0.905
D18S51	13, 17	0.999	1	0.0154	1	0.99915	0.01543	64.677	1.811
D19S433	14, 16.2	0.967	0.948	0.012	1	0.96715	0.01222	79.143	1.898
D21S11	28, 30	0.968	0.98	0.0872	1	0.96809	0.08648	11.194	1.049
D2S1338	23, 24	0.998	1	0.0179	1	0.99831	0.01787	55.866	1.747
D3S1358	15, 17	0.988	0.994	0.1224	1	0.98759	0.12084	8.14	0.911
D5S818	11, 11	0.451	0.394	0.0537	1	0.45103	0.07309	6.17	0.79
D7S820	11, 12	0.984	0.978	0.0356	1	0.98383	0.03617	27.198	1.435
D8S1179	13, 14	0.203	0.9	0.1293	1	0.20267	0.02993	6.771	0.831
FGA	21, 25	0.32	0.356	0.028	1	0.31986	0.01906	16.783	1.225
TH01	7, 7	0.887	0.985	0.1739	1	0.88661	0.15588	5.687	0.755
TPOX	8, 8	1	1	0.1375	1	1	0.13746	7.275	0.862
vWA	15, 20	0.998	0.996	0.0057	1	0.99808	0.00569	174.834	2.243

Results

- Results are expressed as logLR values

$$\text{LR} = 1,000,000 = 10^6$$

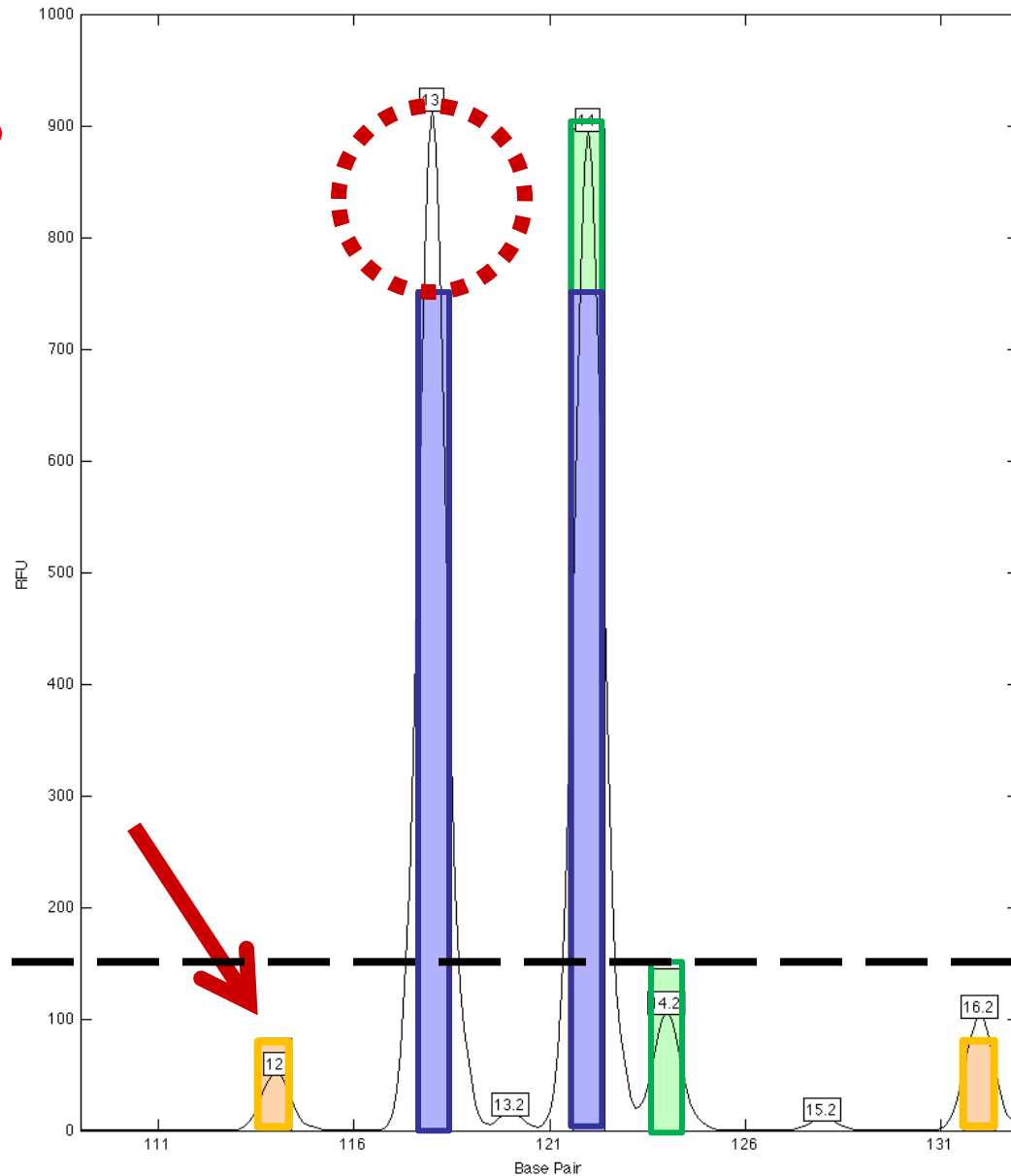
$$\log(\text{LR}) = \log 10^6$$

$$\log(\text{LR}) = 6 * \log 10 (1)$$

$$\log(\text{LR}) = 6$$

Review of One Replicate (of 50K)

D19S433



3P mixture,
3 Unknowns

Poor fit of the
data to the
model

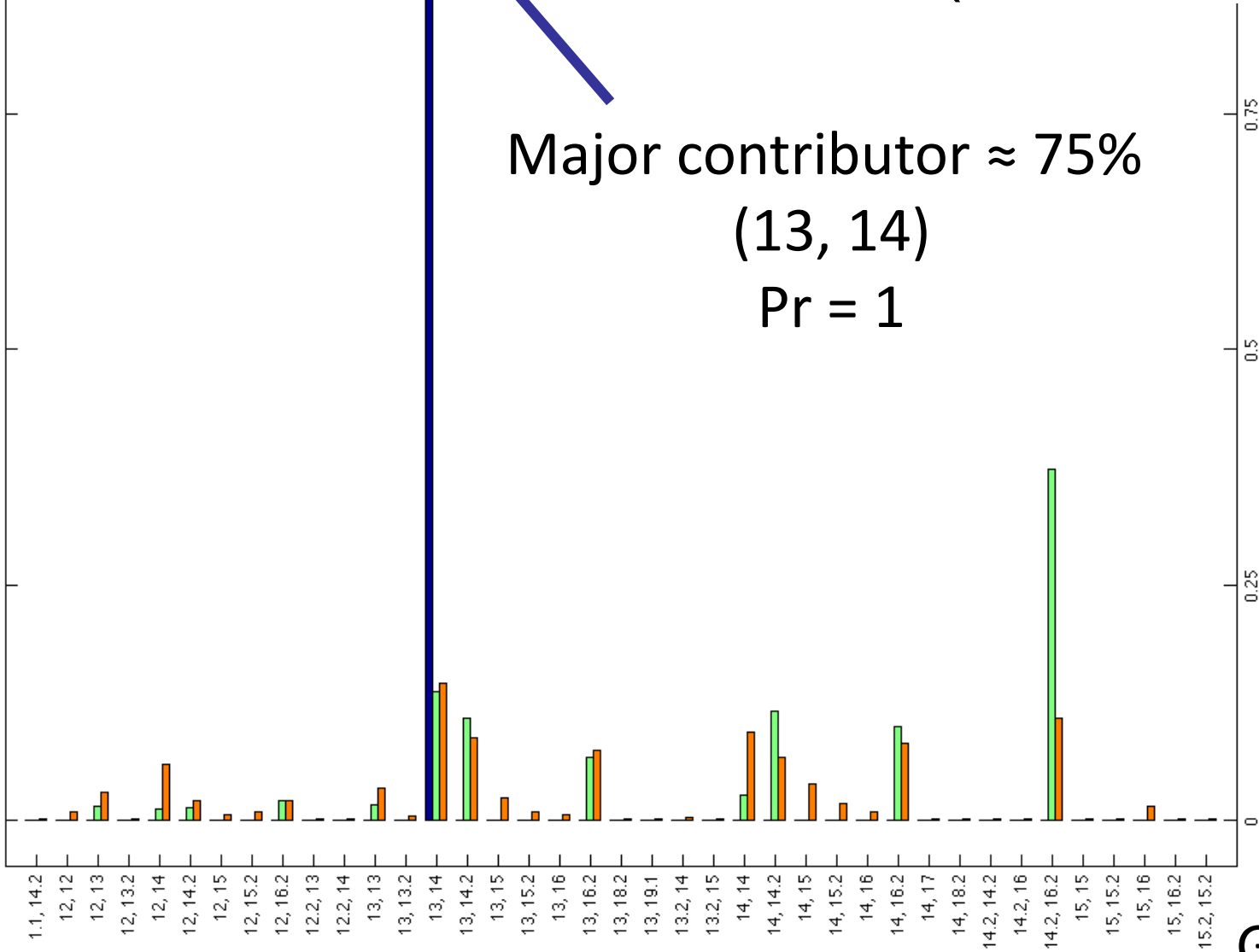
150 RFU

No Conditioning (3 Unknowns)

