Forensic DNA Mixture Interpretation

Probabilistic Genotyping

MAFS Workshop

Milwaukee, WI
September 25, 2012

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

michael.coble@nist.gov

Is there a way forward?

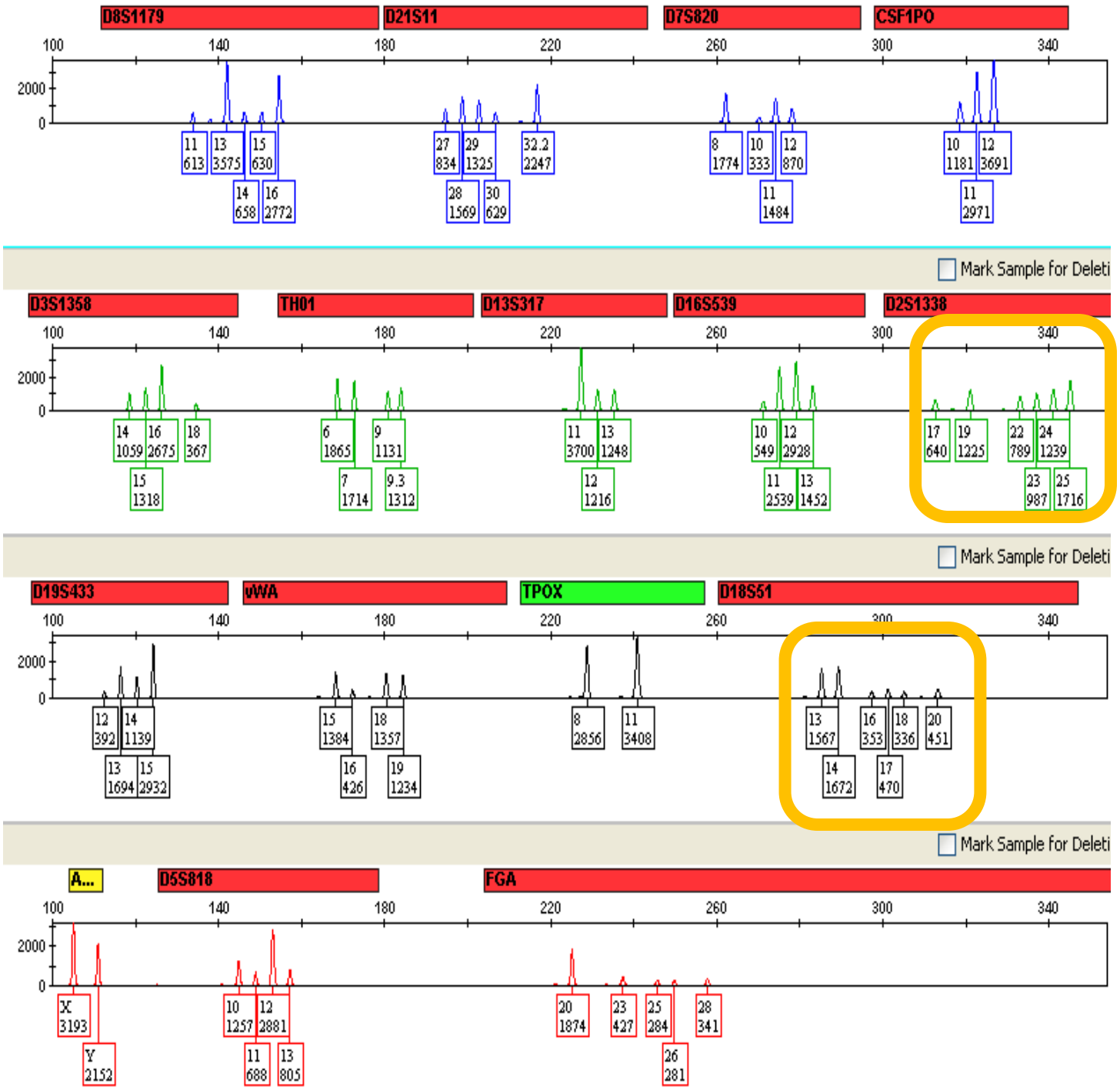
Three Questions

- What were the last words of Julius Caesar before he died?
- Et tu, Brute? Then fall Caesar!
- What is the capital of Bangladesh?
- Dhaka

Three Questions

- How many people are in this mixture?

All alleles are above ST



Do you have any uncertainty in your answer?

Whatever way uncertainty is approached, probability is the *only* sound way to think about it.

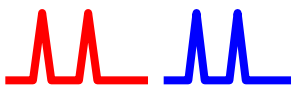
-Dennis Lindley

Two-Person Mixtures

Observed profile

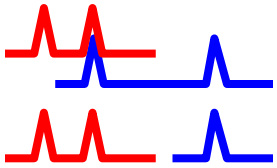
A **B**

14 total combinations



4 alleles

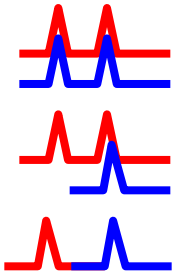
All heterozygotes and non-overlapping alleles



3 alleles

Heterozygote + heterozygote, one overlapping allele

Heterozygote + homozygote, no overlapping alleles

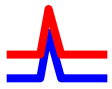


2 alleles

Heterozygote + heterozygote, two overlapping alleles

Heterozygote + homozygote, one overlapping allele

Homozygote + homozygote, no overlapping alleles



1 allele

Homozygote + homozygote, overlapping allele

3-Person Mixtures

Observed profile

150 total combinations

6 alleles

All heterozygotes and non-overlapping alleles

5 alleles

Two heterozygotes and one homozygote

Three heterozygotes, one overlapping allele

4 alleles

Six combinations of heterozygotes, homozygotes and overlapping alleles

3 alleles

Eight combinations of heterozygotes, homozygotes, and overlapping alleles

2 alleles

Five combinations of heterozygotes, homozygotes, and overlapping alleles

1 allele

All homozygotes, overlapping allele

4-Person Mixtures

Observed profile

8 alleles

All heterozygotes and non-overlapping alleles

MANY combinations

7 alleles

Several combinations of heterozygotes, homozygotes, and overlapping alleles

6 alleles

Many combinations

5 alleles

Many combinations

4 alleles

Many combinations

3 alleles

Many combinations

2 alleles

Many combinations

1 allele

All homozygotes, overlapping allele

Four-Person Mixture Studies Summary

>70% of 4-person mixtures would NOT be recognized as 4-person mixtures based on allele count

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28; Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6; Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1; Perez et al., Croat Med J. 2011; 52:314-26

“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.

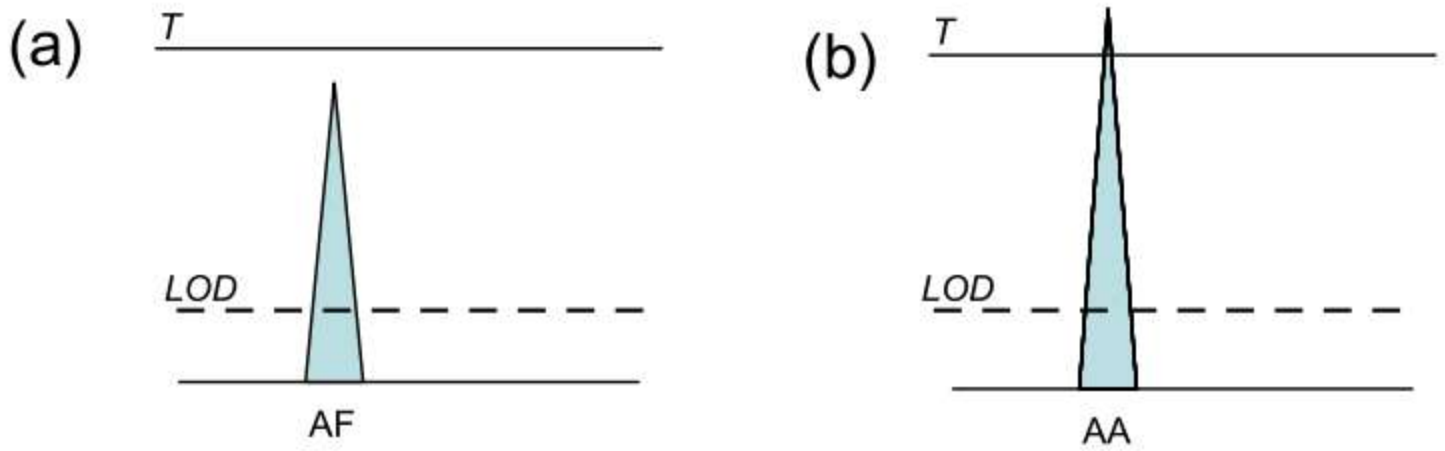
JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2010, Vol. 55, No. 1
doi: 10.1111/j.1556-4029.2009.01257.x
Available online at: interscience.wiley.com

Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman 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.

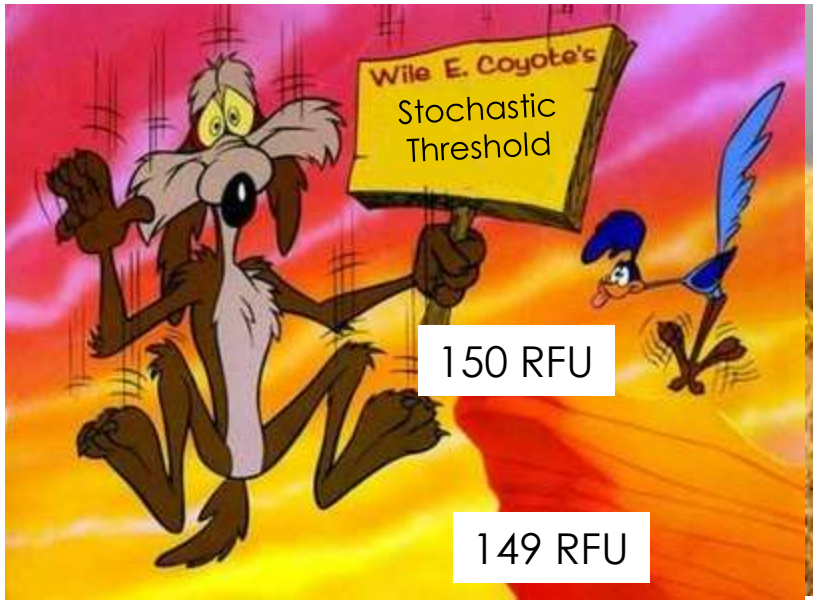
“Falling off the Cliff Effect”

- 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).



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

Falling off the Cliff vs. Gradual Decline



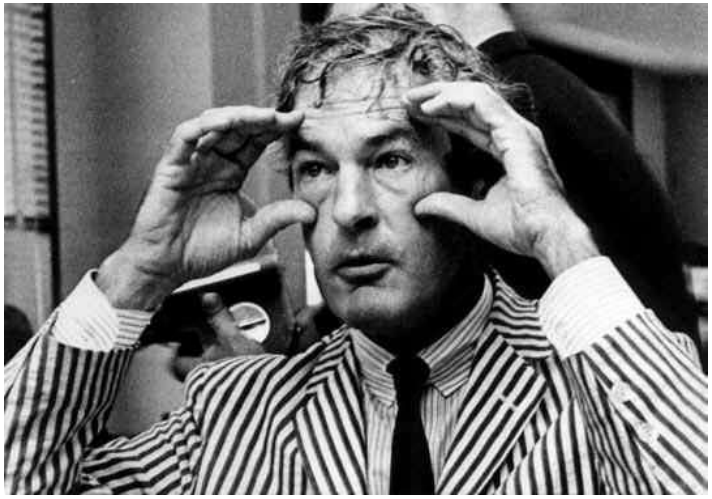
http://blog.sironaconsulting.com/.a/6a00d8341c761a53ef011168cc5ff3970c-pi

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

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

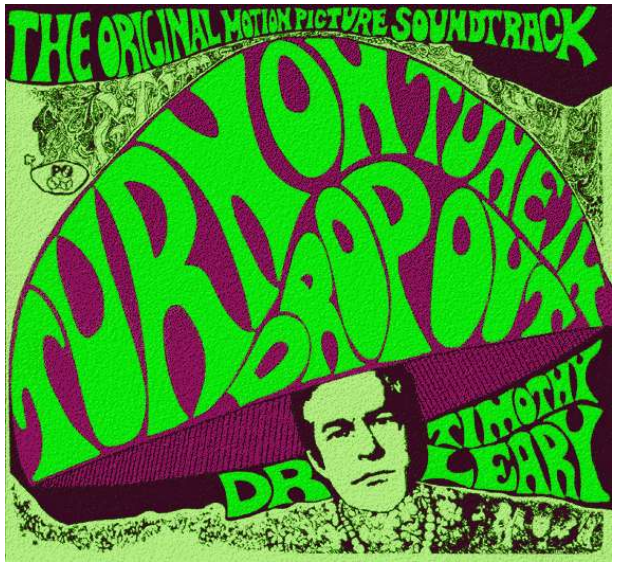
- "The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of ***probabilistic models to circumvent the requirement for a threshold*** and to safeguard the legitimate interests of defendants."

Psychedelic Mixtures



Turn On…

Tune In…



(Talk about) Drop Out

Next Issue of FSI-Genetics

Forensic Science International: Genetics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Editorial

Focus issue—Analysis and biostatistical interpretation of complex and low template DNA samples

Article in press…

Contents lists available at SciVerse ScienceDirect

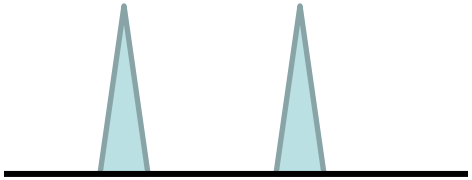
Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

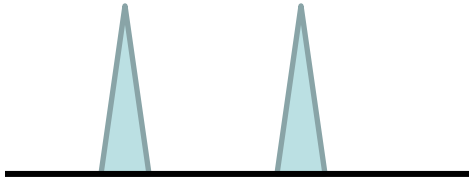
DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods

P. Gill [a,b,*], L. Gusmão [c], H. Haned [d], W.R. Mayr [e], N. Morling [f], W. Parson [g], L. Prieto [h], M. Prinz [i], H. Schneider [j], P.M. Schneider [k], B.S. Weir [l]

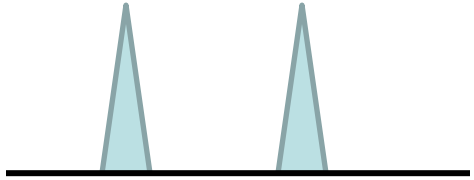
Suspect



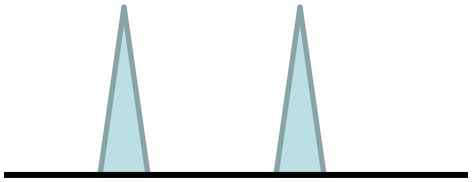
Suspect



Suspect



Evidence



Evidence

Evidence

$$LR = \frac{1}{2pq}$$

$$LR = \frac{0}{2pq}$$

$$LR = \frac{?}{2pq}$$

“2p”

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

Haned *et al.*

Forensic Science International: Genetics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Exploratory data analysis for the interpretation of low template DNA mixtures

H. Haned [a,*], K. Slooten [a,b], P. Gill [c,d]

[a] Netherlands Forensic Institute, Department of Human Biological traces, The Hague, The Netherlands
[b] VU University Amsterdam, Amsterdam, The Netherlands
[c] Norwegian institute of Public Health, Oslo, Norway
[d] University of Oslo, Norway

Mitchell *et al.*

Forensic Science International: Genetics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in

Adele A. Mitchell *, Jeannie Tamariz, Kathleen O'Connell, Nubia Ducasse, Zoran Budimlija, Mechthild Prinz, Theresa Caragine

Department of Forensic Biology, Office of Chief Medical Examiner of The City of New York, 421 E 26th Street, New York, NY 10016, United States

The Drop-out Model

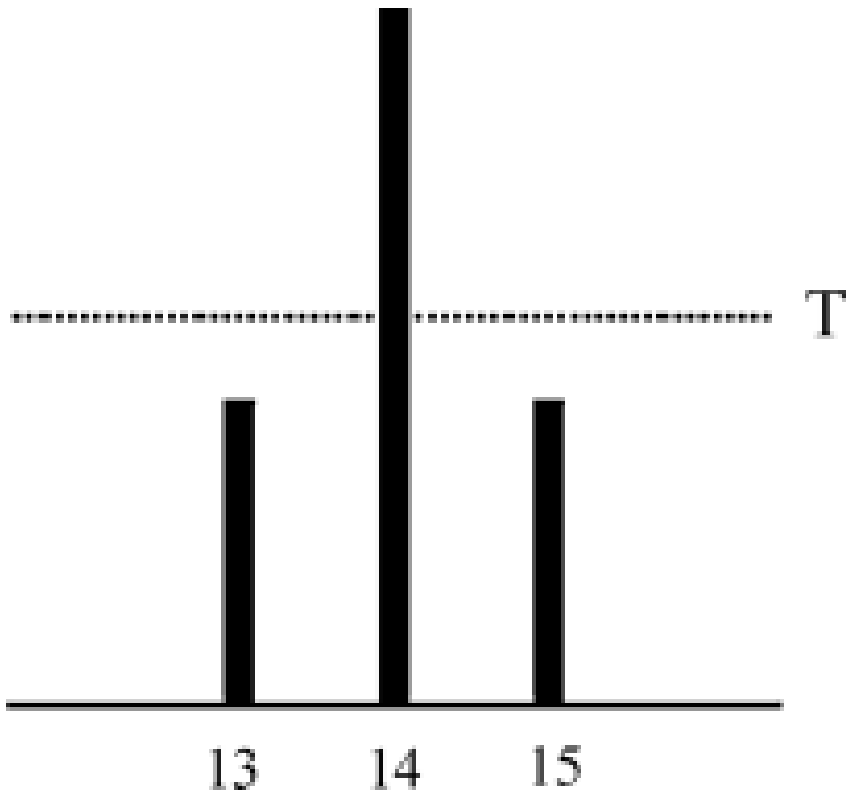
The interpretation of low level DNA mixtures

Hannah Kelly [a,*], Jo-Anne Bright [a], James Curran [b], John Buckleton [a]

[a] ESR, PB 92021 Auckland, New Zealand
[b] Department of Statistics, University of Auckland, PB 92019 Auckland, New Zealand

FSI - Genetics 6 (2012) 191–197

First – Convert Peaks to Alleles



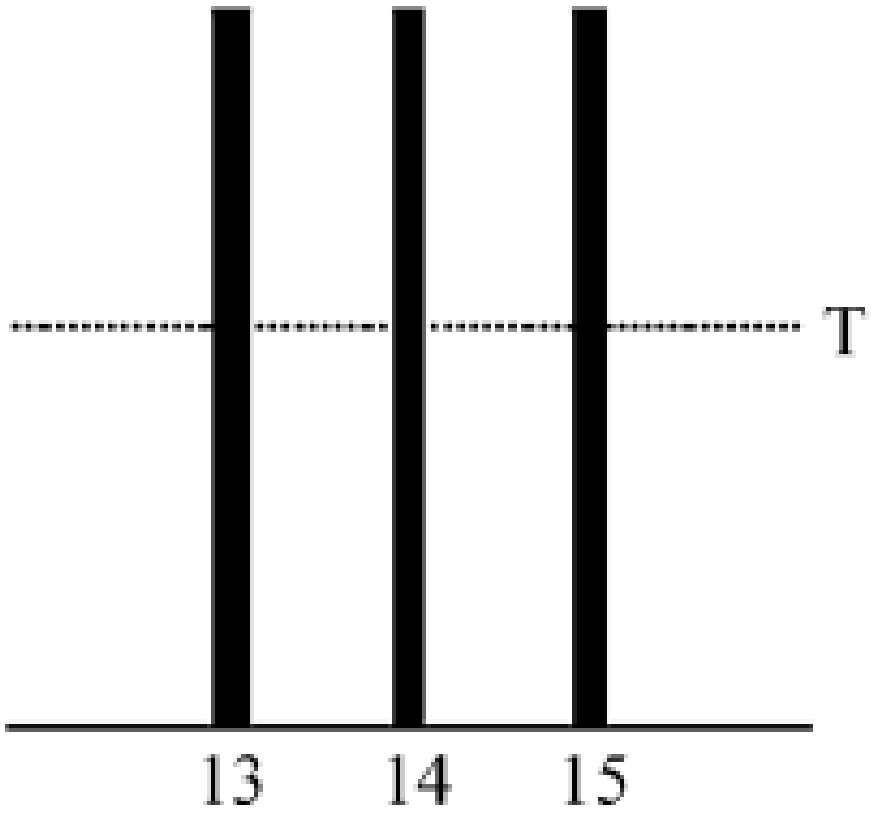
Assume 2 Contributors

3 peaks – 4 alleles

Allelic Vector
13
14
14
15

13,14,14,15

Ambiguity in Determining Vectors



Assume 2 Contributors

Allelic Vectors

13, 13, 14, 15

13, 14, 14, 15

13, 14, 15, 15

3 possibilities

Permutations

- The number of permutations is the number of ways that the alleles can be arranged as pairs.