D19S433

Genotype Probability

Major contributor $\approx 75\%$
(13, 14)
Pr = 1

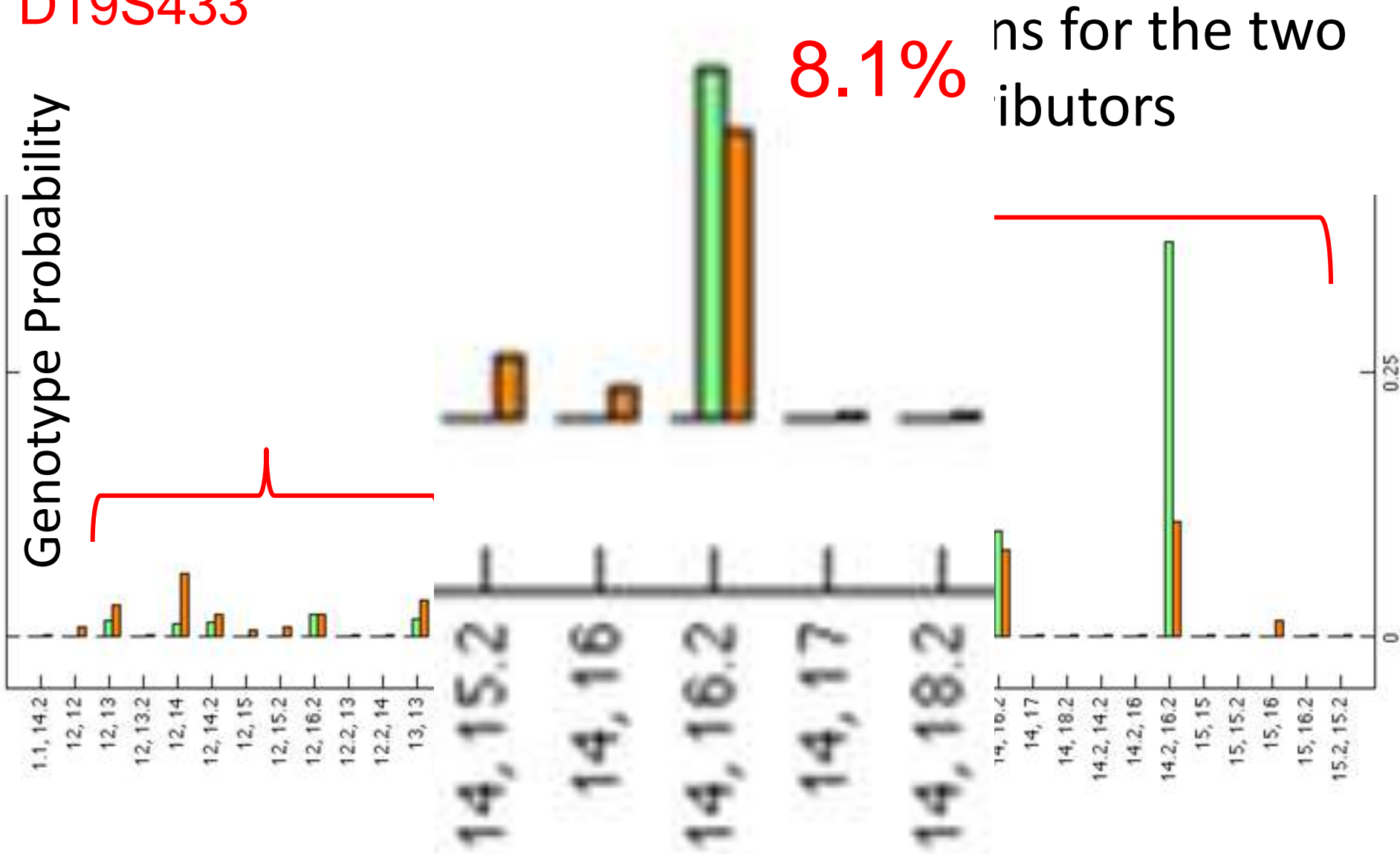


Genotypes

No Conditioning (3 Unknowns)

D19S433

Genotype Probability



locus	allele pair	L	Q	R	S	L*S	L*R	LR	log(LR)
D19S433	13 , 14	0.002	0.146	0.1082			0.00020		
	14.2, 16.2	0.270	0.109	0.0044			0.00118		
	14 , 14	0.002	0.093	0.0498			0.00008		
	13 , 14.2	0.017	0.088	0.0392			0.00068		
	14 , 16.2	0.013	0.081	0.0120	1	0.01295	0.00016		
	13 , 16.2	0.018	0.074	0.0131			0.00023		
	14 , 14.2	0.009	0.067	0.0361			0.00031		
	12 , 14	0.002	0.059	0.0498			0.00012		
	14 , 15	0.001	0.038	0.0343			0.00002		
	13 , 13	0.001	0.034	0.0587			0.00007		
	12 , 13	0.002	0.029	0.0541			0.00010		
	13 , 15	0.001	0.024	0.0373			0.00002		
	12 , 16.2	0.017	0.021	0.0060			0.00010		
	12 , 14.2	0.013	0.020	0.0180			0.00023		
	14 , 15.2	0.001	0.018	0.0275			0.00003		
	15 , 16	0.002	0.015	0.0006			0.00000		
	13 , 15.2	0.001	0.009	0.0299			0.00003		
	12 , 15.2	0.003	0.009	0.0137			0.00004		
	14 , 16	0.000	0.009	0.0017			0.00000		
	12 , 12	0.004	0.009	0.0125			0.00004		
	12 , 15	0.001	0.006	0.0172			0.00001		
	13 , 16	0.000	0.006	0.0019			0.00000		
	13 , 13.2	0.001	0.004	0.0261			0.00003		
	13.2, 14	0.001	0.003	0.0240			0.00002		
	13.2, 15	0.001	0.002	0.0083			0.00001		
	14 , 18.2	0.002	0.002	0.0017			0.00000		
	13 , 19.1	0.019	0.002	0.0000			0.00000		
	12 , 13.2	0.002	0.002	0.0120			0.00003		
	14.2, 16	0.001	0.002	0.0006			0.00000		
	12.2, 13	0.001	0.002	0.0168			0.00002		
	13 , 18.2	0.002	0.001	0.0019			0.00000		
	12.2, 14	0.001	0.001	0.0155			0.00001		
	14.2, 14.2	0.004	0.001	0.0065			0.00003		
	15 , 15	0.000	0.001	0.0059			0.00000		
	15 , 15.2	0.000	0.001	0.0095			0.00000		
	14 , 17	0.001	0.001	0.0000			0.00000		
	15 , 16.2	0.000	0.001	0.0042			0.00000		
	15.2, 15.2	0.001	0.001	0.0038			0.00000		
	1.1, 14.2	0.072	0.001	0.0097			0.00069		
						0.01295	0.00385	3.367	0.527

Suspect "A"
Genotype

39 probable
genotypes

D19S433

Suspect A = 14, 16.2

$$H_P = 0.013$$

Allele Pair	Probability	Genotype Frequency	Prob * GenFreq
13,14	0.002	0.1082	0.00020
14.2, 16.2	0.270	0.0044	0.00118
14, 14	0.002	0.0498	0.00008
13, 14.2	0.017	0.0392	0.00068
14, 16.2	0.013	0.0120	0.00016
13, 16.2	0.018	0.0131	0.00023
etc...	etc...	etc...	etc...
		Sum	0.00385

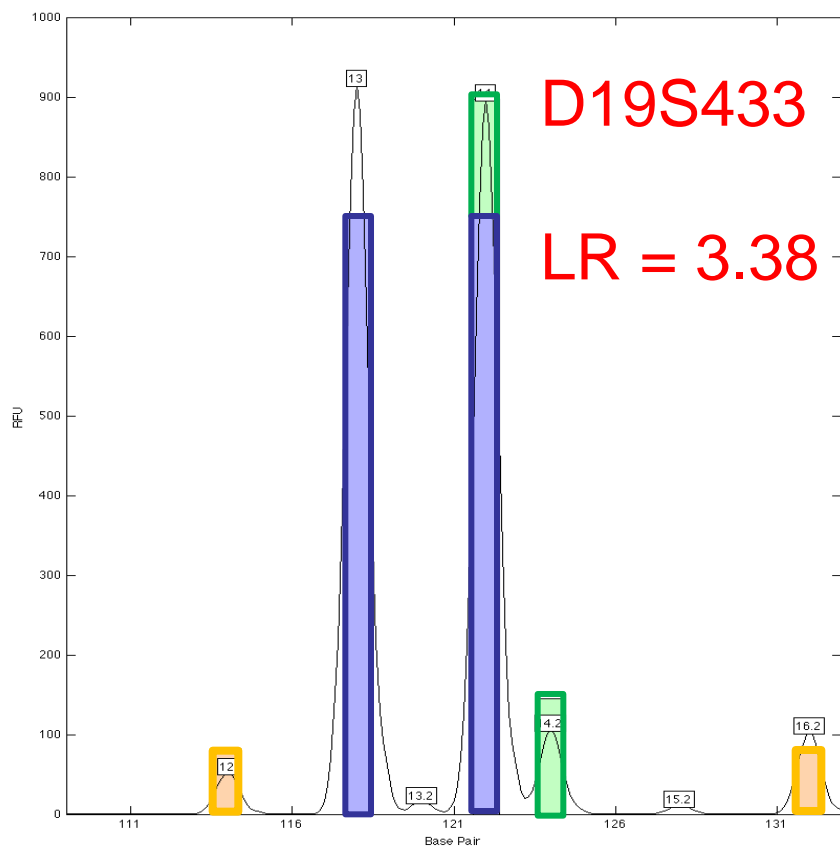
H_D

$$LR = \frac{0.013}{0.00385} = 3.38$$

D19S433

No Conditioning (3 Unknowns)