Permutations

- An easier way to compute using factorials.

$$\binom{n}{m_1, m_2, \ldots m_l} = \frac{n!}{m_1! m_2!, \ldots, m_l!}$$

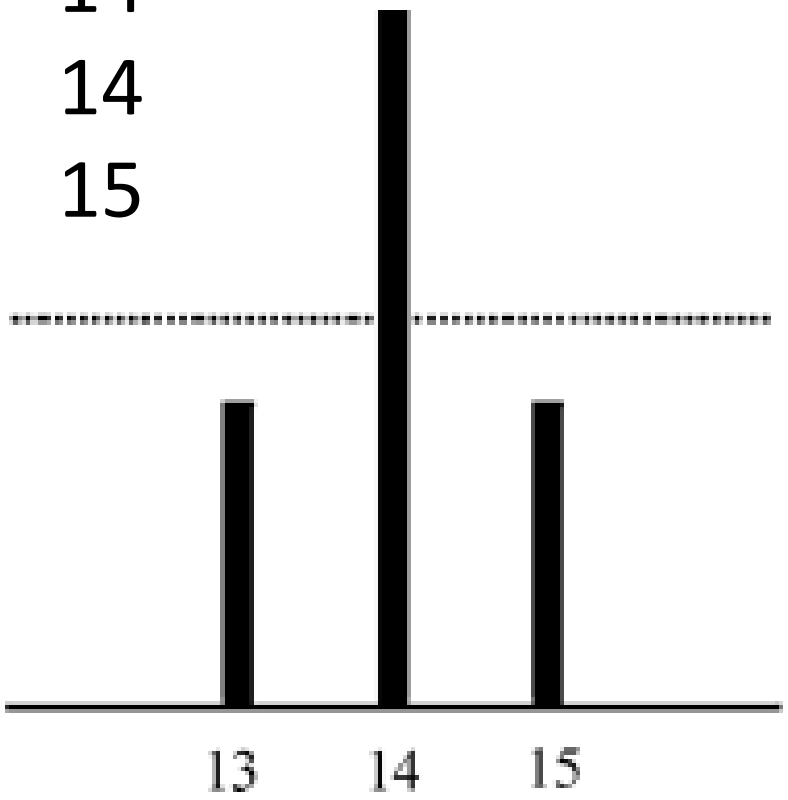
n = total number of alleles at the locus.

m = number of times each allele is seen.

Determine the Permutations for this example

Allelic Vectors

13
14
14
15



$$= \frac{4!}{1!2!1!}$$

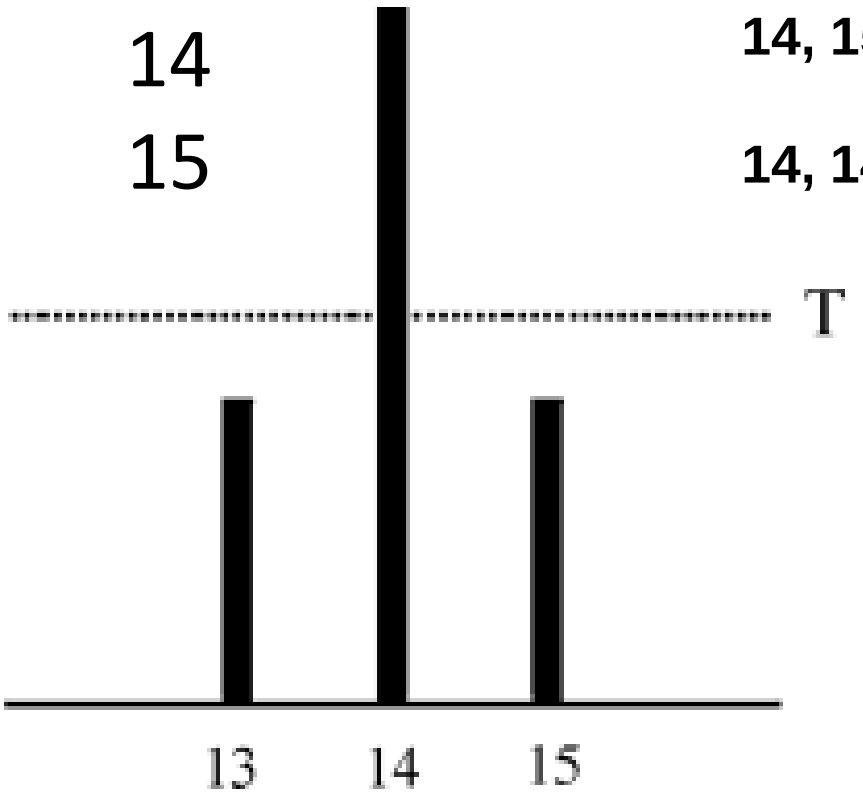
$$= \frac{4x3x2x1}{1x2x1}$$

$$= 12$$

Let's Prove It!

Allelic Vectors

13
14
14
15



13, 14 and 14, 15 = 2ab x 2bc = $4ab^2c$

13, 15 and 14, 14 = 2ac x b^2 = $2ab^2c$

14, 15 and 13, 14 = 2bc x 2ab = $4ab^2c$

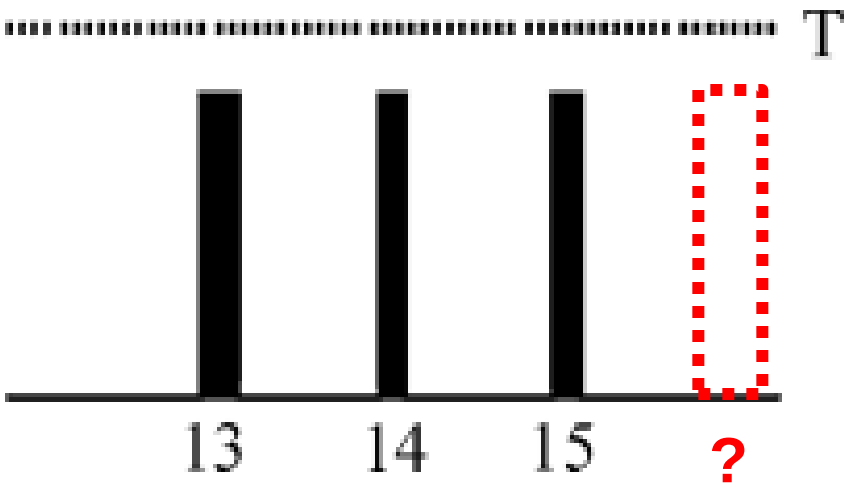
14, 14 and 13, 15 = b^2 x 2bc = $2ab^2c$

= $12ab^2c$

= 12

Assign Allele Designations

- Use “F” as a placeholder to consider alleles that may have dropout.



Assume 2 Contributors

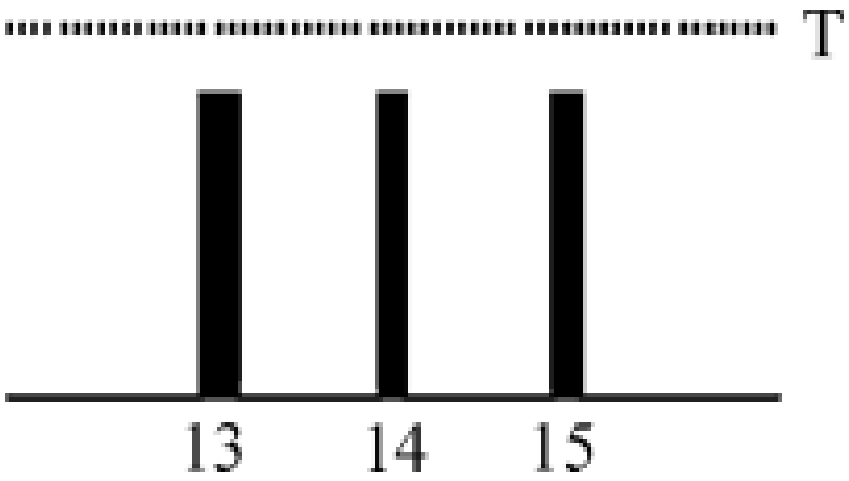
3 peaks – 3 alleles

<u>Allelic Vector</u>

13,14,15,F

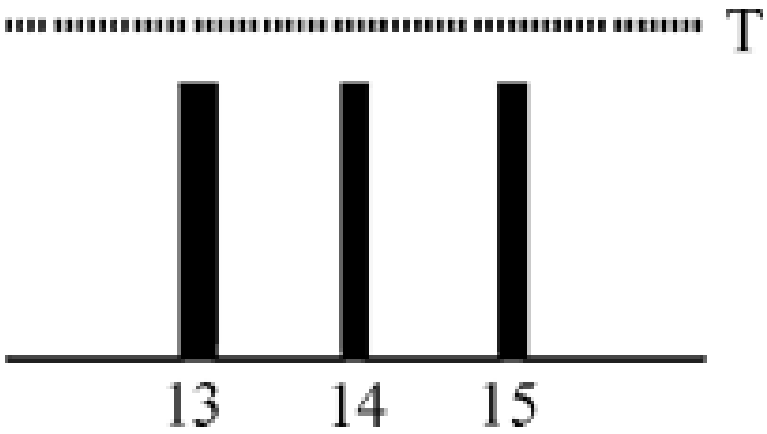
Assign Probability using the F-model

- Calculate the number of permutations using “F” as a placeholder and then drop it from the equation.



Assign Probability using the F-model

$$\Pr(13,14,15,F|X) = \frac{4!}{1!1!1!1!}\Pr(13,14,15,F|X)$$



$$= 24\Pr(13,14,15|X)$$

Apply the Sampling Formula (Balding and Nichols 1994)

$$\frac{x\,\theta + (1-\theta)p_A}{1 + (n-1)\,\theta}$$

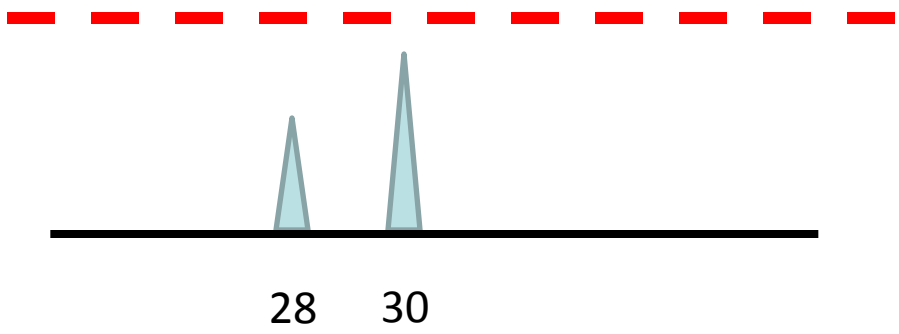
x = value calculated from the F-model.

p_a = frequency of the “a” allele.

Θ = coancestry coefficient (F_{ST}).

n = number of alleles.

A Worked Example



POI = 28, 30

2 peaks – 4 alleles

D21
Assume 2 contributors
Allele 28 = 107 RFU
Allele 30 = 198 RFU
ST = 200 RFU

Allelic Vector
28,30,F,F

Permutations and Probability

$$\Pr(28,30,F,F \mid 28,30) =$$

$$\frac{4!}{1!1!2!} \Pr(28,30,F,F \mid 28,20)$$

$$= 12\Pr(28,30 \mid 28,30)$$

Apply the Sampling Formula (Balding and Nichols 1994)

$$\Pr(A|X) = \frac{x\theta + (1-\theta)\,p_a}{1+(n-1)\theta}$$

Pr(E|Hp) =1

Pr(E|Hd) =12Pr(28,30|28,30)

$$\frac{12(\theta(1-\theta)\,p_{28})(\theta+(1-\theta)\,p_{30})}{(1+\theta)(1+2\theta)}$$

LR = 1.86

Kelly *et al.*

- Other models including the “Q” method and the Unconstrained Combinatorial “UC” method (no peak height info).

- The UC method overestimates the LR and is not appropriate. The “Q” model performs better than the “F” model, but is more mathematically intense…

The “Q” Model for D21 (28,30)

Allelic vector (28,30)

$\Pr(E|Hp) = 1$

$$4\Pr(28,28,28,30|28,30) + 6\Pr(28,28,30,30|28,30) + 4\Pr(28,30,30,30|28,30) + 12\Pr(28,28,30,Q|28,30) + 12\Pr(28,30,30,Q|28,30) + 12\Pr(28,30,Q,Q|28,30)$$

$$\Pr(E|Hd) = 2\Pr(28,30|28,30) \times \begin{bmatrix} 6 - 6\Pr(28|28,28,30,30) - 6\Pr(30|28,28,30,30) + 2\Pr(28,28|28,28,30,30) \\ +2\Pr(30,30|28,28,30,30) \\ +3\Pr(28,30|28,28,30,30) \end{bmatrix}$$

$$\frac{2(\theta(1-\theta)p_{28})(\theta+(1-\theta)p_{30})}{(1+\theta)(1+2\theta)} \times \left[6 - \frac{6(2\theta+(1-\theta)p_{28})}{(1+3\theta)} - \frac{6(2\theta+(1-\theta)p_{30})}{(1+3\theta)} + \frac{2(2\theta+(1-\theta)p_{28})(3\theta+(1-\theta)p_{28})}{(1+3\theta)(1+4\theta)} + \frac{2(2\theta+(1-\theta)p_{30})(3\theta(1-\theta)p_{30})}{(1+3\theta)(1+4\theta)} + \frac{3(2\theta+(1-\theta)p_{28})(2\theta+(1-\theta)p_{30})}{(1+3\theta)(1+4\theta)} \right]$$

LR with Pr(Drop-out)

Forensic Science International: Genetics 4 (2009) 1–10

Contents lists available at ScienceDirect

Forensic Science International: Genetics

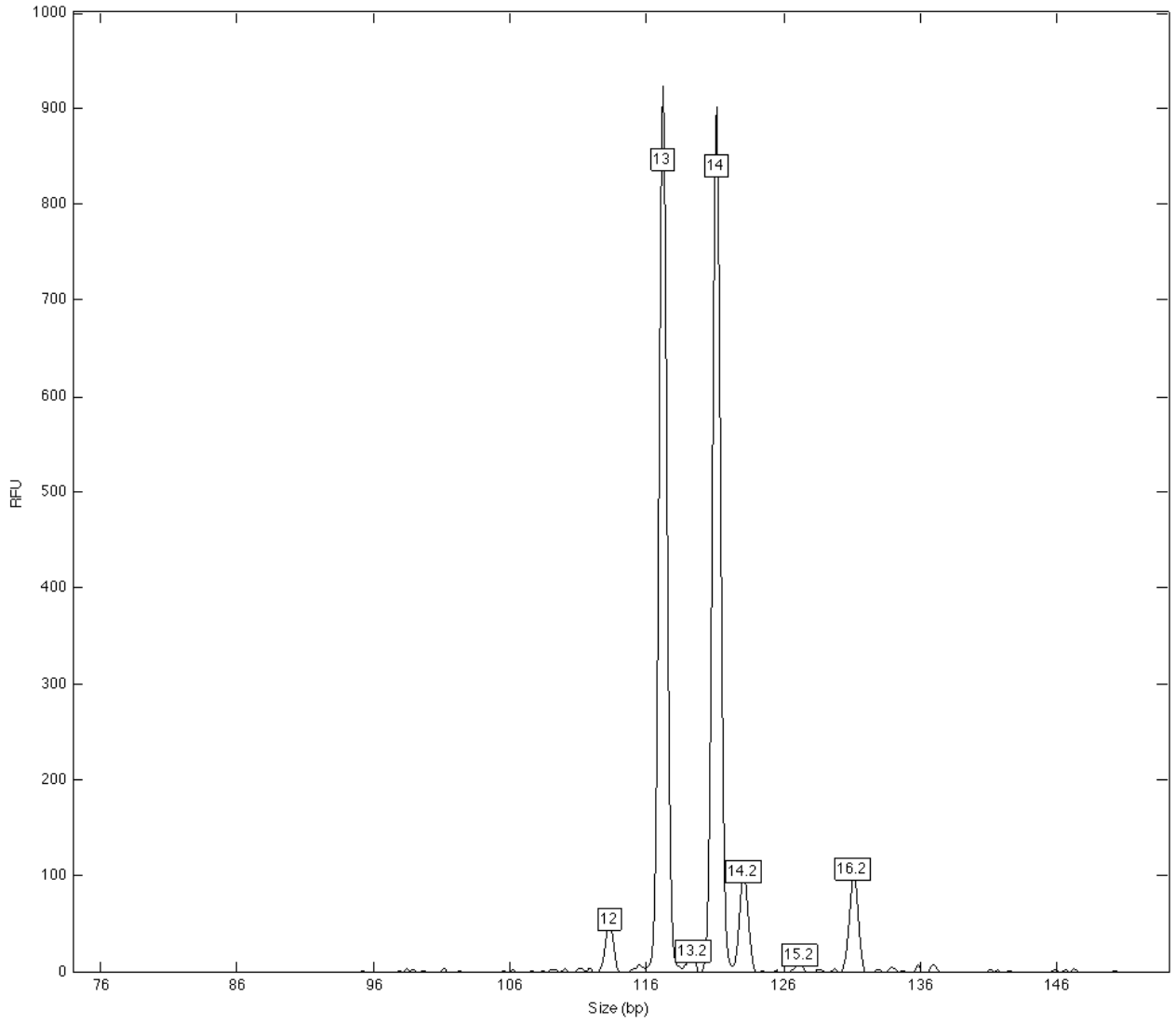
journal homepage: www.elsevier.com/locate/fsig