No Conditioning

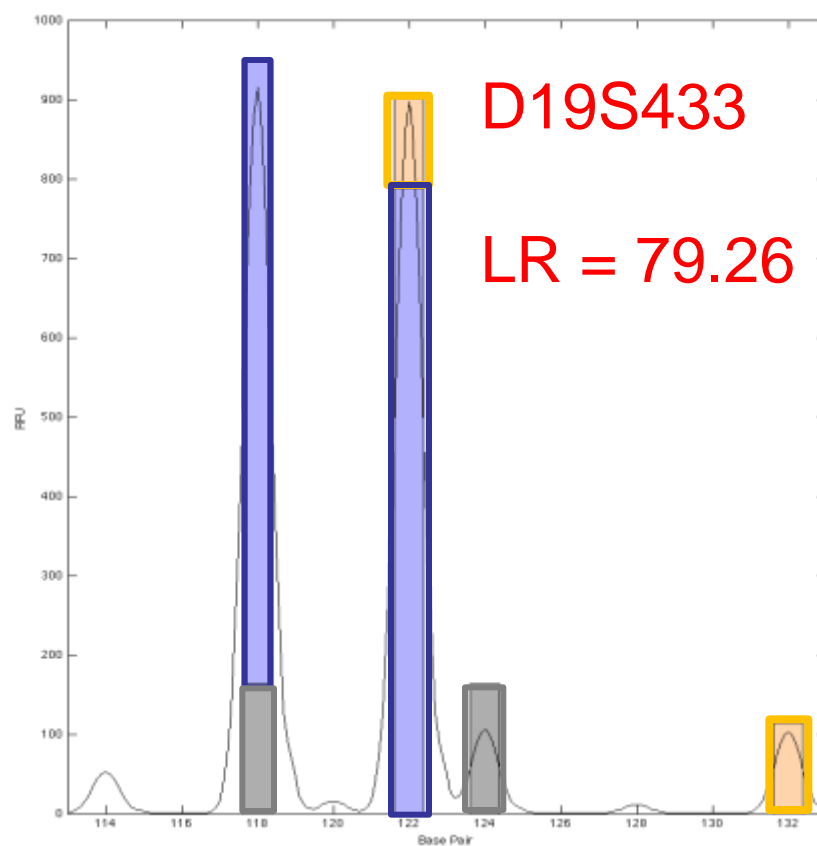


Profile - Combined $\log(\text{LR})$

Suspect A $\log(\text{LR}) = 8.03$

Suspect B $\log(\text{LR}) = 7.84$

Conditioned on Victim



Profile - Combined $\log(\text{LR})$

Suspect A $\log(\text{LR}) = 18.72$

Suspect B $\log(\text{LR}) = 19.45$

Exploring the Capabilities

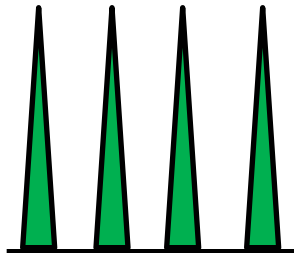
- **Degree of Allele Sharing**
- **Mixture Ratios**
- DNA Quantity

Mixture Data Set

- Mixtures of pristine male and female DNA amplified at a total concentration of 1.0 ng/ μ L using Identifiler (standard conditions).
- Mixture ratios ranged from 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90
- Each sample was amplified twice.

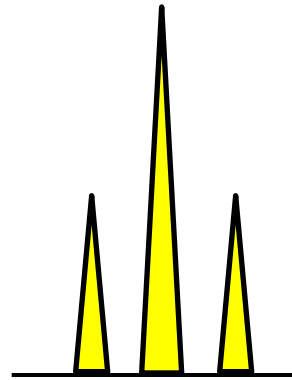
Mixture Data Set

- Three different combinations:



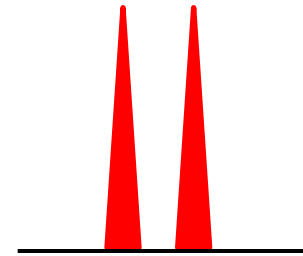
“Low” Sharing

4 alleles – 10 loci
3 alleles – 5 loci
2 alleles – 0 loci
1 allele – 0 loci



“Medium” Sharing

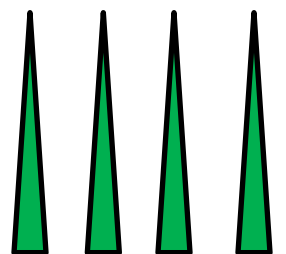
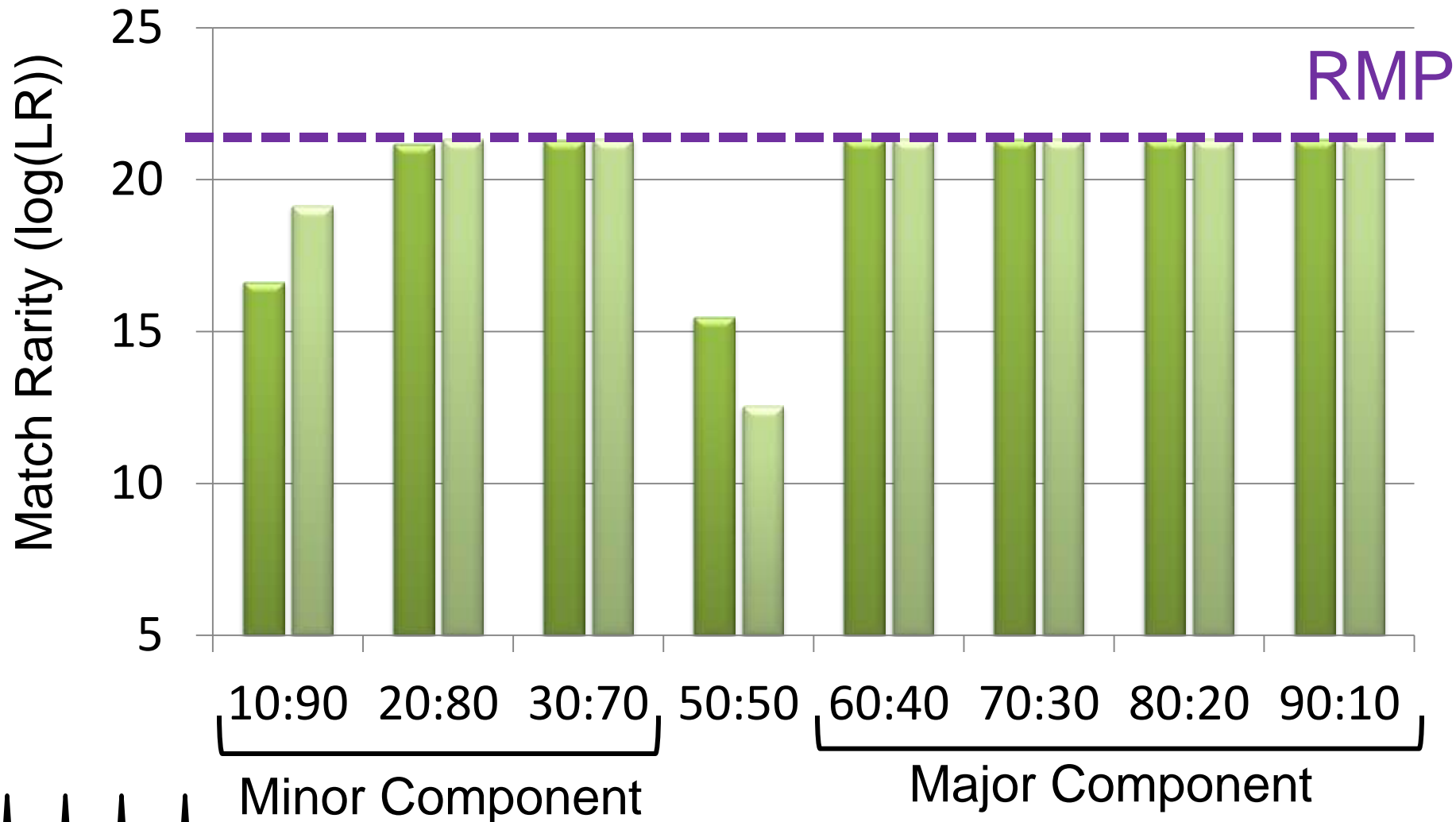
4 alleles – 3 loci
3 alleles – 8 loci
2 alleles – 4 loci
1 allele – 0 loci



“High” Sharing

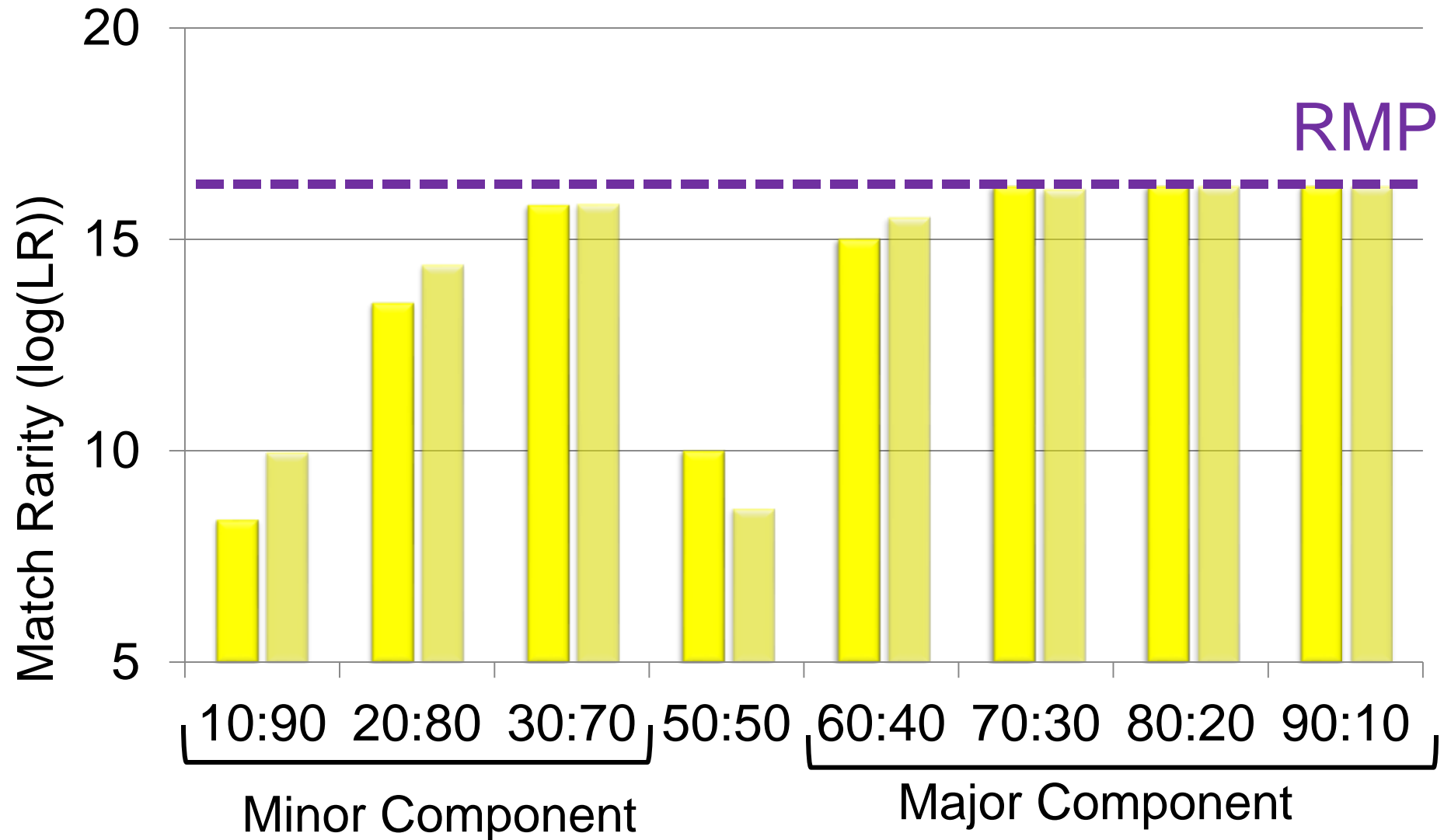
4 alleles – 0 loci
3 alleles – 6 loci
2 alleles – 8 loci
1 allele – 1 loci

Match Score in Duplicate Runs

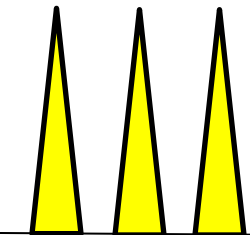


“Easy” for
Deconvolution

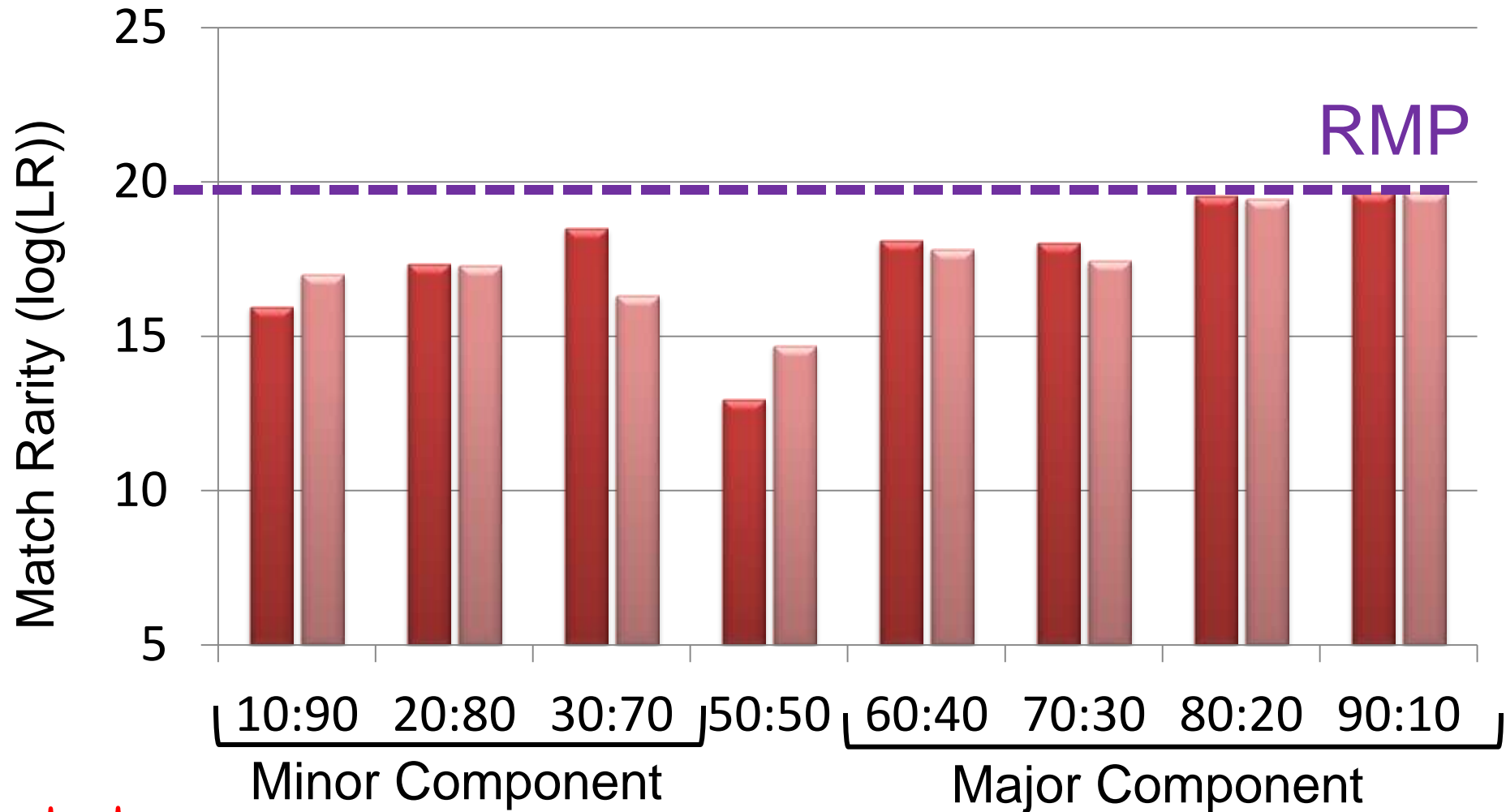
Match Score in Duplicate Runs



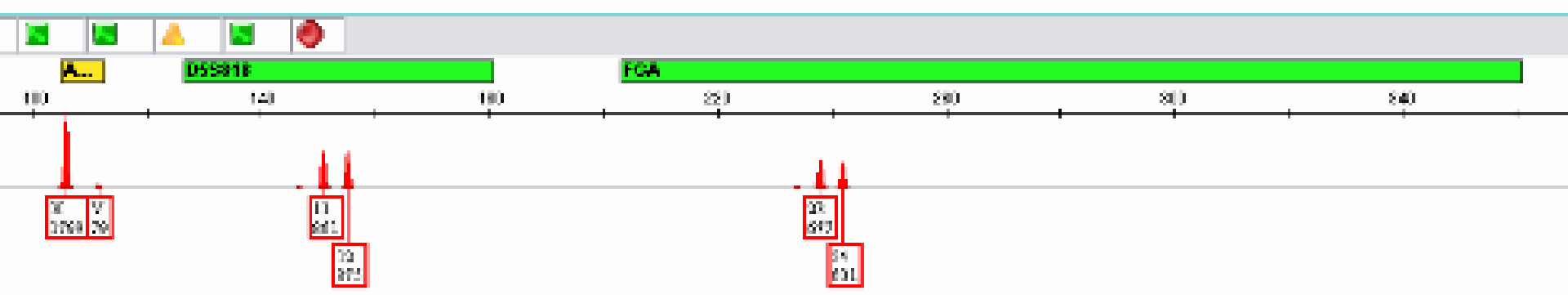
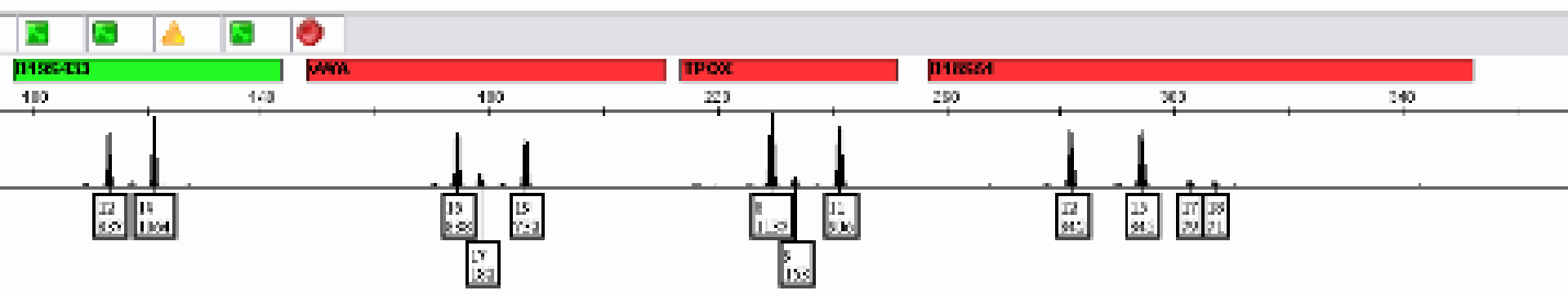
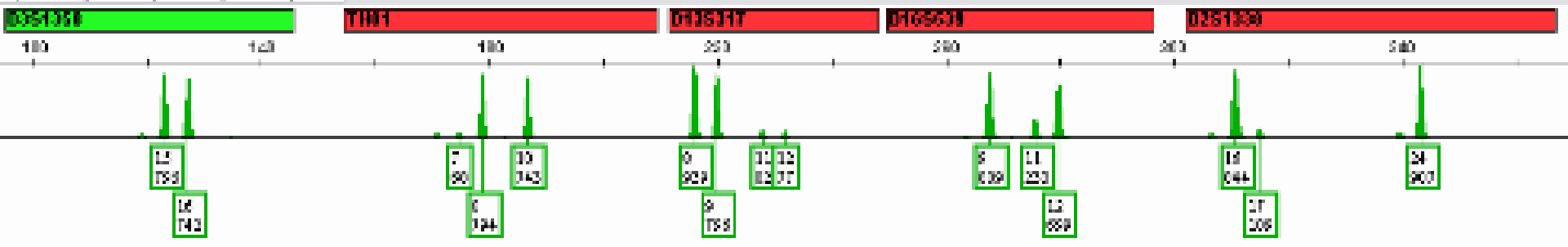
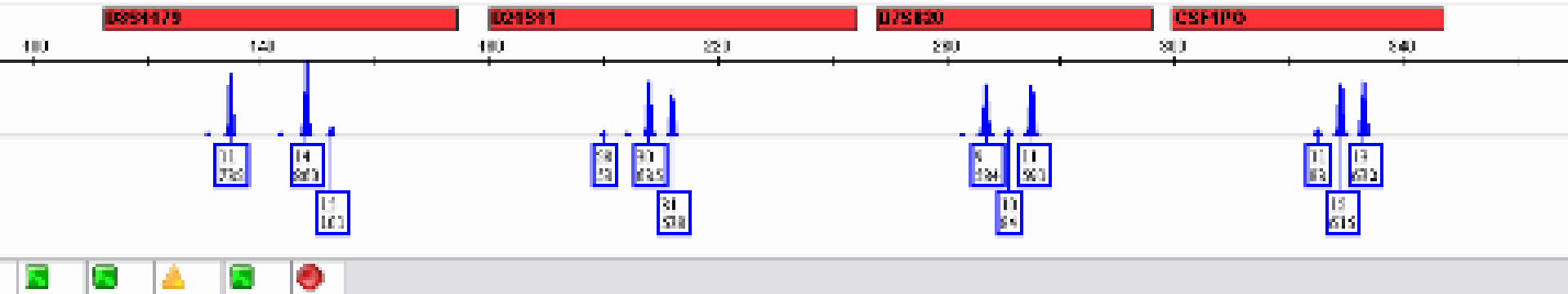
“Challenging” for
Deconvolution