Interpreting low template DNA profiles

David J. Balding [a,*], John Buckleton [b]

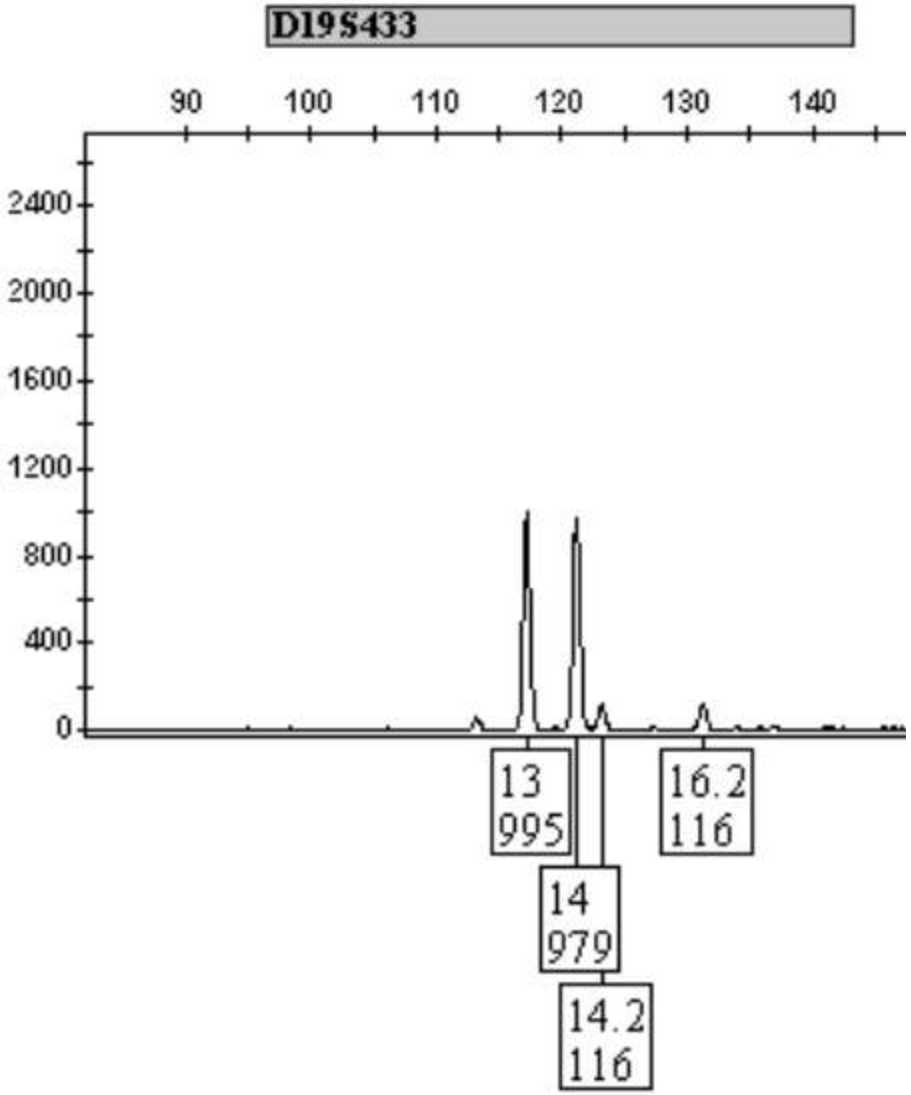
[a] Department of Epidemiology and Public Health, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK
[b] ESR Private Bag 92021, Auckland, New Zealand

3 person mixture – 1 major, 2 minor



D19S433

3 Person Mixture

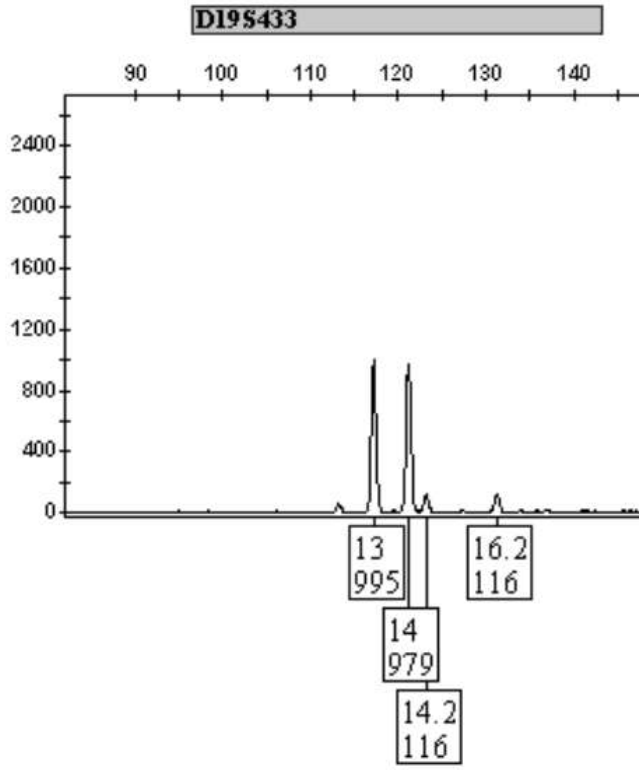


V = 13, 14

CP = 13, 14.2

S = **15**, 16.2

$$\frac{P(E \mid H_1)}{P(E \mid H_2)}$$



V = 13, 14

CP = 13, 14.2

S = **15**, 16.2

Pr(Drop-out) = 10%

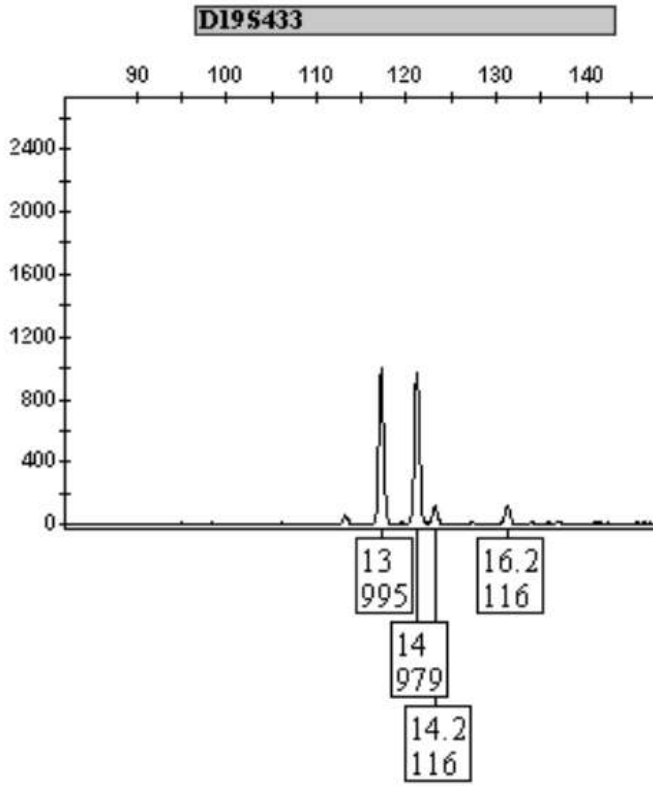
Pr(Drop-in) = 1%

$P(E \mid H_1)$ = Pr(No Drop-out at 16.2) Pr(Drop-out at 15) Pr(No Drop-in)

= 0.90 0.10 0.99

= 0.0891

3 Person Mixture



V = 13, 14

CP = 13, 14.2

S = 15, 16.2

$$\frac{P(E \mid H_1)}{P(E \mid H_2)} \quad 0.0891$$

Keith Inman, Norah Rudin and Kirk Lohmueller have modified the Balding program to incorporate your own data for estimating Pr(Drop-out).

JOURNAL OF FORENSIC SCIENCES

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

PAPER

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. Duceman,[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

Monte Carlo

What is a Markov Chain?

"A mathematical system that undergoes transitions from one state to another, between a finite or countable number of possible states. It is a random process usually characterized as memoryless: the next state depends only on the current state and not on the sequence of events that preceded it."

Andrey Markov

http://en.wikipedia.org/wiki/File:AAMarkov.jpg

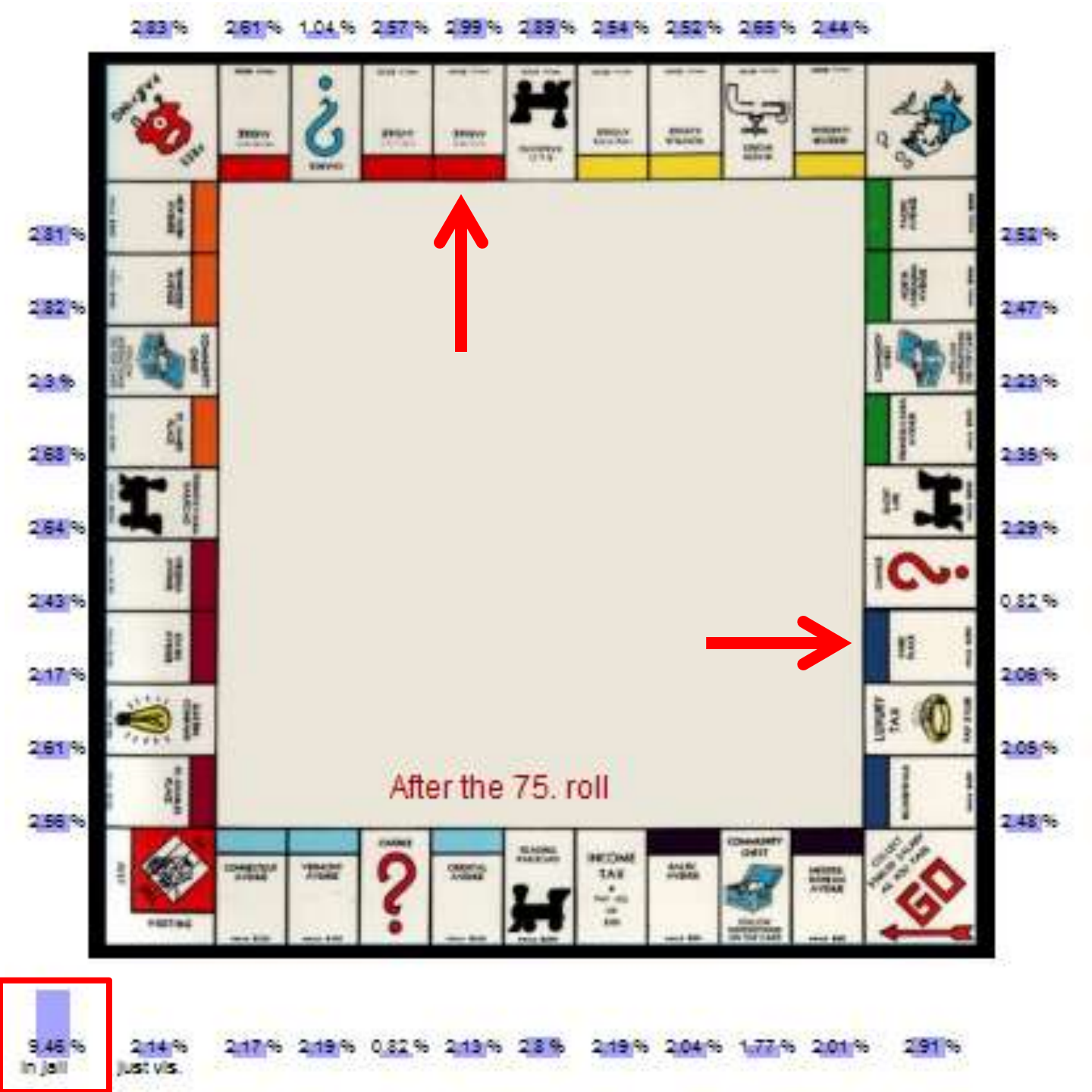
http://en.wikipedia.org/wiki/Markov_chain

Is Blackjack a Markov Chain?

Monopoly is a Markov Chain

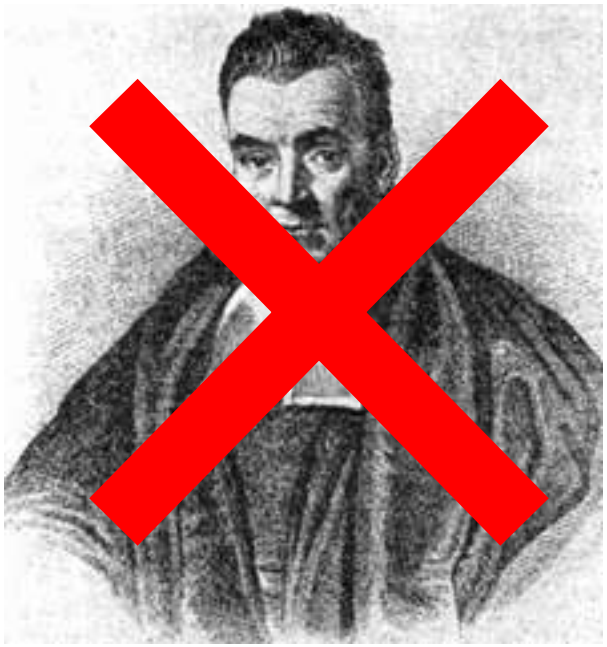
Monopoly simulation

- http://www.bewersdorff-online.de/amonopoly/monopoly_m.htm



Higher Prob. of being in jail

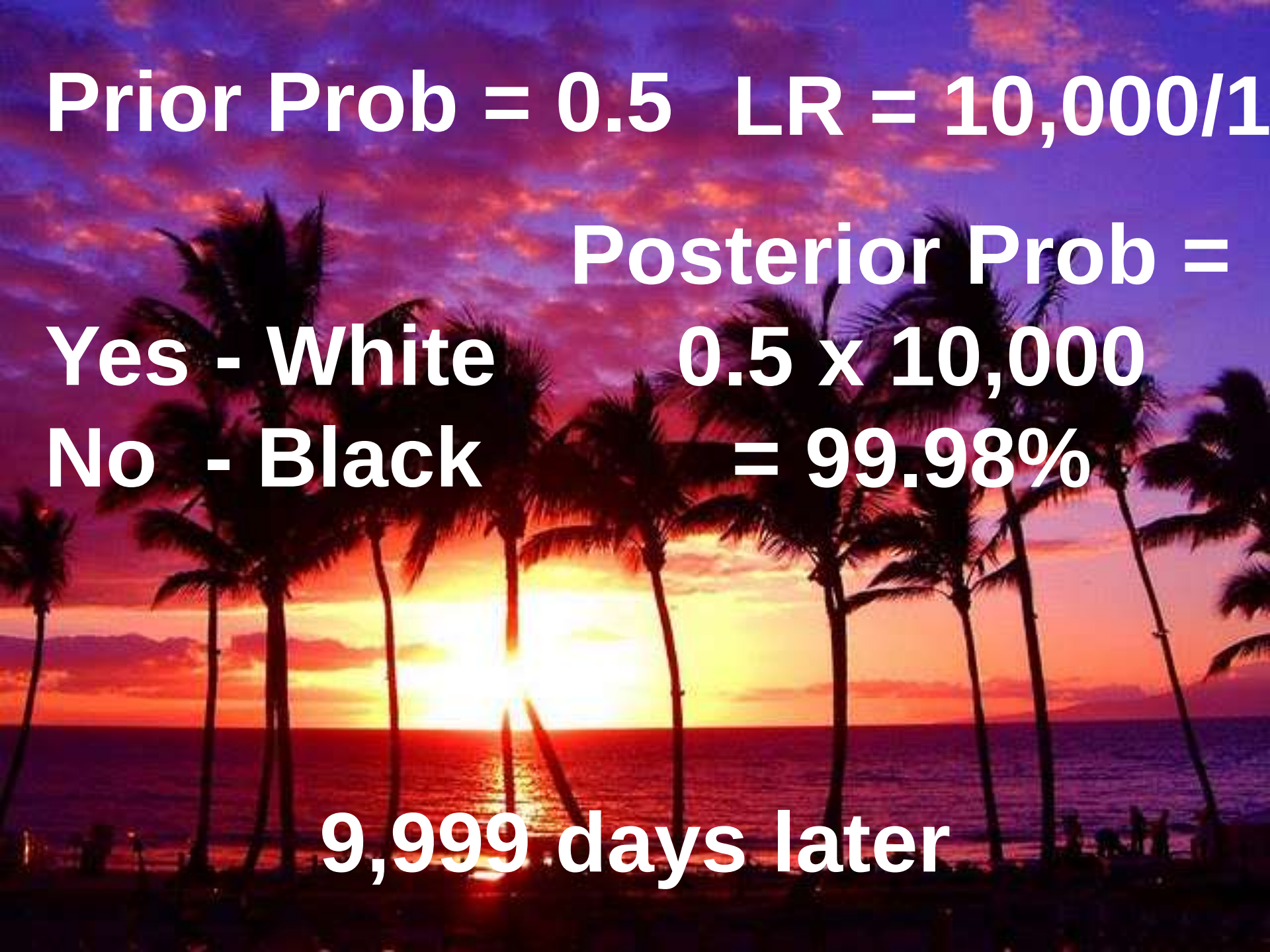
True Allele also uses a Bayesian Analysis of the data