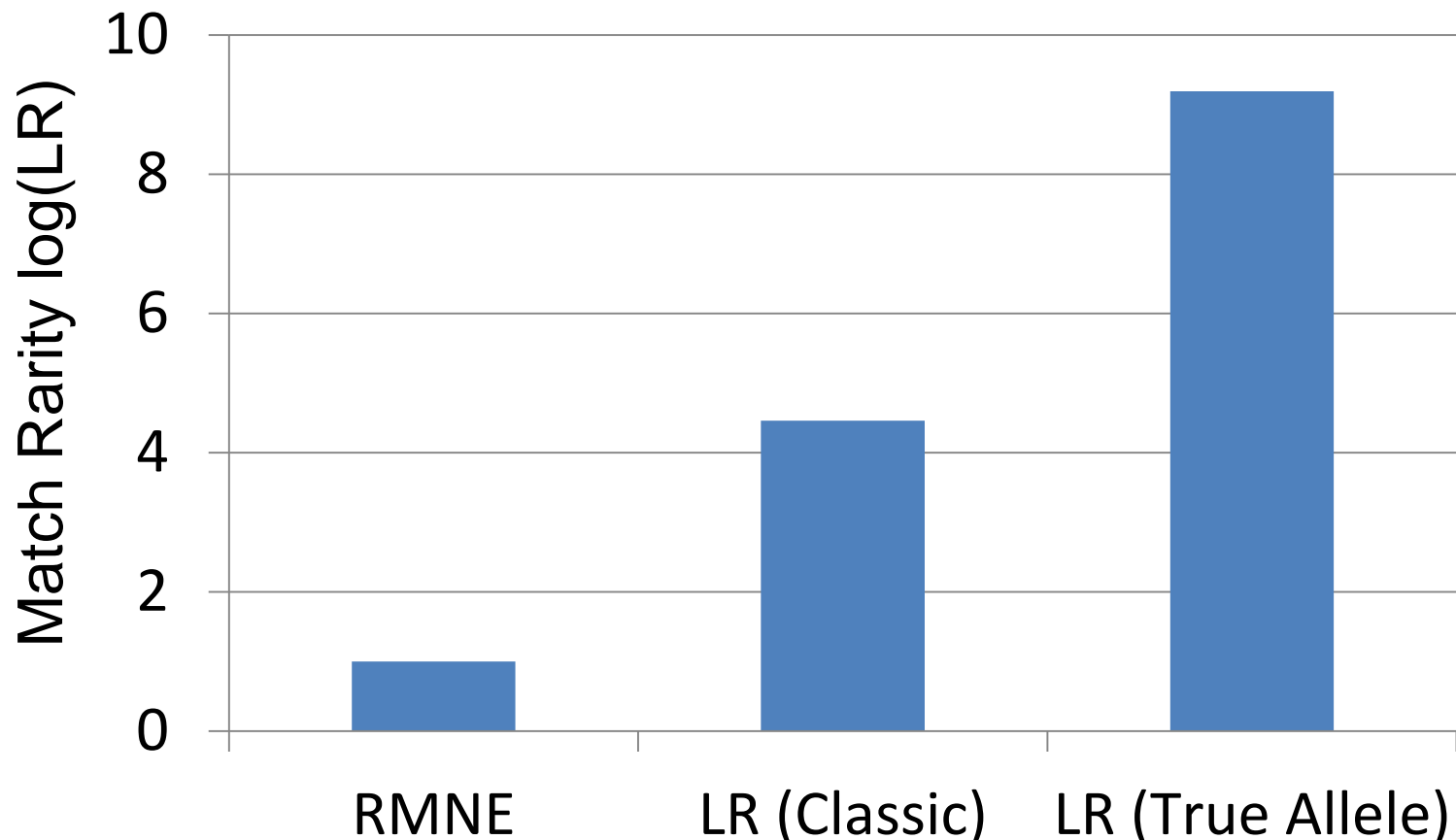
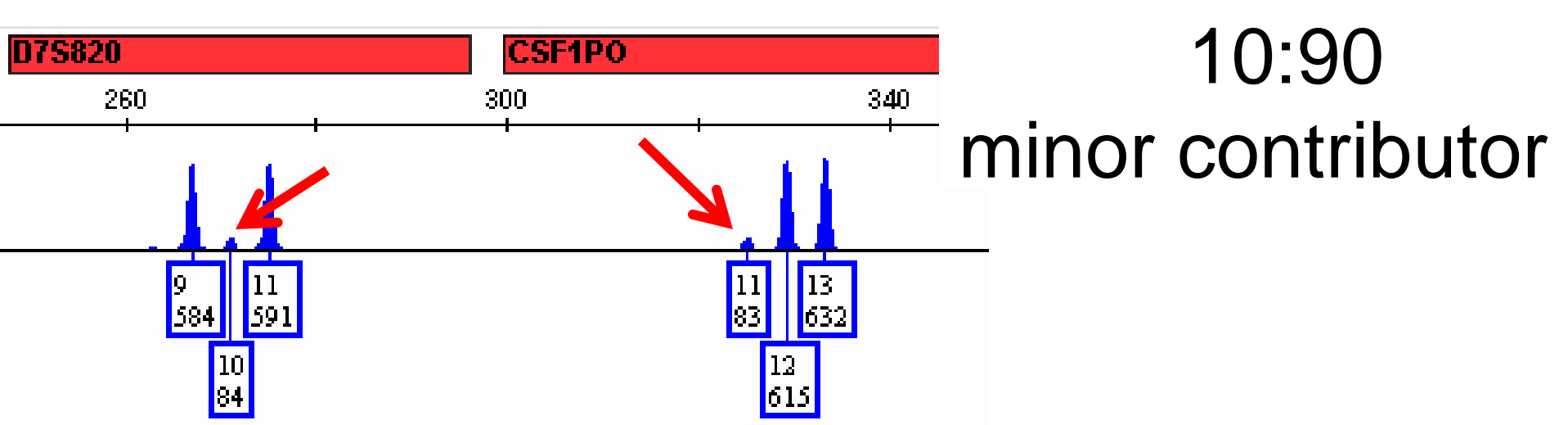