Bayes' Theorem

$$\frac{P(H_1|E)}{P(H_2|E)} = \frac{P(H_1)}{P(H_2)} \cdot \frac{P(E|H_1)}{P(E|H_2)}$$

Posterior Probability	Prior Probability	Likelihood Ratio


Prior Prob = 0.5 LR = 10,000/1
Posterior Prob =
Yes - White
No - Black
0.5 x 10,000
= 99.98%
9,999 days later

Little Orphan Alien…

The sun'll come out tomorrow

With a 99.98% probability

tomorrow there'll be sun

Real-life Example

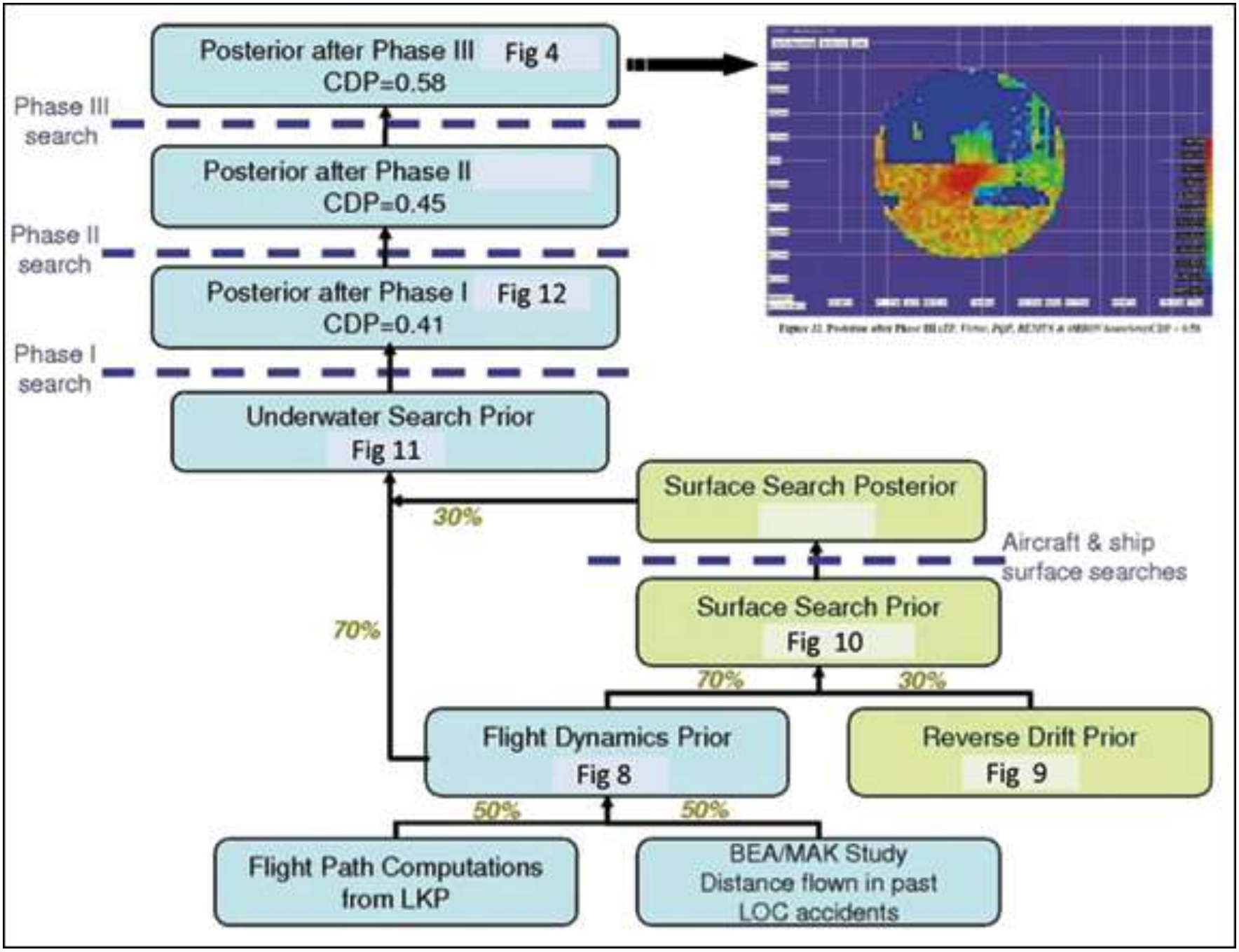
Air France Flight 447

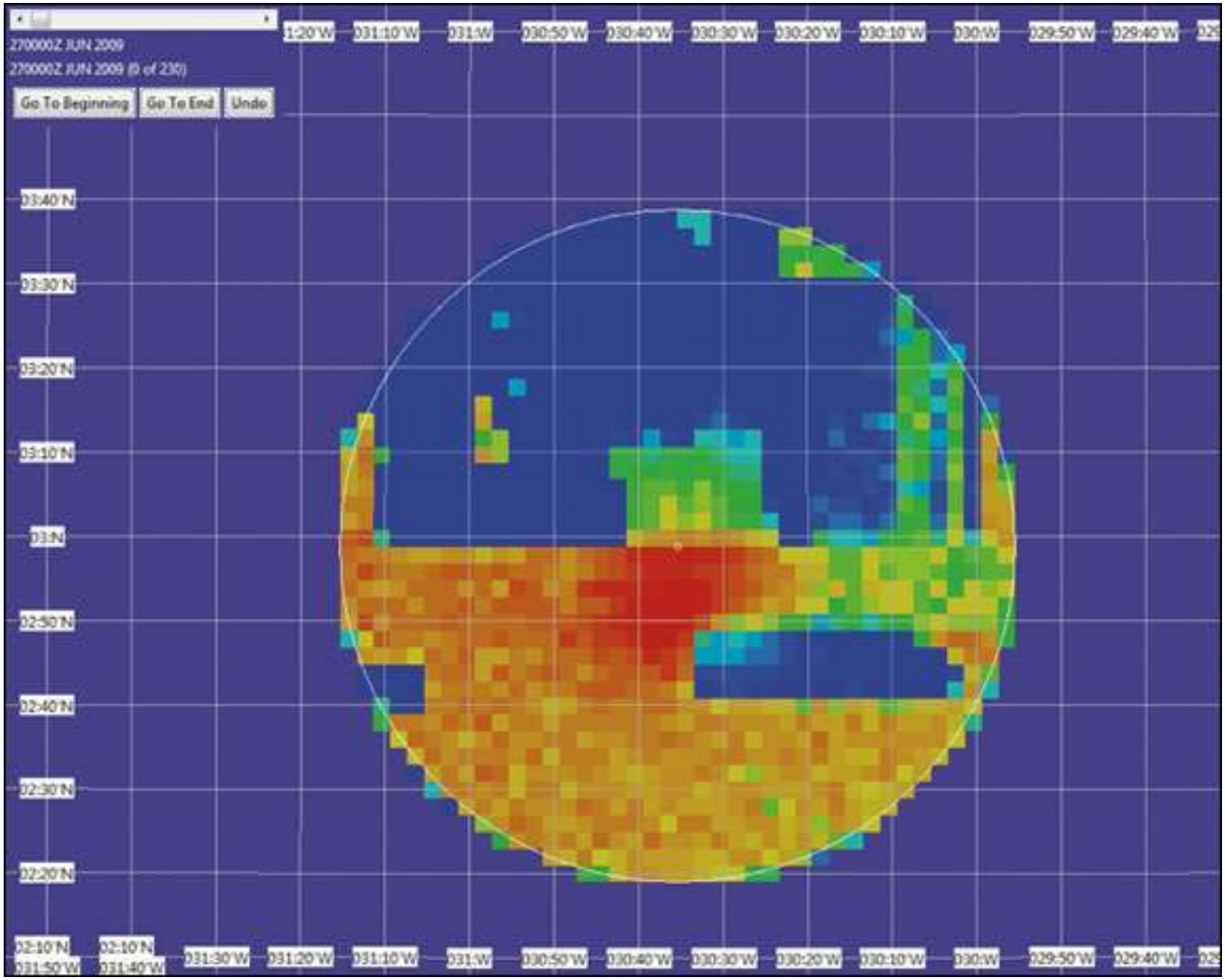
- June 1, 2009, Air France Flight 447, (Rio de Janeiro to Paris) with 228 passengers and crew disappeared over the South Atlantic.
- 33 bodies were located from June 6-10, 2009.
- By June 17, 50 bodies had been recovered in two distinct groups more than 50 miles apart.

Air France Flight 447

- Initial searches conclude at the end of August.
- More searches in 2009 and 2010.
- In July 2010, the US-based search consultancy Metron was asked by BEA (France) to examine the results. Metron uses a Bayesian approach to find the potential crash site.

- http://www.informs.org/ORMS-Today/Public-Articles/August-Volume-38-Number-4/In-Search-of-Air-France-Flight-447


Posterior after Phase III Fig 4
CDP=0.58
Phase III search
Posterior after Phase II
CDP=0.45
Phase II search
Posterior after Phase I Fig 12
CDP=0.41
Phase I search
Underwater Search Prior
Fig 11
Surface Search Posterior
30%
Aircraft & ship surface searches
Surface Search Prior
Fig 10
70%
70%
30%
Flight Dynamics Prior
Fig 8
Reverse Drift Prior
Fig 9
50%
50%
Flight Path Computations from LKP
BEA/MAK Study
Distance flown in past
LOC accidents

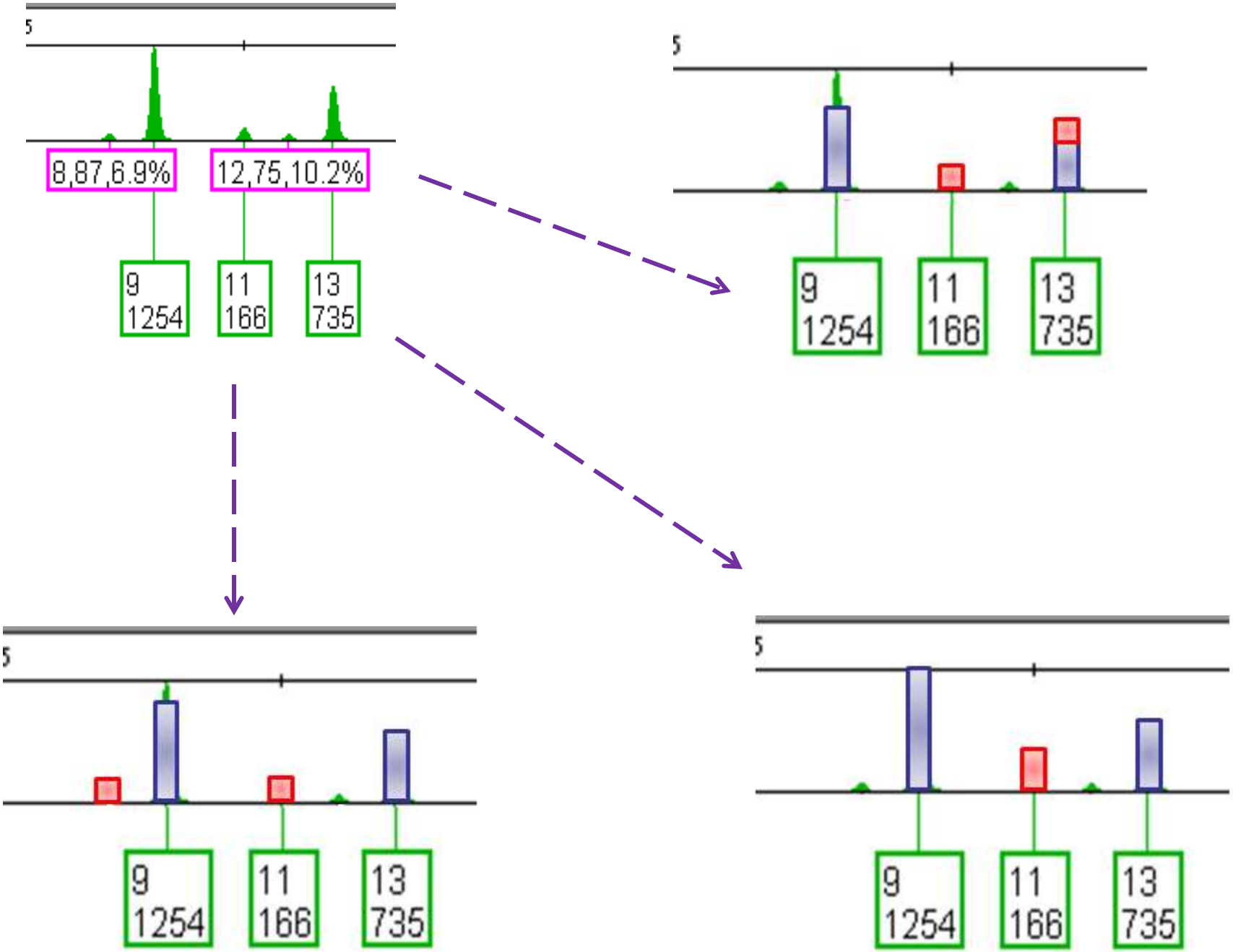

270000Z JUN 2009
270000Z JUN 2009 (0 of 230)
Go To Beginning
Go To End
Undo
1:20'W
031:10'W
031'W
030:50'W
030:40'W
030:30'W
030:20'W
030:10'W
030'W
029:50'W
029:40'W
03:40'N
03:30'N
03:20'N
03:10'N
03'N
02:50'N
02:40'N
02:30'N
02:20'N
02:10'N
031:50'W
02:10'N
031:40'W
031:30'W
031:20'W
031:10'W
031'W
030:50'W
030:40'W
030:30'W
030:20'W
030:10'W
030'W
029:50'W
029:40'W

Air France Flight 447

- January 2011 – Metron published their findings on the BEA website using a Bayesian approach to find the potential crash site.

- Fourth phase initiated in April 2011 – debris field was found within a week. Flight recorders were found in May 2011.

- http://www.informs.org/ORMS-Today/Public-Articles/August-Volume-38-Number-4/In-Search-of-Air-France-Flight-447


5
8,87,6.9%
12,75,10.2%
9
1254
11
166
13
735
5
9
1254
11
166
13
735
5
9
1254
11
166
13
735
5
9
1254
11
166
13
735

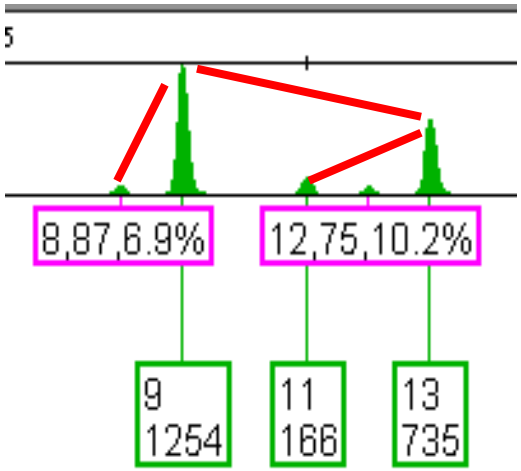
Probabilistic Modeling of TA

Mathematical Modeling of the Data

50-100,000 Simulations (MCMC)

Probable ***Genotypes*** to explain the mixture

PHR, Mix Ratio, Stutter etc…



Genotypes	Probability
9,11	76%
11,11	15%
11,13	2%
8,11	2%
11,12	2%
9,9	1%
9,12	<1%
10,11	<1%
8,12	<1%
8,9	<1%

True Allele Software (Cybergenetics)

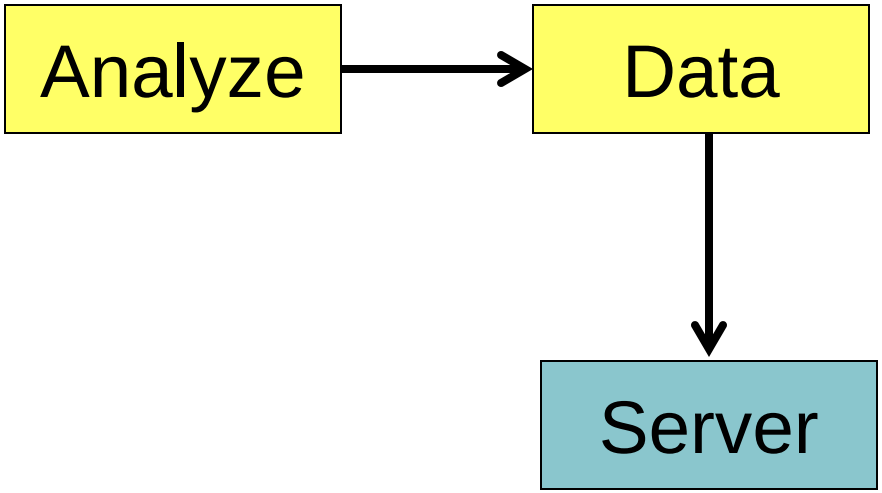
- We purchased the software in September 2010.
- Three day training at Cybergenetics (Pittsburgh, PA) in October.
- Software runs on a Linux Server with a Mac interface.