Match Score in Duplicate Runs



“Difficult” for
Deconvolution





Exploring the Capabilities

- Degree of Allele Sharing
- Mixture Ratios
- **DNA Quantity**

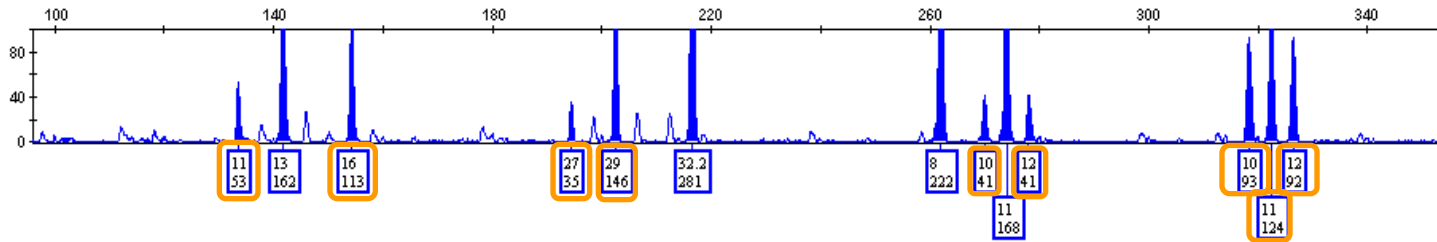
Identifiler

125 pg total DNA

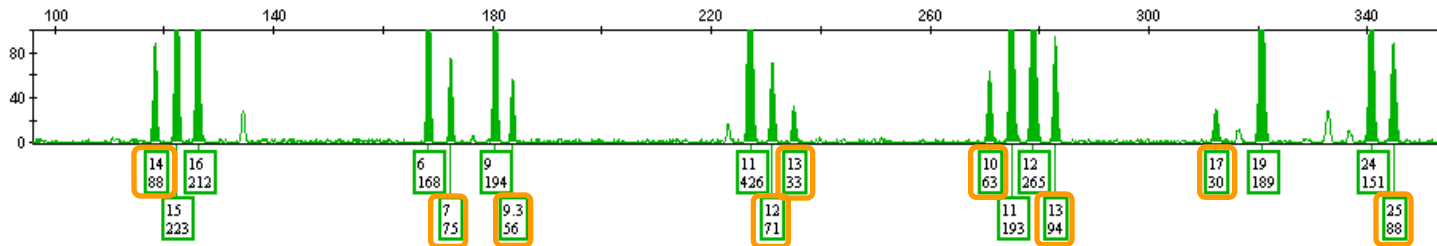
AT = 30 RFU

ST = 150 RFU

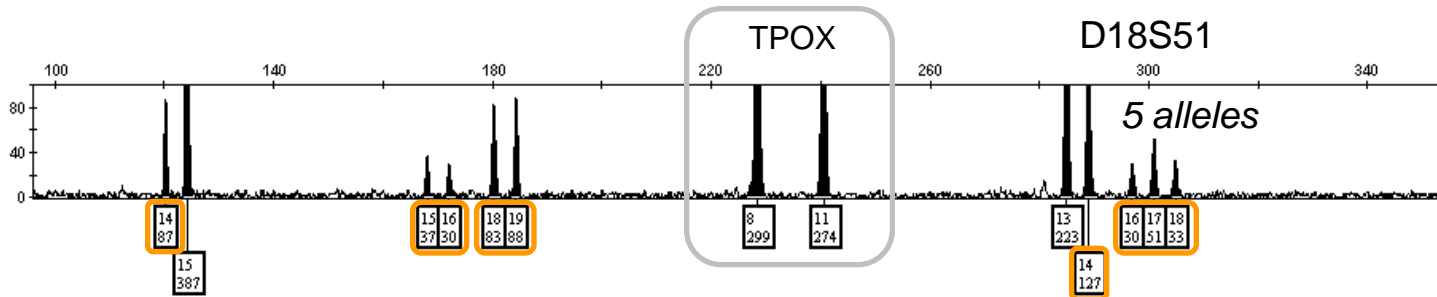
Stutter filter off



Peaks below stochastic threshold



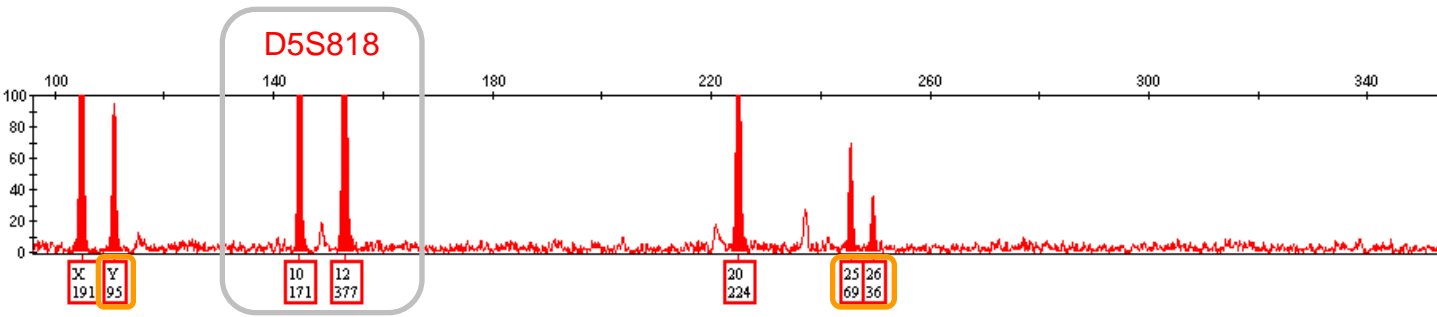
y-axis
zoom to
100 RFU



TPOX

D18S51

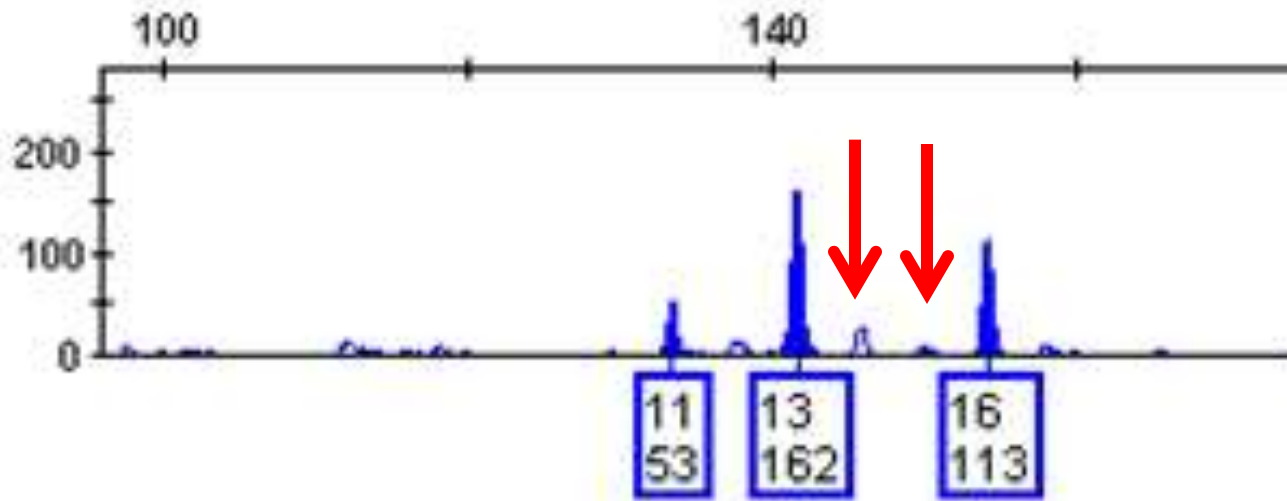
5 alleles



D5S818

D8S1179

“True Genotypes”

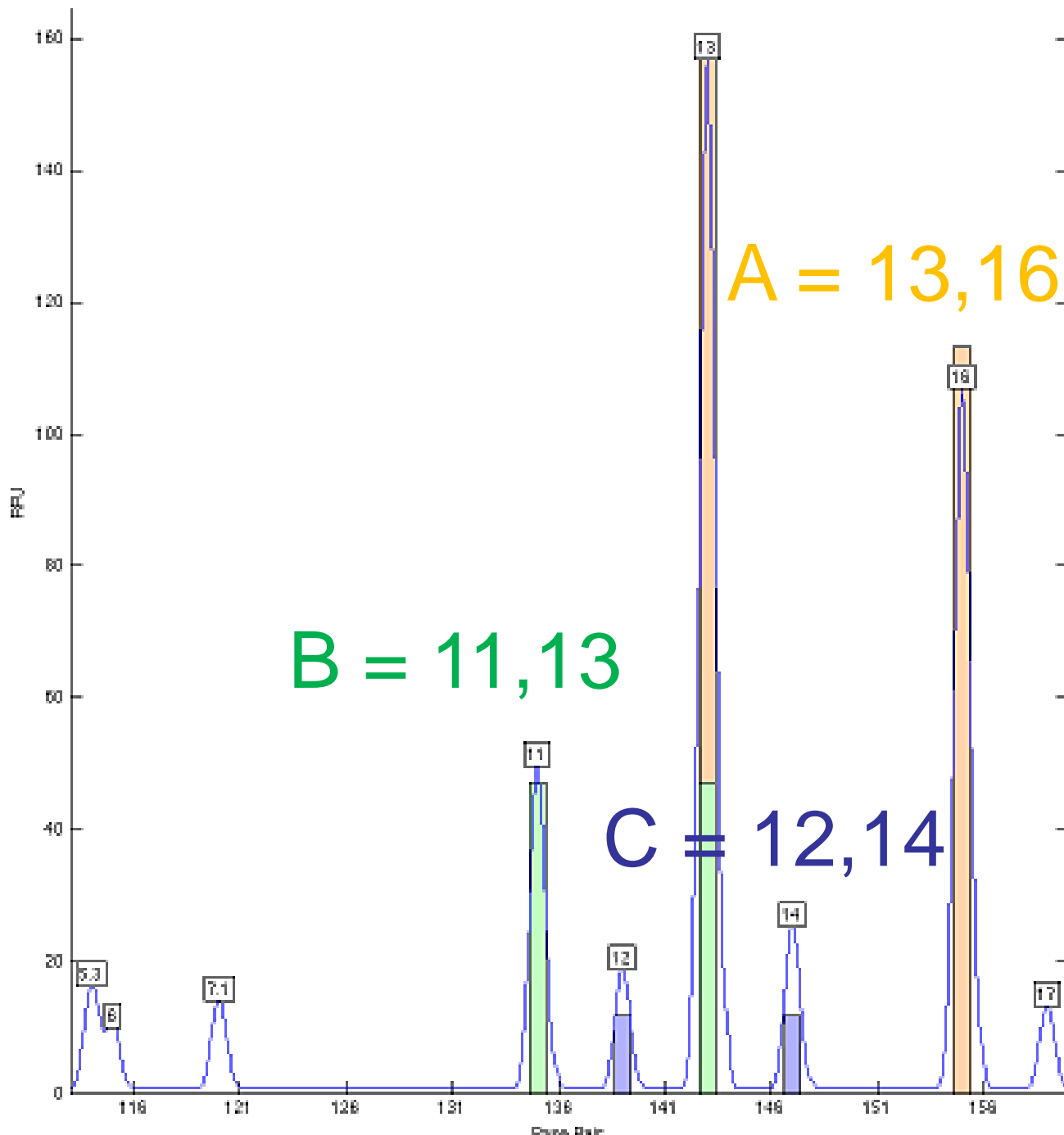


A = 13, 16

B = 11, 13

C = 14, 15

3 person Mixture – No Conditioning
Major Contributor \approx 83 pg input DNA
2 Minor Contributors \approx 21 pg input DNA



“True Genotypes”

A = 13,16

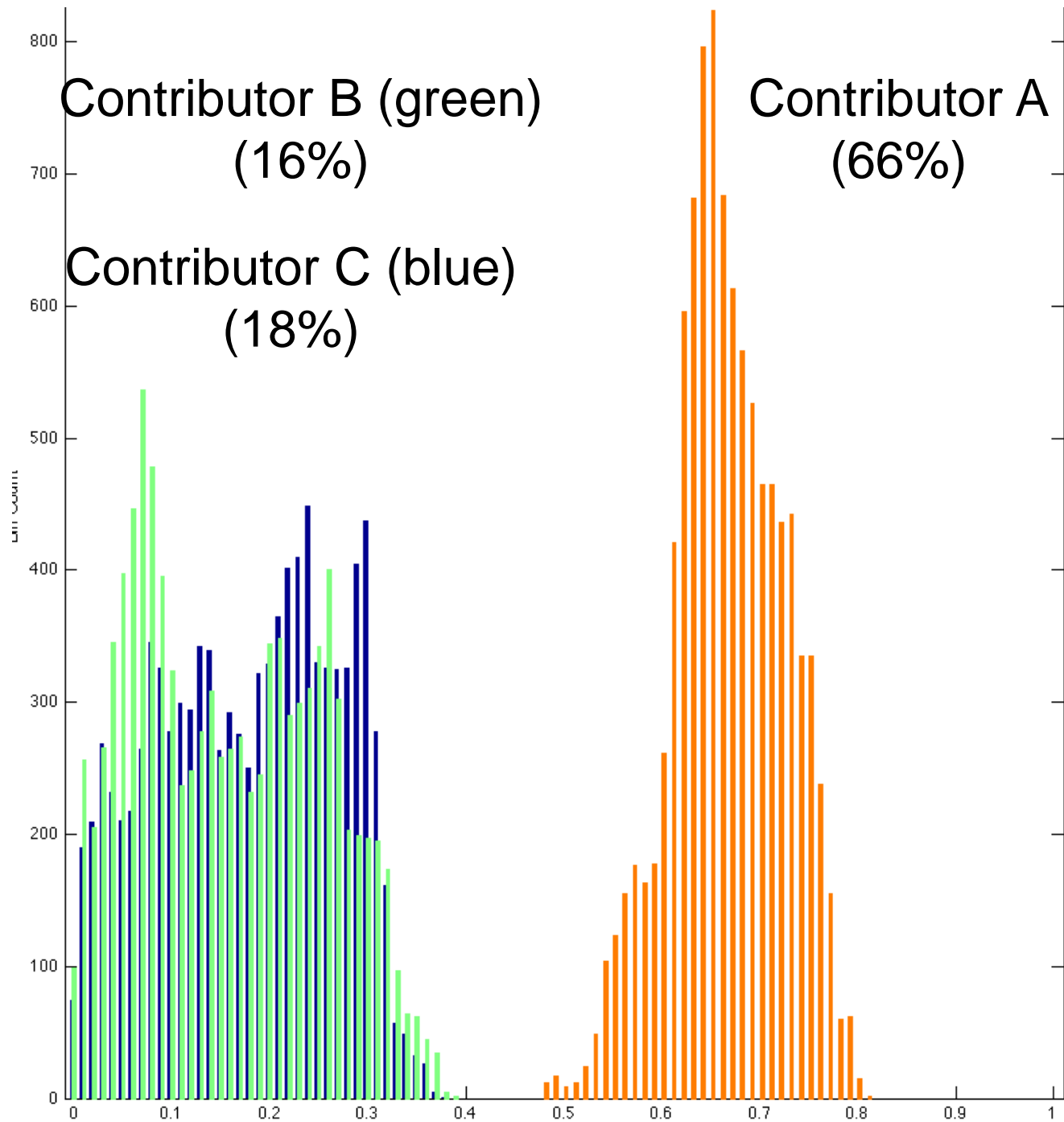
A = 13,16

B = 11,13

B = 11,13

C = 12,14

C = 14,15

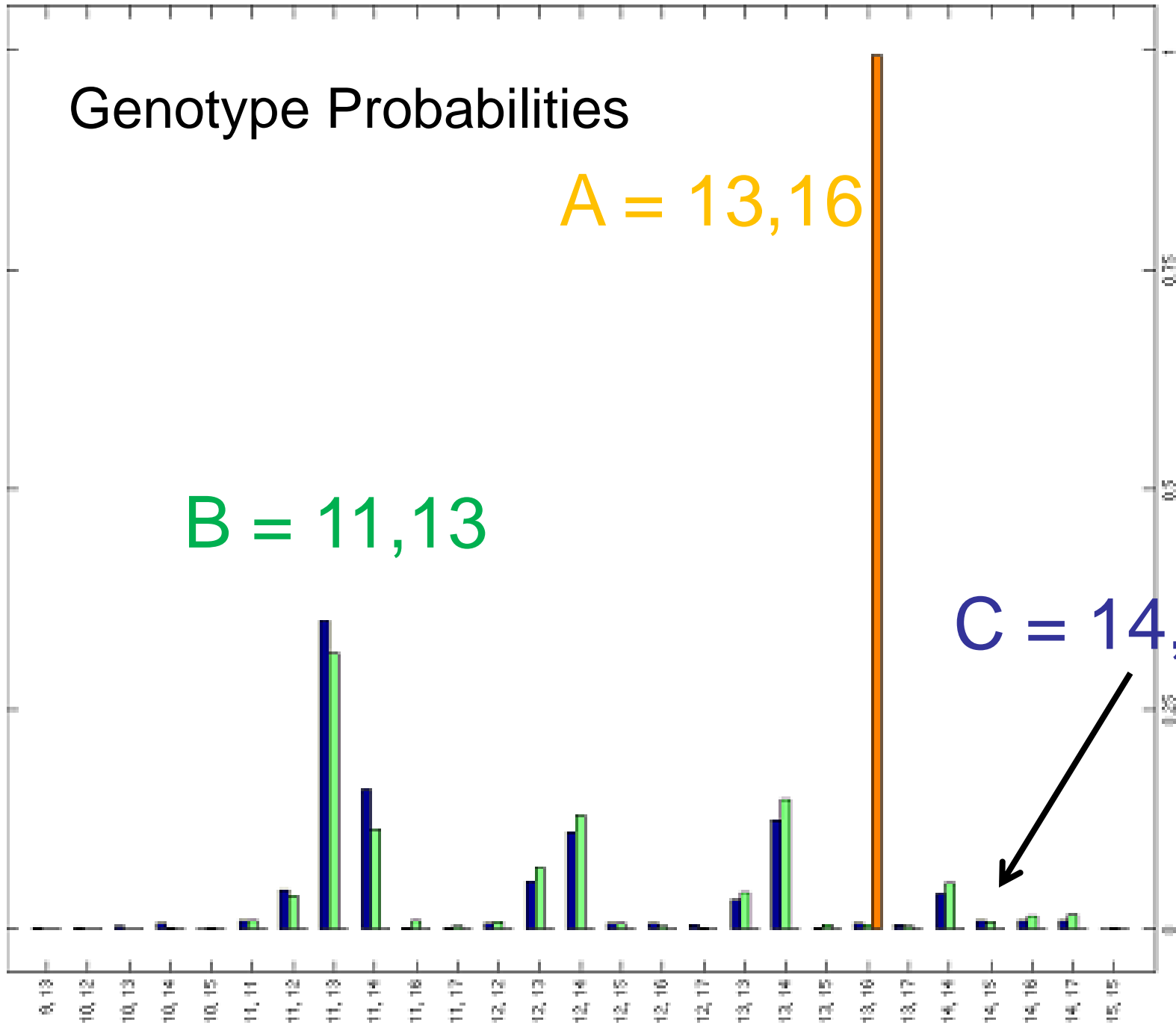


Genotype Probabilities

A = 13,16

B = 11,13

C = 14,15



Results for Contributor A (male)

		Probability	Genotype		H_p	H_d	
Locus	Allele Pair	Likelihood	Frequency	Suspect	Numerator	Denominator	LR
CSF1PO	10, 11	0.572	0.1292			0.07395	
	11, 12	0.306	0.2133	1	0.30563	0.0652	
	10, 12	0.12	0.1547			0.01861	
					0.30563	0.15791	1.935
D13S317	11, 11	1	0.1149	1	1	0.11488	8.704
D8S1179	13, 16	0.998	0.0199	1	0.99786	0.0199	49.668

The match rarity between the evidence and suspect is 1.21 quintillion

Results for Contributor B (female)

Locus	Allele Pair	Probability	Genotype	Suspect	H_p	H_d	LR
		Likelihood	Frequency		Numerator	Denominator	
D8S1179	11, 13	0.073	0.0498	1	0.07338	0.00366	
	11, 14	0.034	0.0271			0.00092	
	13, 14	0.006	0.0996			0.00065	
	12, 14	0.011	0.0606			0.00068	
	12, 13	0.005	0.1115			0.0006	
	11, 12	0.018	0.0303			0.00054	
	14, 14	0.004	0.0271			0.00012	
	13, 13	0.003	0.0916			0.00031	
	14, 16	0.003	0.0108			0.00003	
	14, 15	0.001	0.0379			0.00003	

etc...