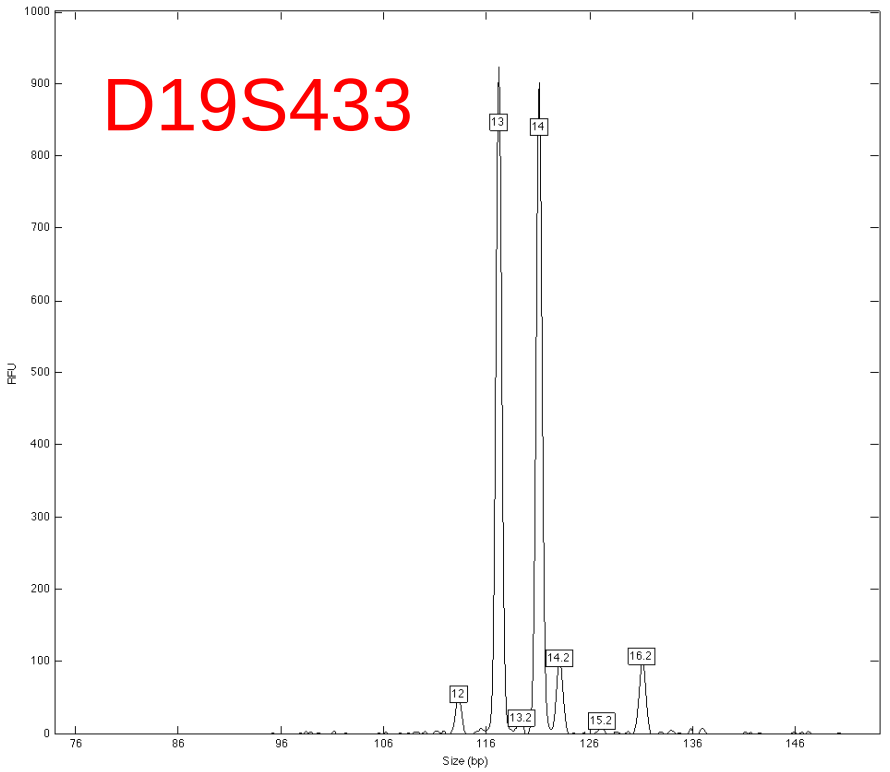
True Allele Casework Workflow
5 Modules

Analyze

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

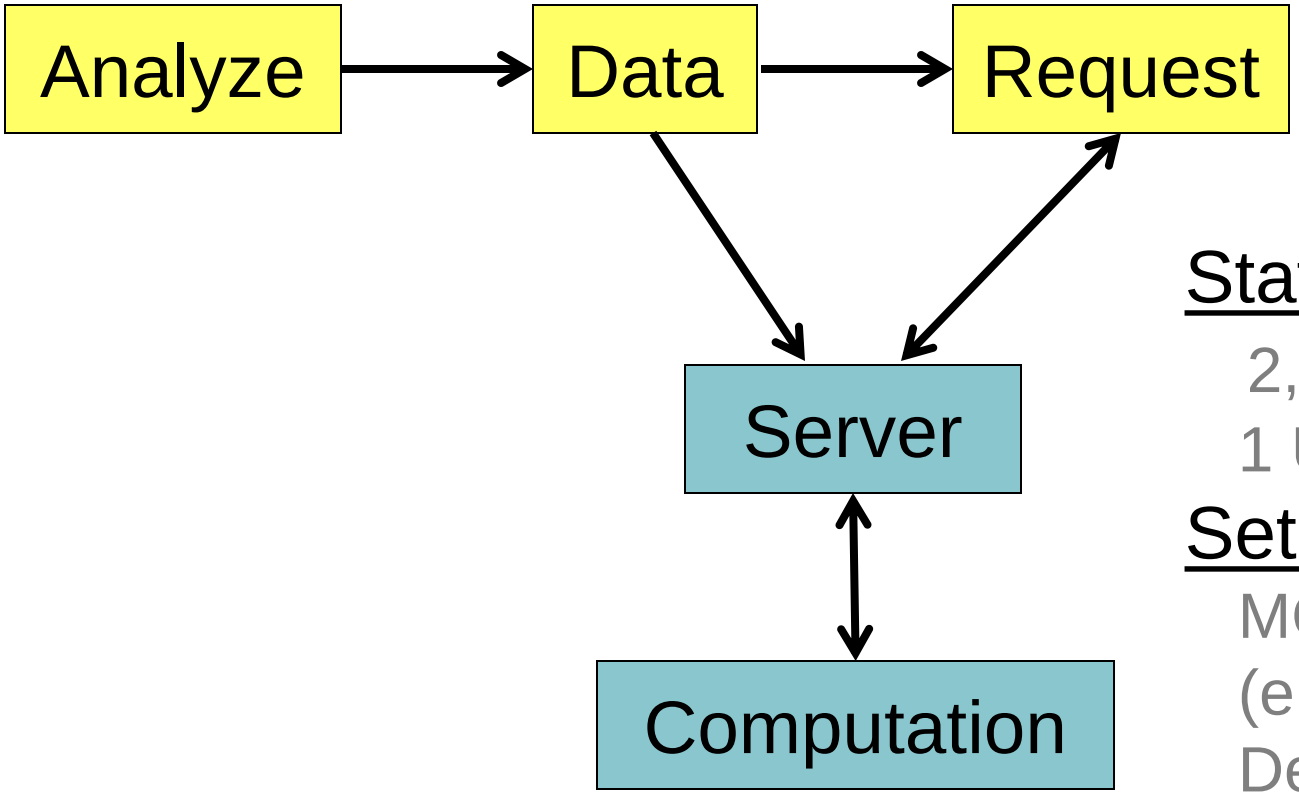
True Allele Casework Workflow
5 Modules





All Peaks above 10 RFU are considered

True Allele Casework Workflow
5 Modules



State Assumptions

2, 3, 4 unknowns

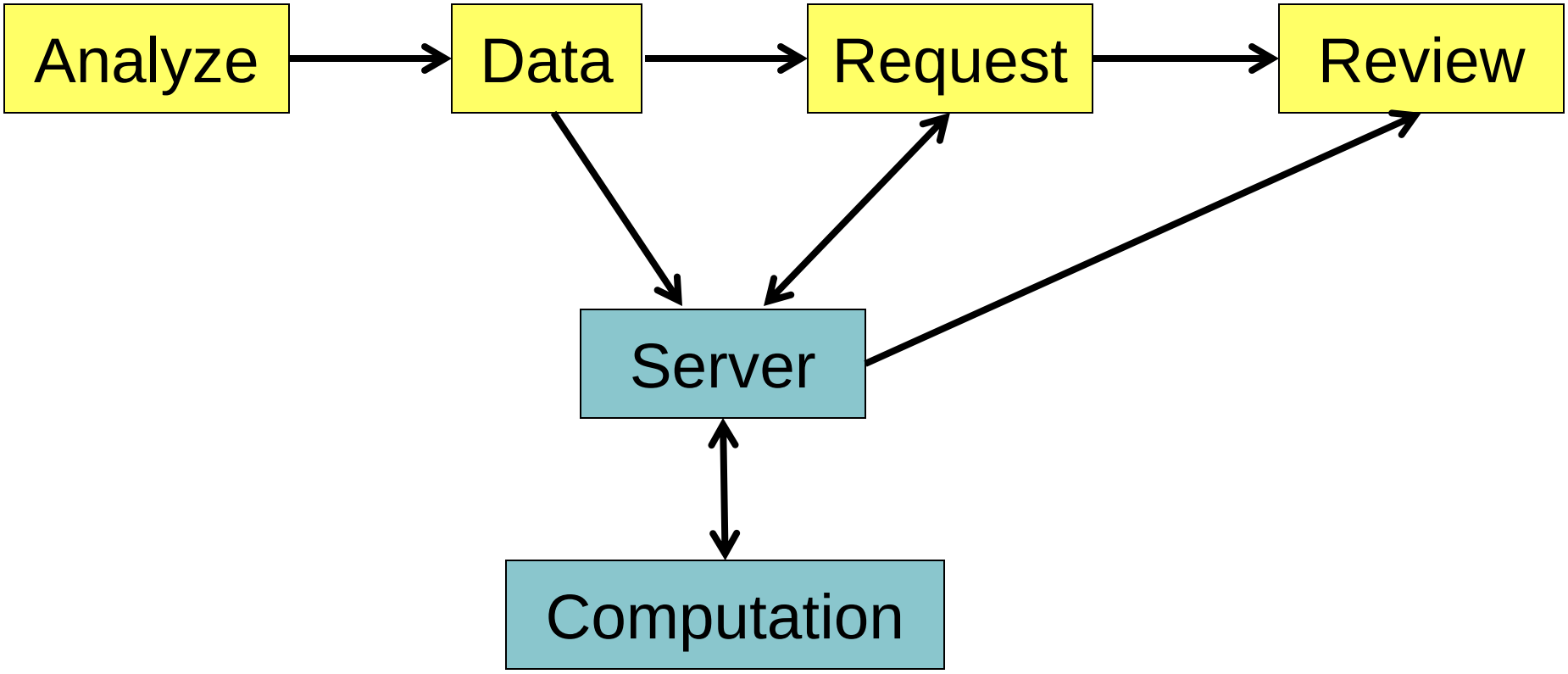
1 Unk with Victim?

Set Parameters

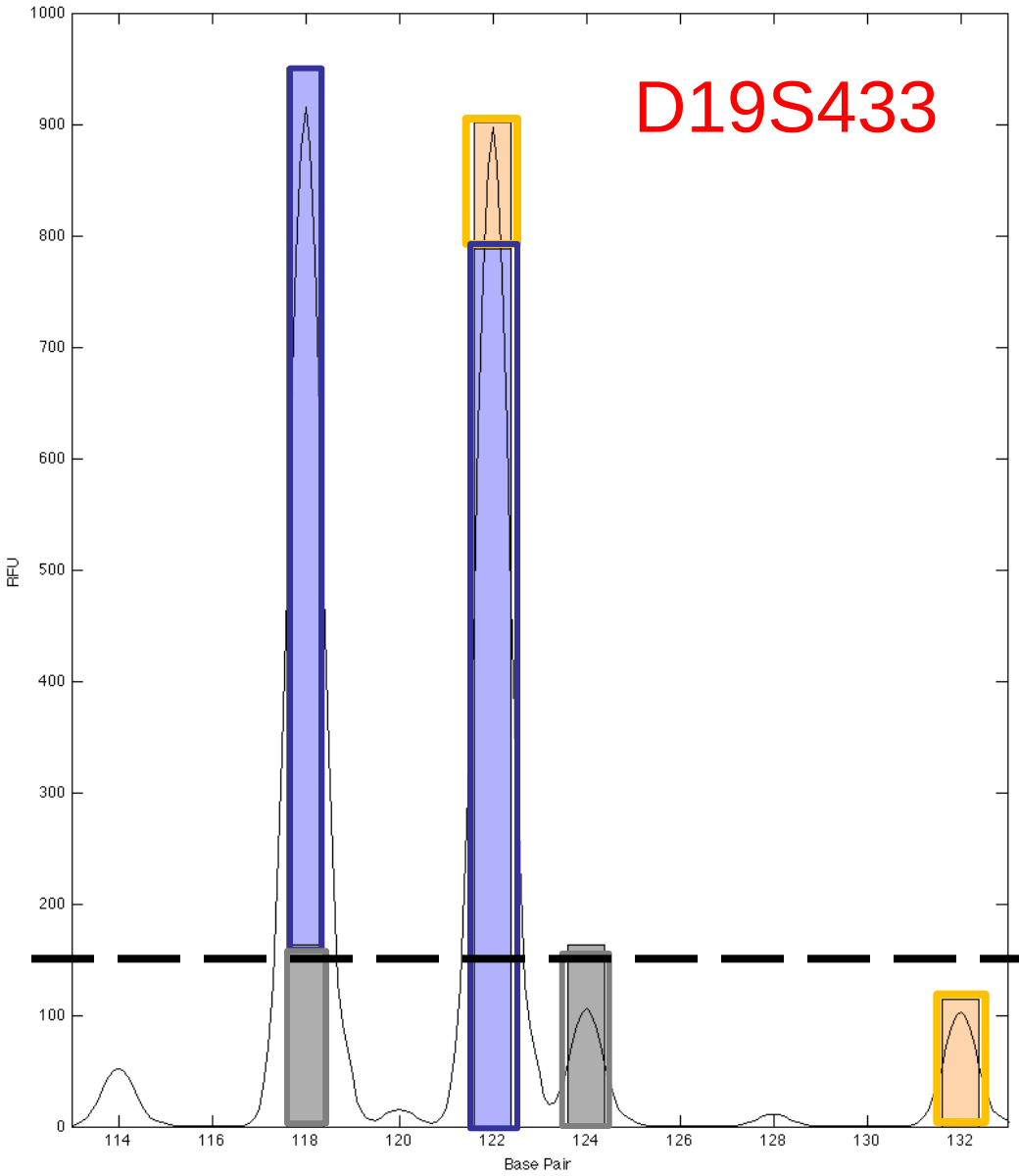
MCMC modeling (e.g.50K)

Degradation?

True Allele Casework Workflow
5 Modules



Review of One Replicate (of 50K)

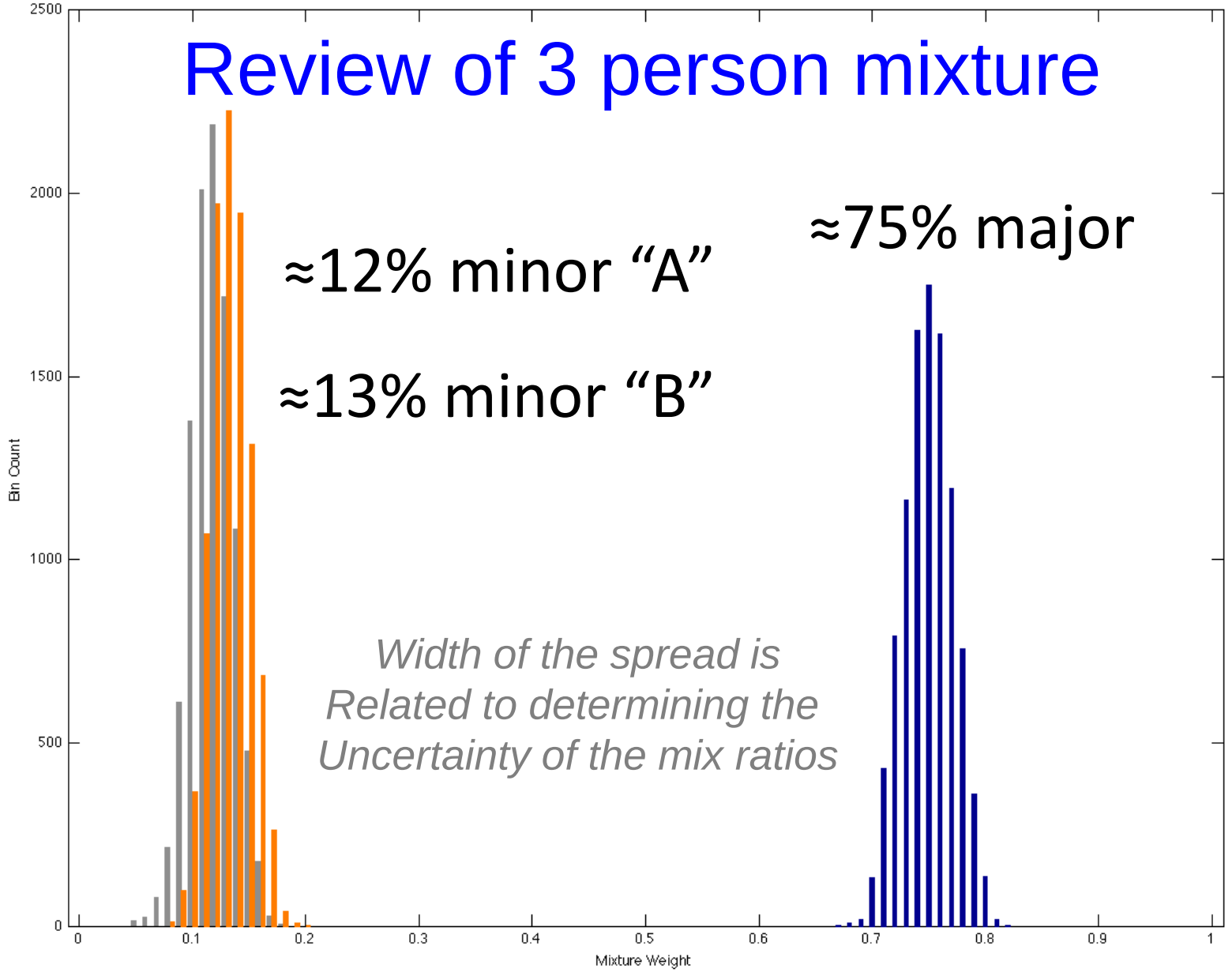


3P mixture,
2 Unknowns,

Conditioned on the Victim (major)

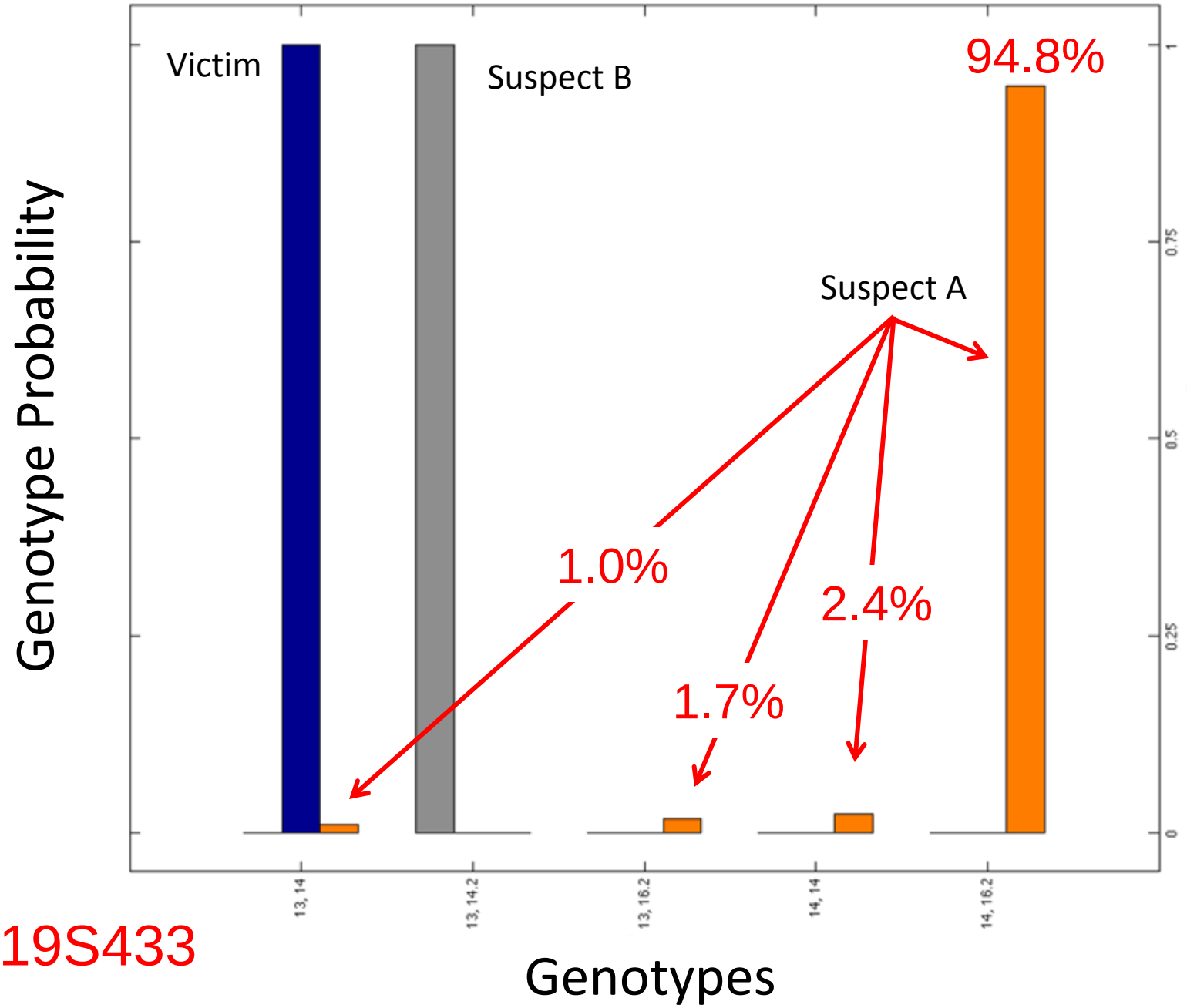
Good fit of the data to the model

Review of 3 person mixture



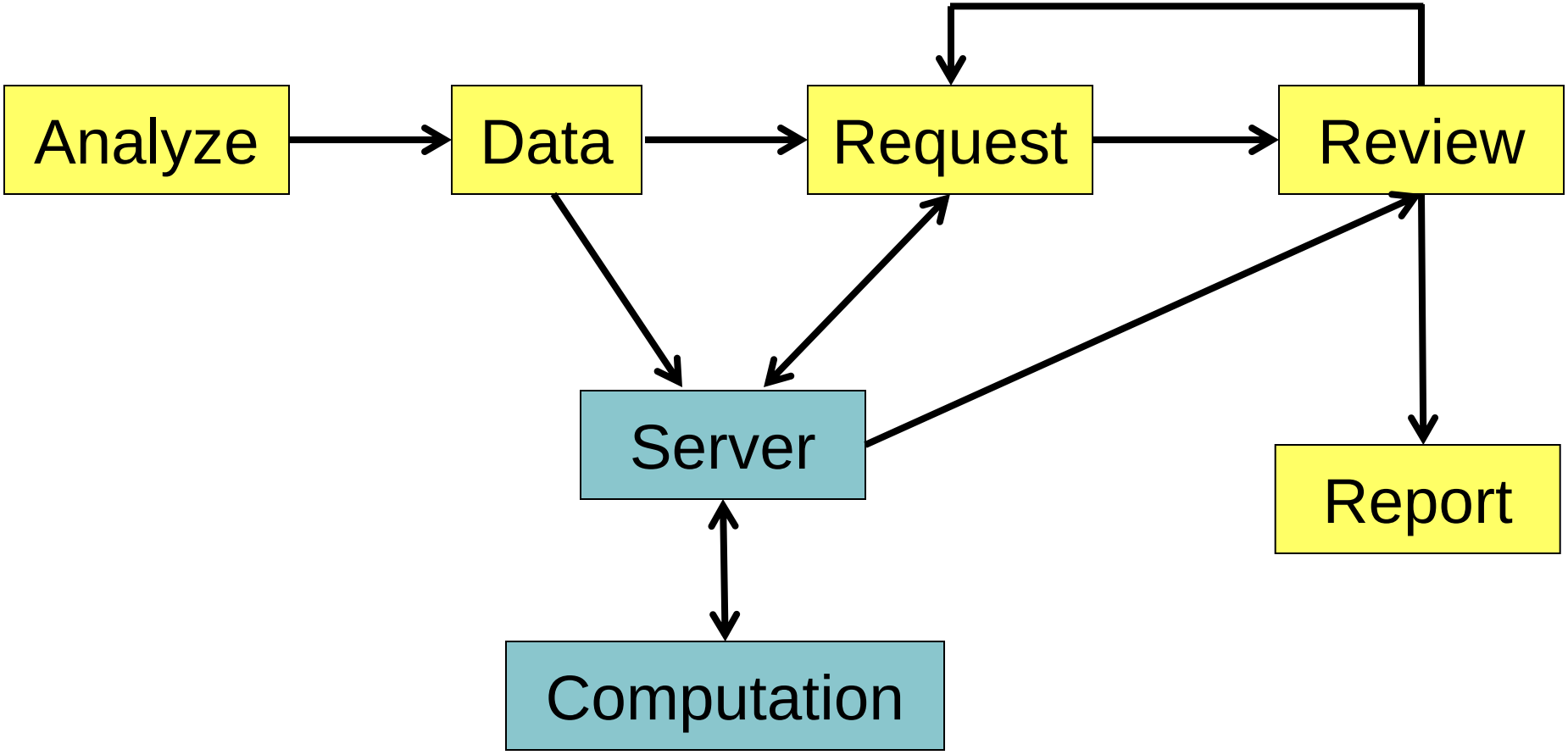
Bin Count

Mixture Weight


Victim
Suspect B
94.8%
Suspect A
1.0%
2.4%
1.7%
Genotype Probability
0
0.25
0.5
0.75
1
13, 14
13, 14.2
13, 16.2
14, 14
14, 16.2
Genotypes

D19S433

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**

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

Combined LR = 5.6 Quintillion

			Genotype Probability Distribution			Weighted Likelihood	Likelihood Ratio		
	allele pair	Likelihood	Questioned	Reference	Suspect	Numerator	Denominator	LR	log(LR)
locus	x	l(x)	q(x)	r(x)	s(x)	l(x)*s(x)	l(x)*r(x)		
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

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

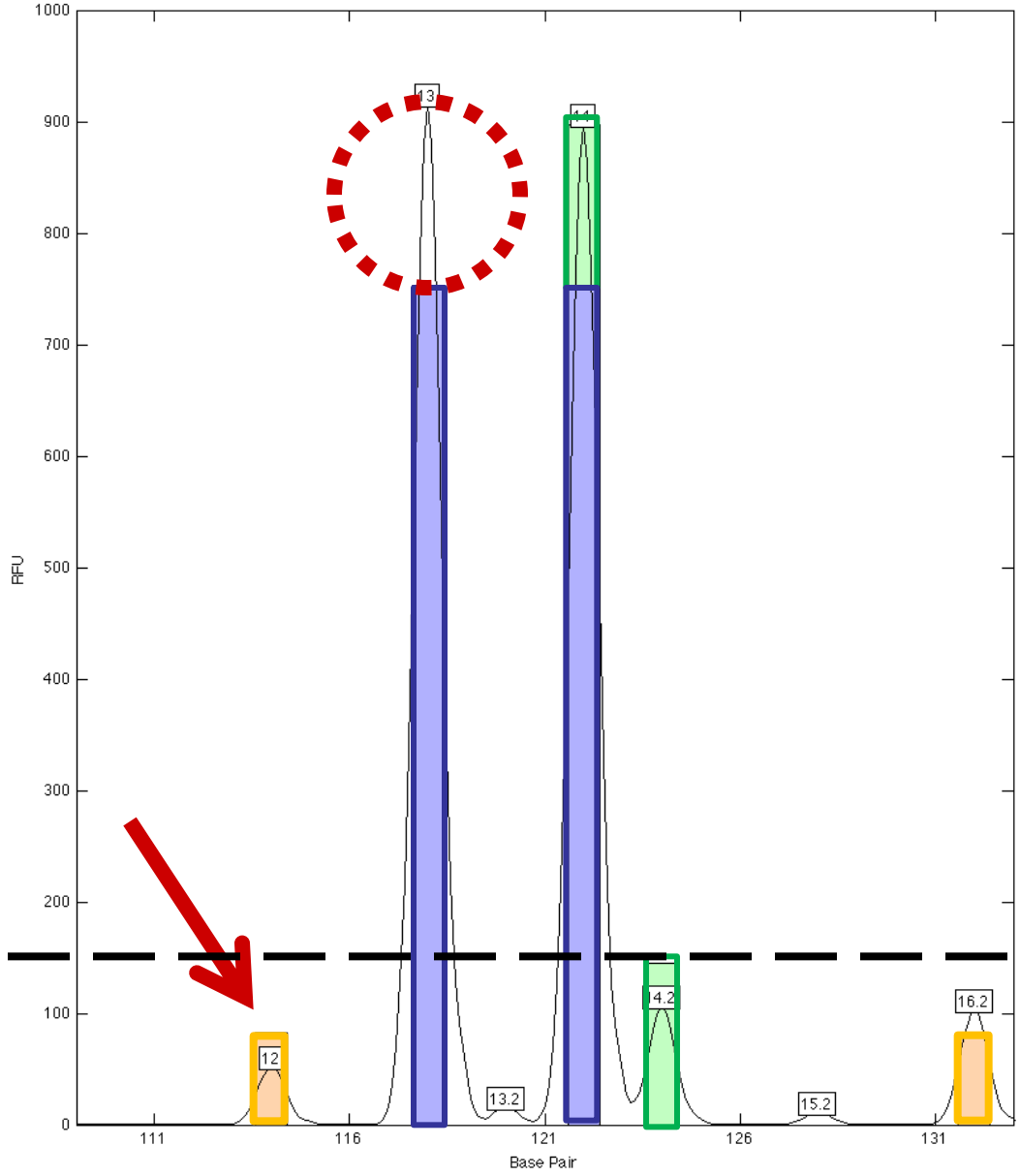
$$\log(LR) = \log 10^6$$

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

$$\log(LR) = 6$$

Review of One Replicate (of 50K)

D19S433

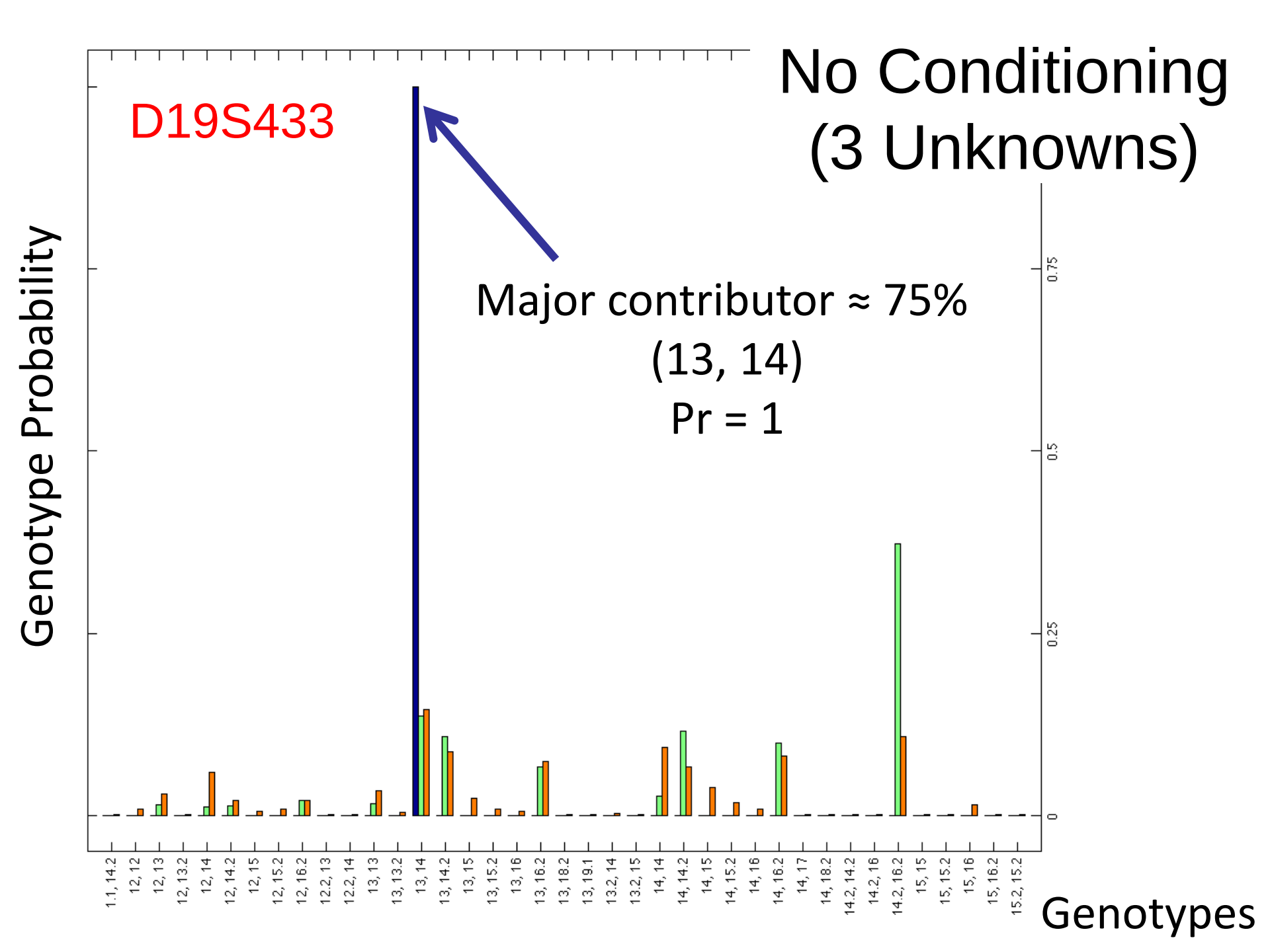


150 RFU

3P mixture,

3 Unknowns

Poor fit of the data to the model

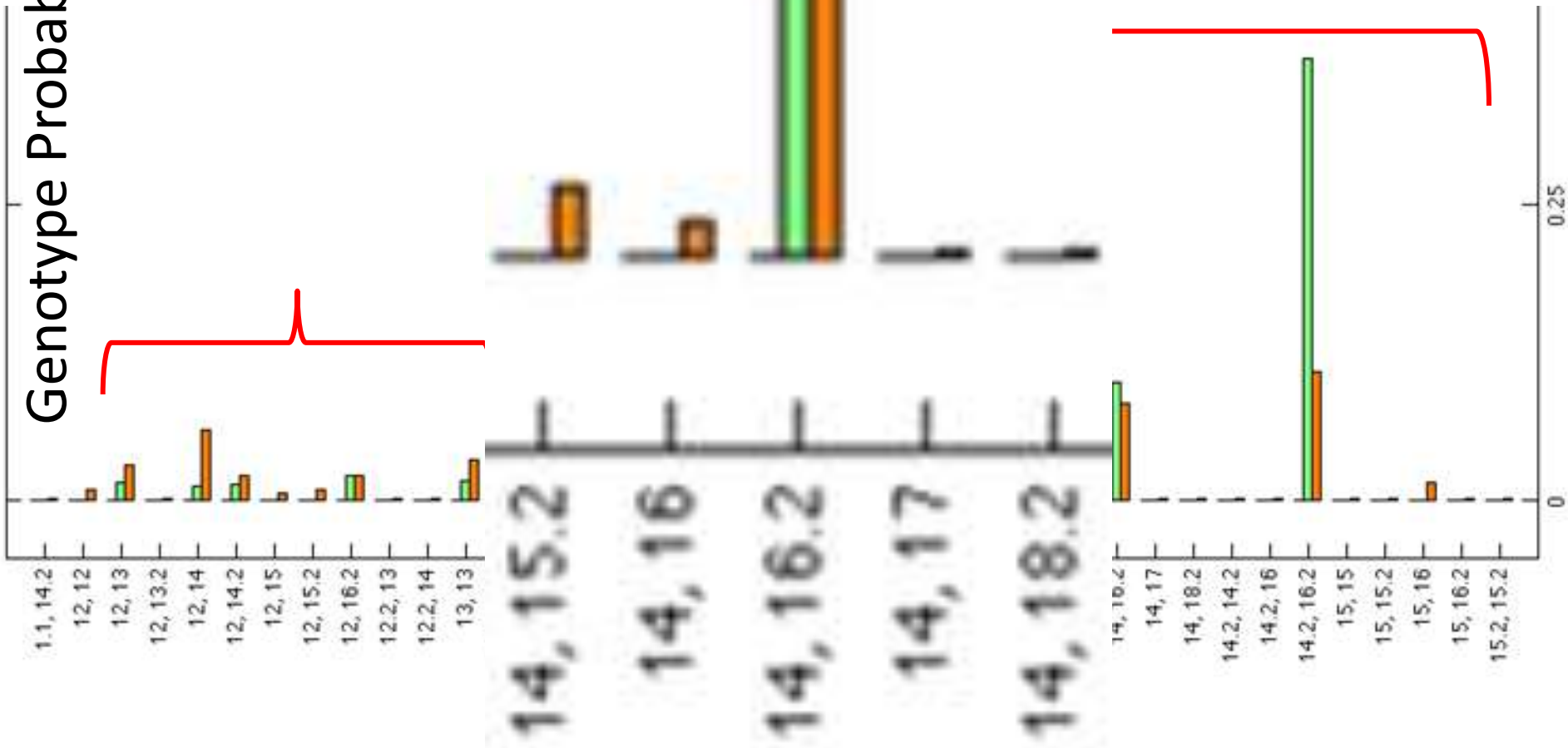

No Conditioning
(3 Unknowns)
D19S433
Major contributor ≈ 75%
(13, 14)
Pr = 1
Genotype Probability
Genotypes
0
0.25
0.5
0.75
1.1, 14.2
12, 12
12, 13
12, 13.2
12, 14
12, 14.2
12, 15
12, 15.2
12, 16.2
12.2, 13
12.2, 14
13, 13
13, 13.2
13, 14
13, 14.2
13, 15
13, 15.2
13, 16
13, 16.2
13, 18.2
13, 19.1
13.2, 14
13.2, 15
14, 14
14, 14.2
14, 15
14, 15.2
14, 16
14, 16.2
14, 17
14, 18.2
14.2, 14.2
14.2, 16
14.2, 16.2
15, 15
15, 15.2
15, 16
15, 16.2
15.2, 15.2

No Conditioning (3 Unknowns)