9.197

The match rarity between the evidence and suspect is 1.43 million

Results for Contributor C (male)

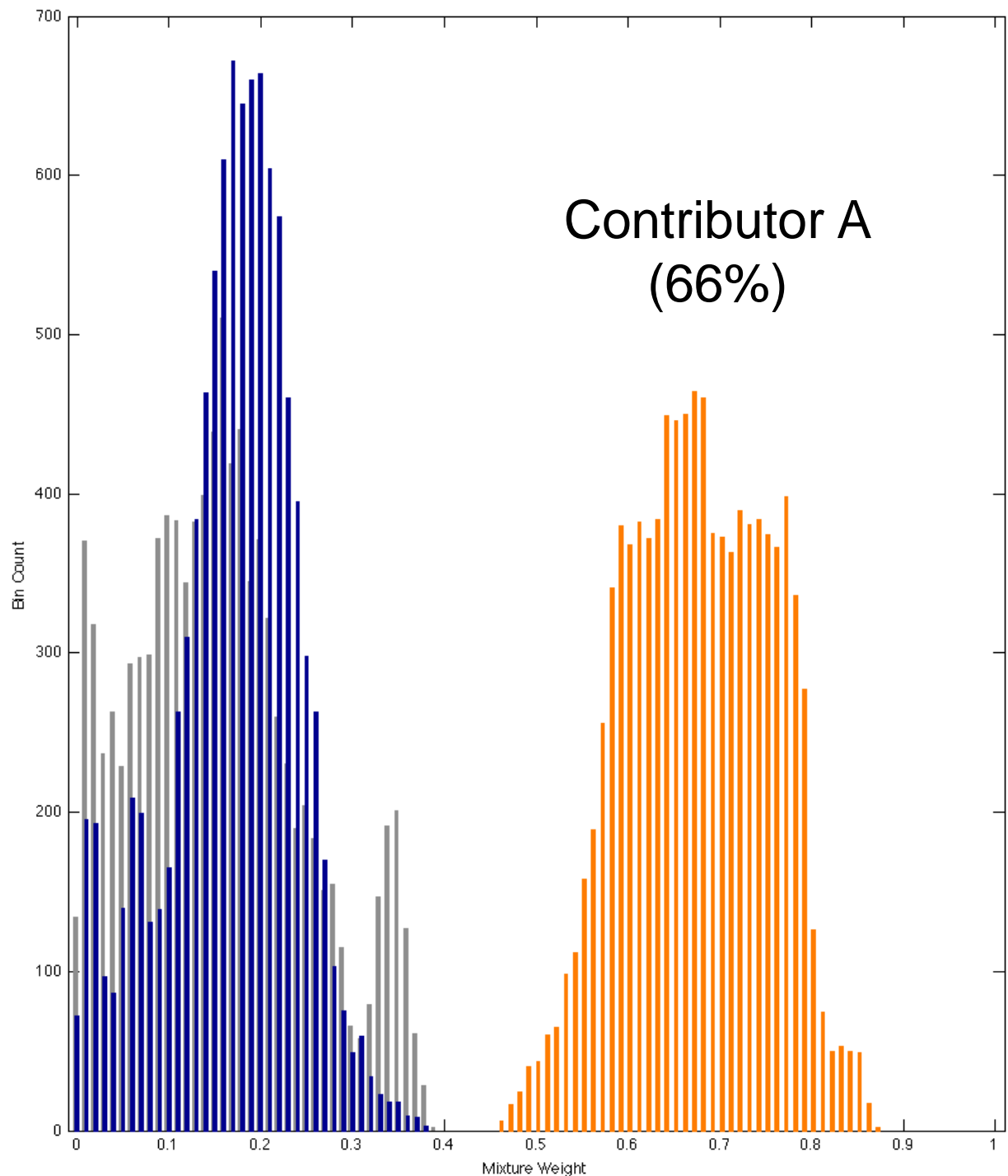
		Probability	Genotype		H _p	H _d	
Locus	Allele Pair	Likelihood	Frequency	Suspect	Numerator	Denominator	LR
D8S1179	11, 13	0.056	0.0498			0.00279	
	13, 14	0.007	0.0996			0.00066	
	12, 14	0.011	0.0606			0.00068	
	11, 14	0.021	0.0271			0.00056	
	12, 13	0.006	0.1115			0.00066	
	14, 14	0.005	0.0271			0.00013	
	etc...	etc...	etc...			etc...	
	14, 15	0.001	0.0379	1	0.00056	0.00002	
	12, 15	0.001	0.0424			0.00003	
	etc...	etc...	etc...			etc...	
	10, 15	0	0.0227			0.00001	
					0.00056	0.00665	0.084

The match rarity between the evidence and suspect is 9.16 thousand

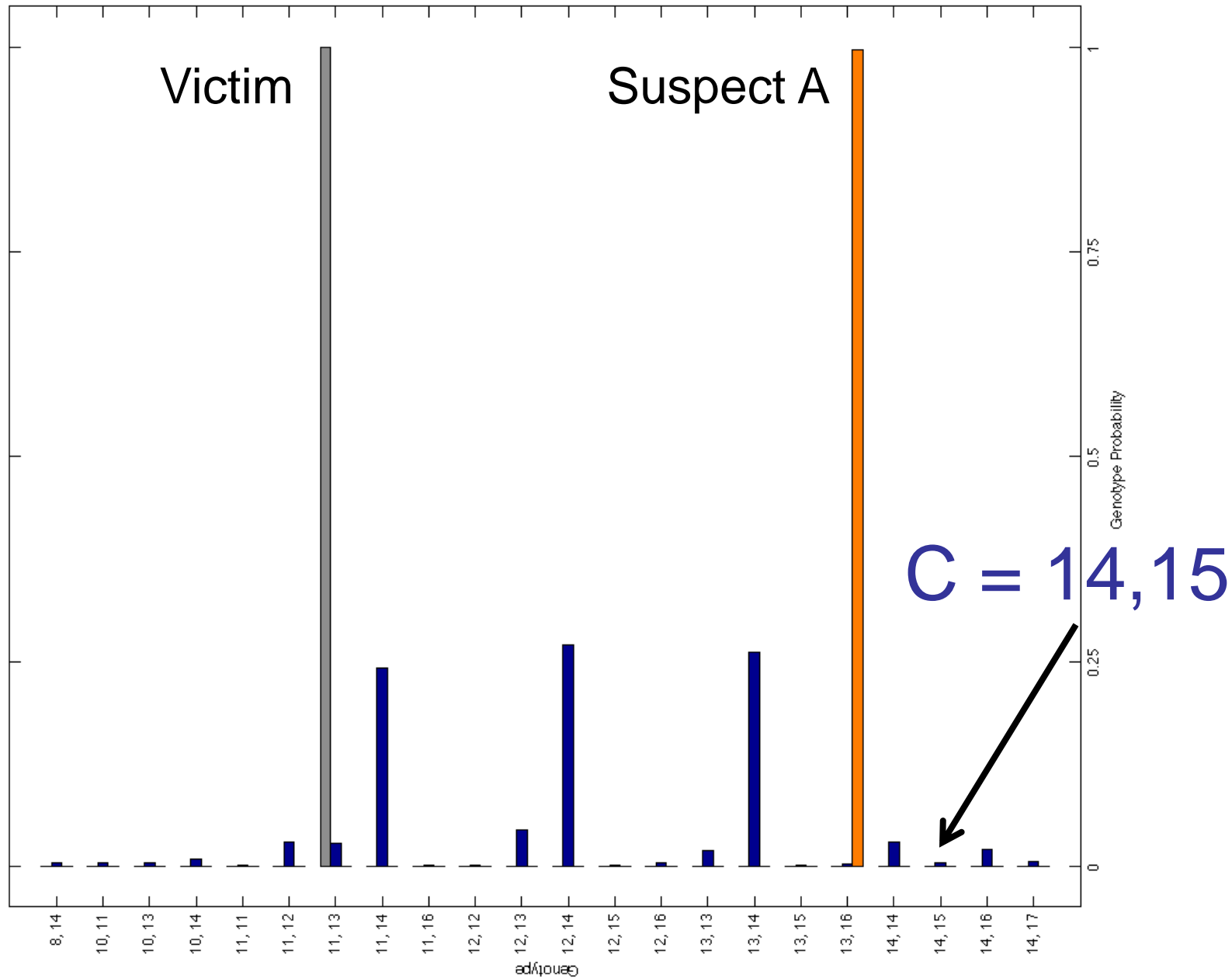
Contributor B (gray)
(16%)

Contributor C (blue)
(18%)

Conditioned on the Victim



The Power of Conditioning



The Power of Conditioning

	LR (no conditioning, 3unk)
Contributor A	1.21 Quintillion
Contributor B (victim)	1.43 Million
Contributor C	9.16 Thousand

	LR (conditioned on victim + 2unk)
Contributor A	1.32 Quintillion
Contributor B (victim)	2.19 Million
Contributor C	59.8 Thousand



Ranged from 1.13 to 800K

Summary

- True Allele utilizes probabilistic genotyping and makes better use of the data than the RMNE approach.
- However, the software is computer intensive. On our 4 processor system, it can take 12-16 hours to run up to four 3-person mixture samples.

Summary

- **Allele Sharing:** Stacking of alleles due to sharing creates more uncertainty.
- **Mixture Ratio:** With “distance” between the two contributors, there is greater certainty. Generally, True Allele performs better than RMNE and the classic LR with low level contributors.

Summary

- **DNA Quantity:** Generally, with high DNA signal, replicates runs on True Allele are very reproducible.
- However, with low DNA signal, higher levels of uncertainty are observed (as expected).
- There is a need to determine an appropriate threshold for an inclusion $\log(\text{LR})$.

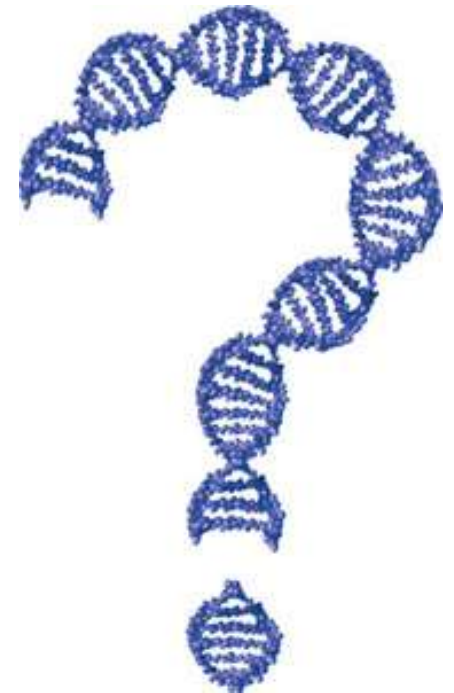
Thank You!

Our team publications and presentations are available at:
<http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm>

Funding from the **National
Institute of Justice (NIJ)**
through NIST Office of Law
Enforcement Standards



Questions?



john.butler@nist.gov

301-975-4049

michael.coble@nist.gov

301-975-4330