D19S433

8.1%

ns for the two
ibutors



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**

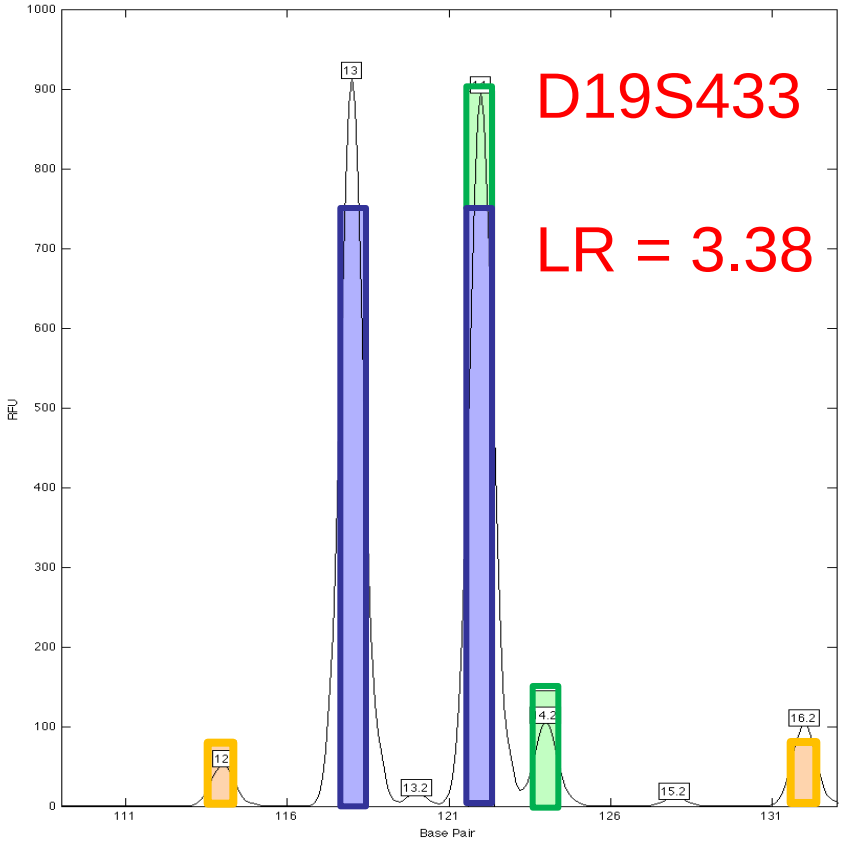
H_D

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

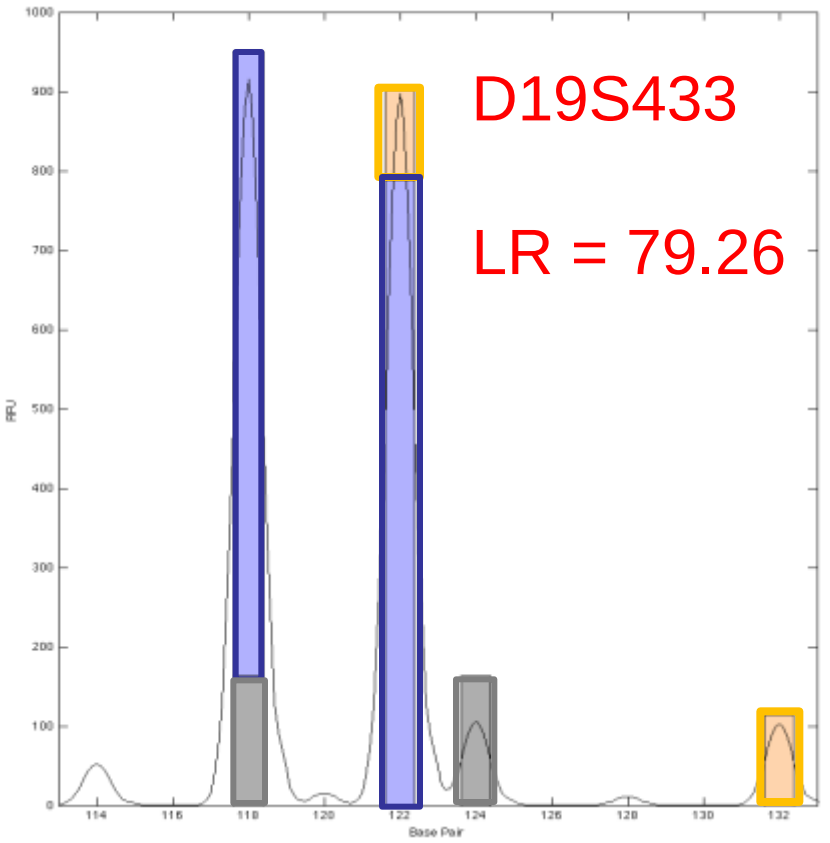
D19S433

No Conditioning (3 Unknowns)

No Conditioning



Conditioned on Victim



Profile - Combined log(LR)

Suspect A log(LR) = 8.03

Suspect B log(LR) = 7.84

Profile - Combined log(LR)

Suspect A log(LR) = 18.72

Suspect B log(LR) = 19.45

Exploring the Capabilities

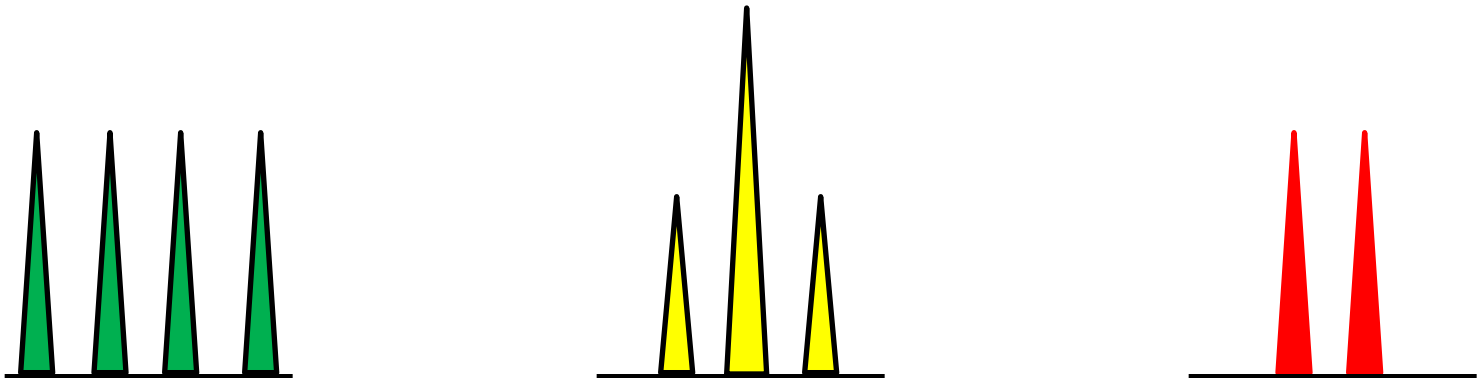
- **Degree of Allele Sharing**
- **Mixture Ratios**
- DNA Quantity

Mixture Data Set

- Mixtures of pristine male and female DNA amplified at a total concentration of 1.0 ng/μL 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

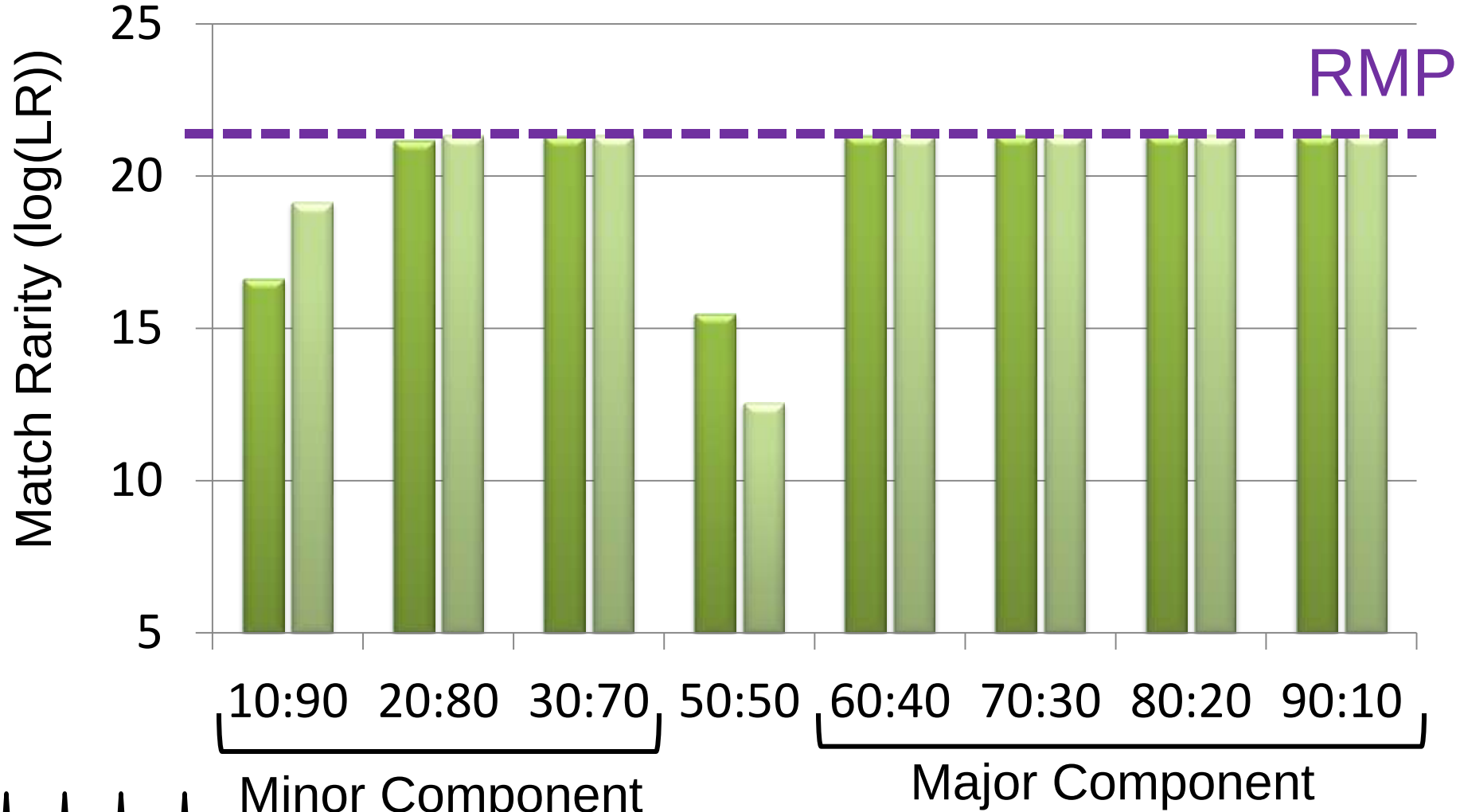
- Three different combinations:

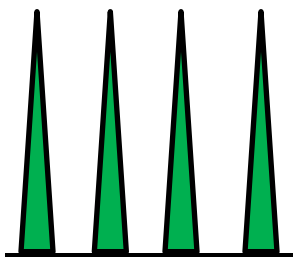


"Low" Sharing	"Medium" Sharing	"High" Sharing
4 alleles – 10 loci	4 alleles – 3 loci	4 alleles – 0 loci
3 alleles – 5 loci	3 alleles – 8 loci	3 alleles – 6 loci
2 alleles – 0 loci	2 alleles – 4 loci	2 alleles – 8 loci
1 allele – 0 loci	1 allele – 0 loci	1 allele – 1 loci

Virtual MixtureMaker - http://www.cstl.nist.gov/strbase/software.htm

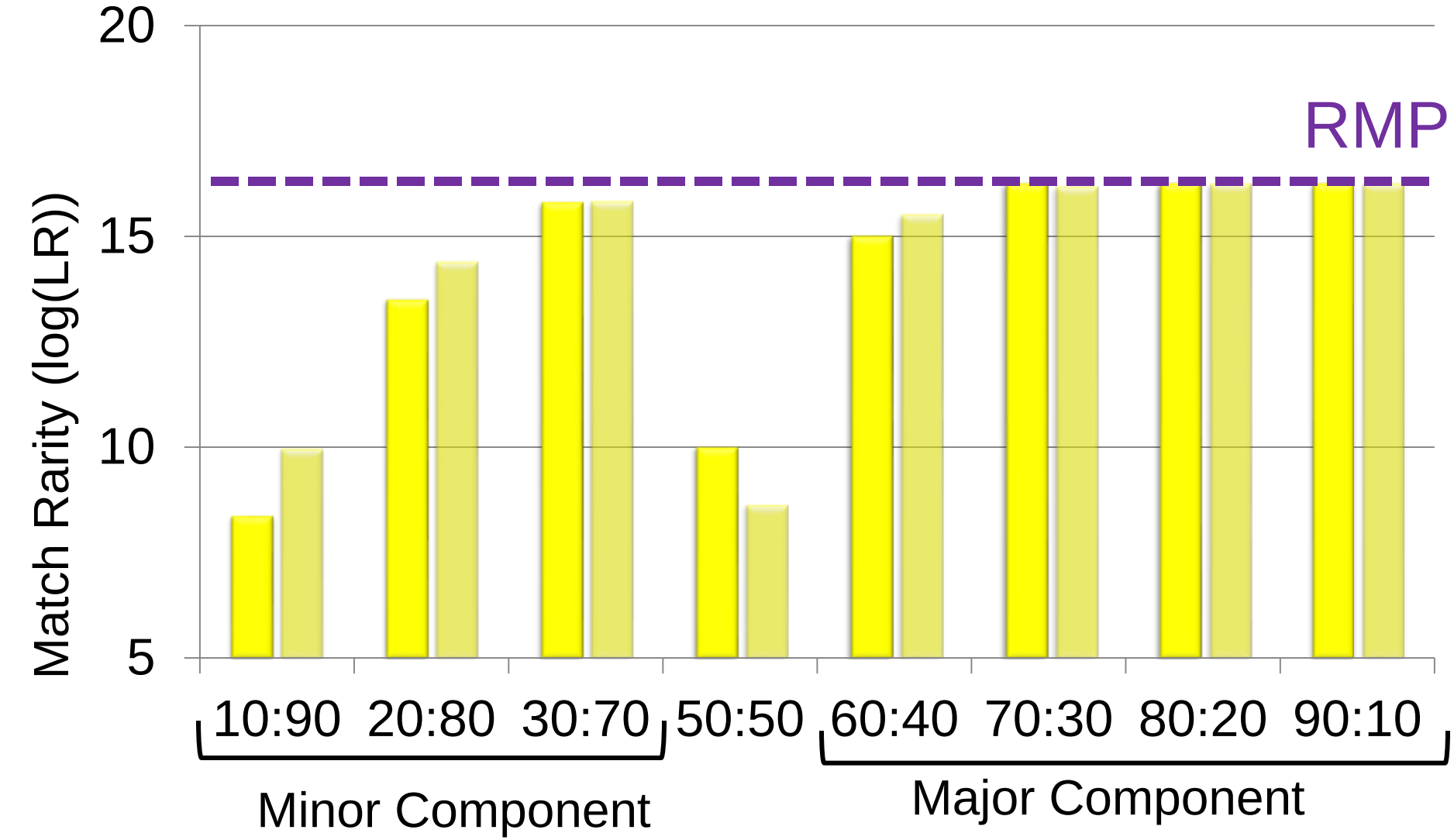
Match Score in Duplicate Runs

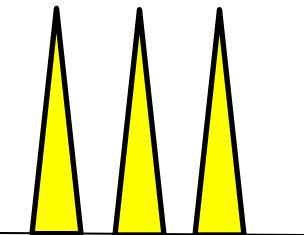




"Easy" for Deconvolution

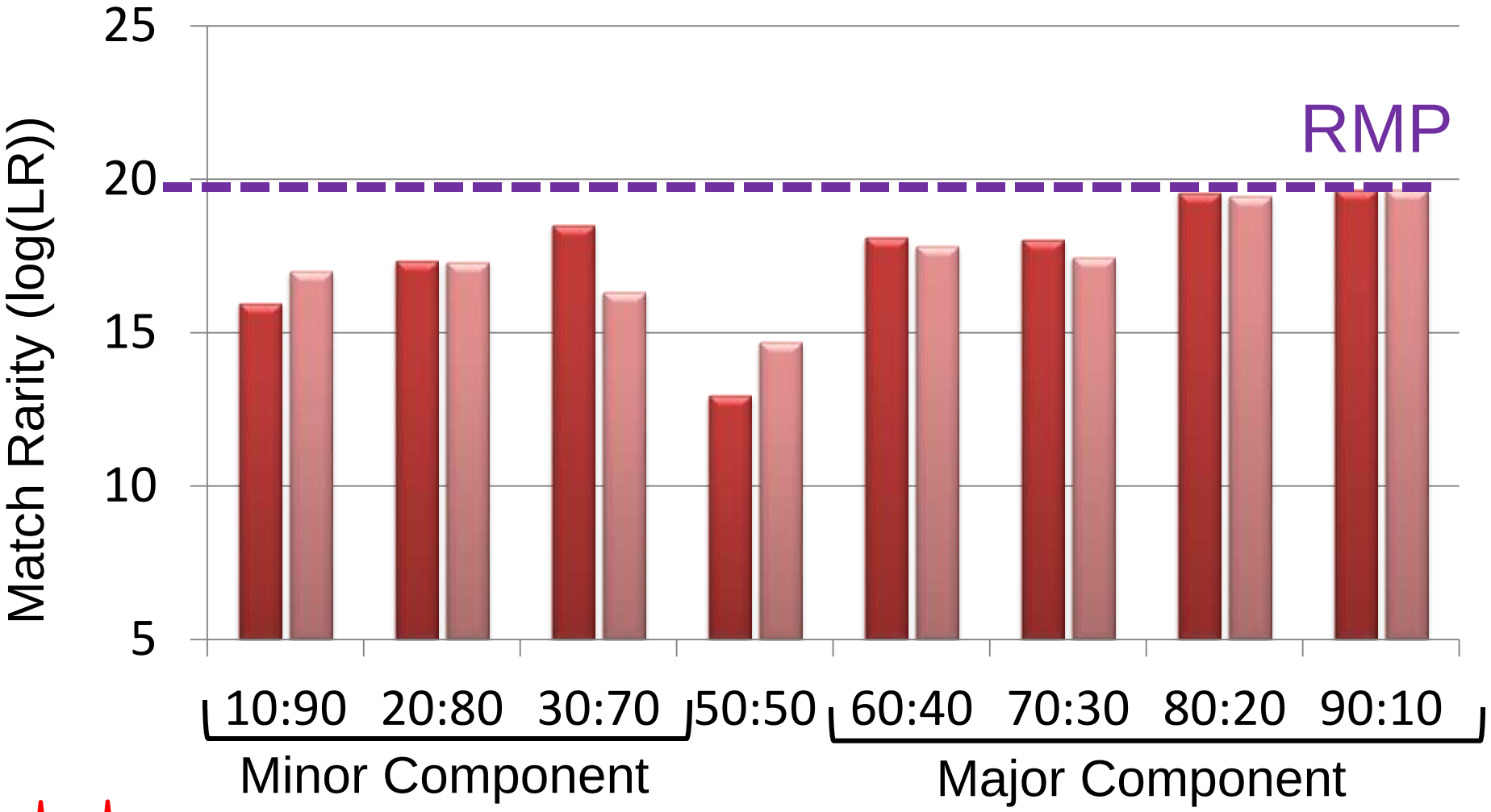
Match Score in Duplicate Runs



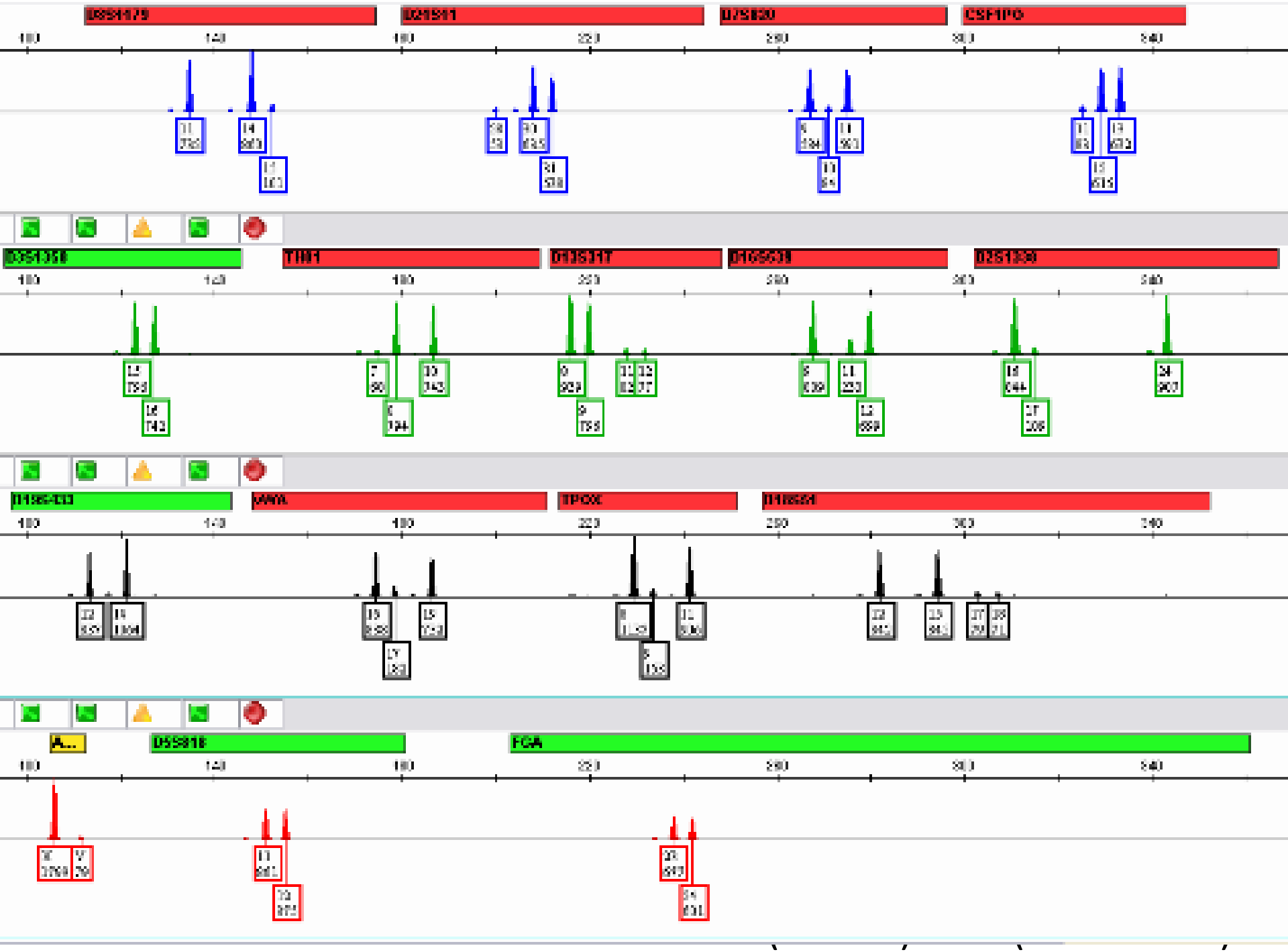


"Challenging" for Deconvolution

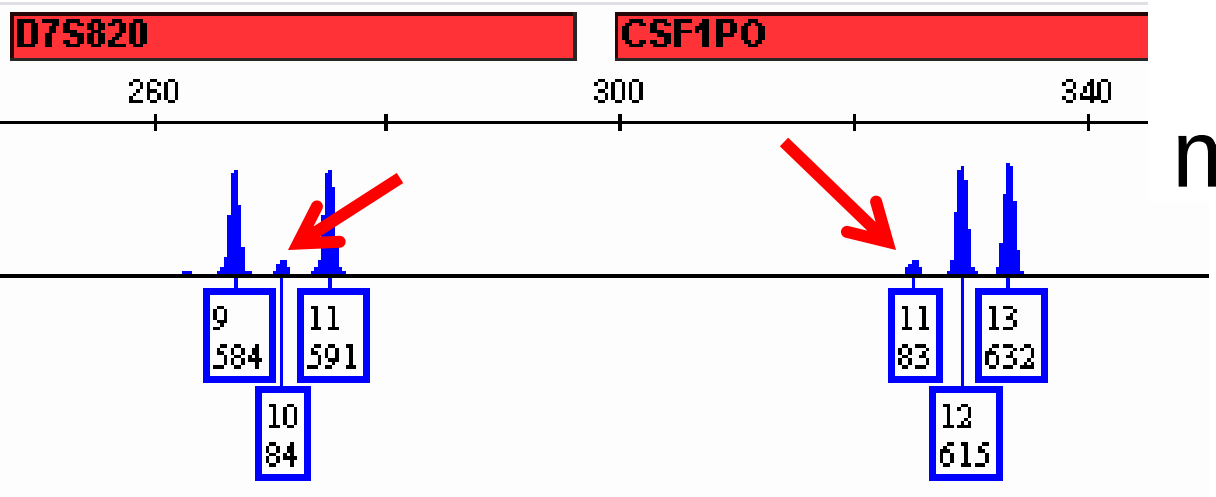
Match Score in Duplicate Runs

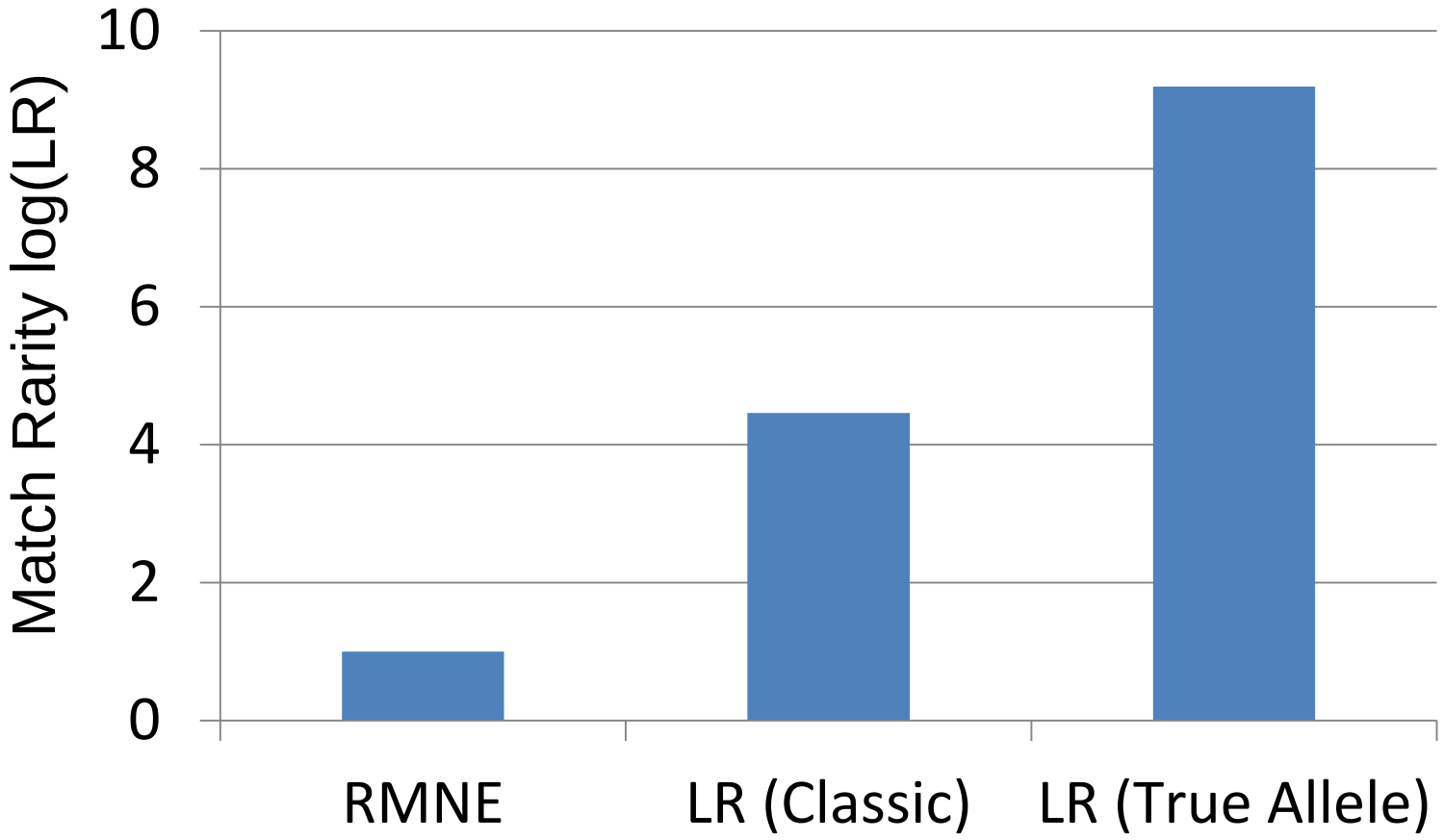


"Difficult" for Deconvolution


D5S818
D21S11
D7S820
CSF1PO
TH01
D13S317
D16S539
D2S1338
AMEL
TPOX
FGA

10:90 minor contributor





Exploring the Capabilities

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

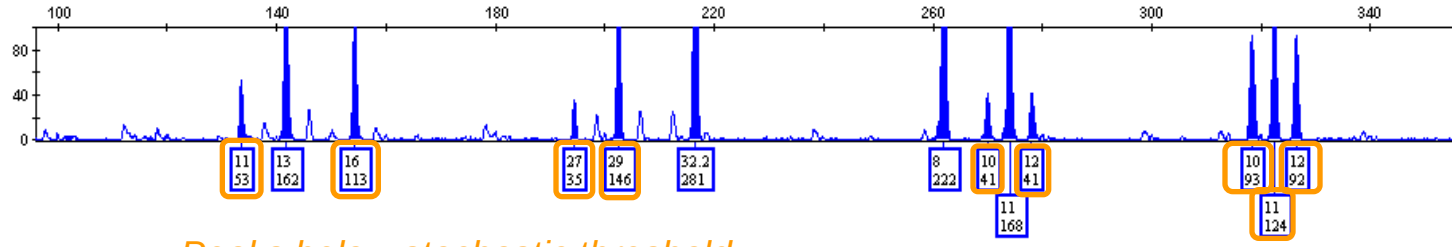
Identifiler

125 pg total DNA

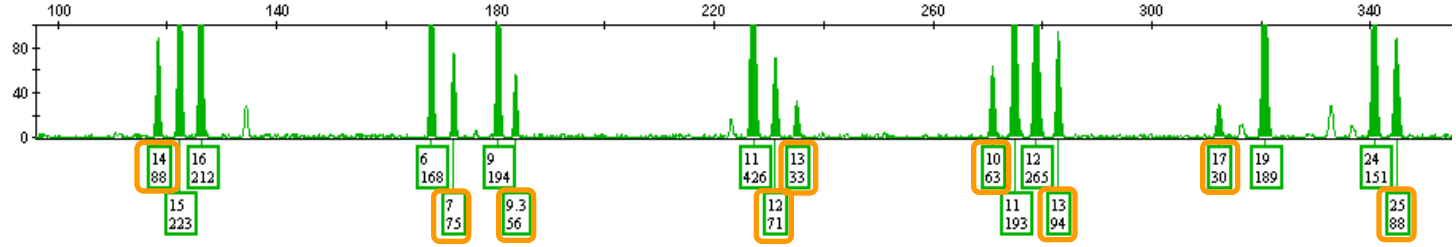
AT = 30 RFU

ST = 150 RFU

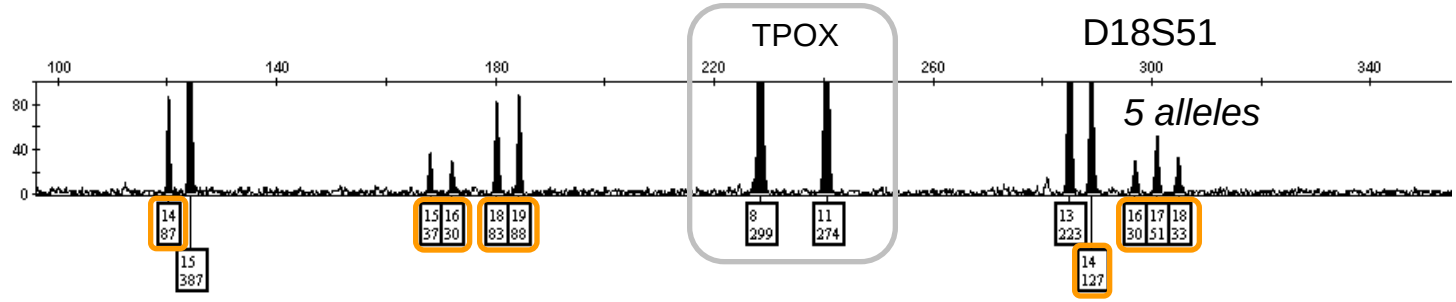
Stutter filter off

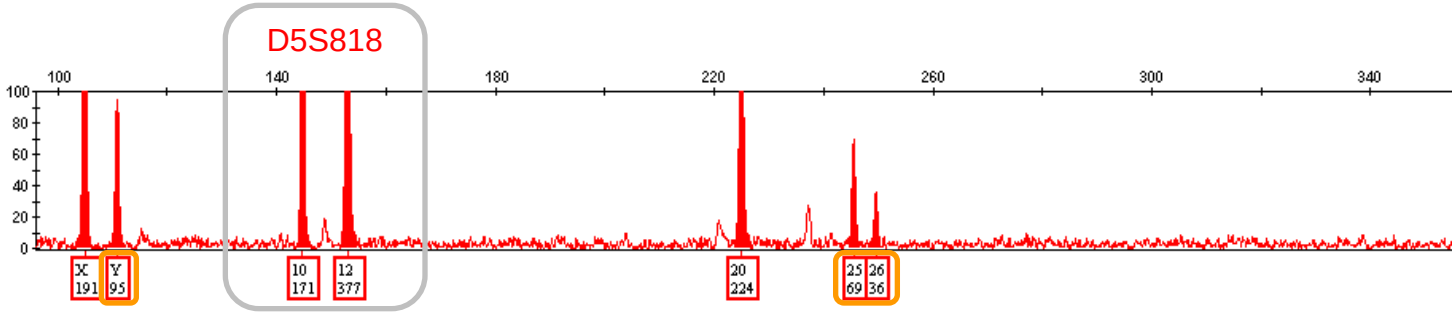


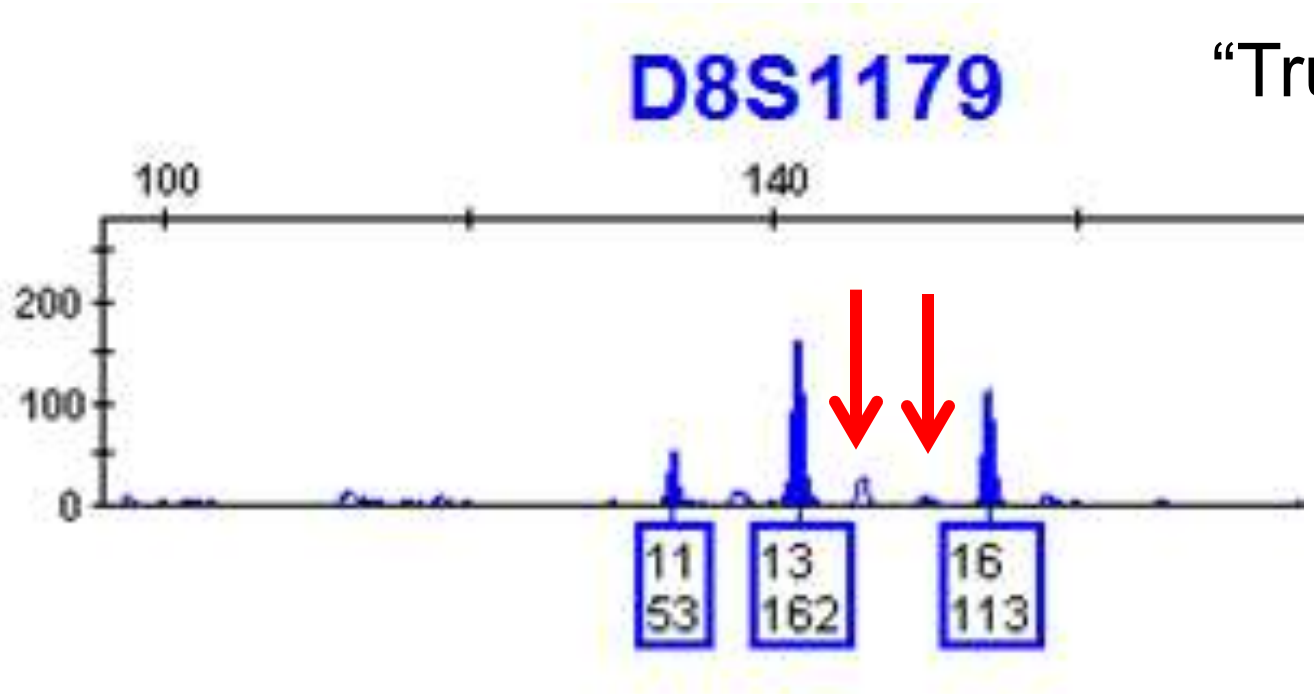
Peaks below stochastic threshold



y-axis zoom to 100 RFU







"True Genotypes"

A = 13, 16

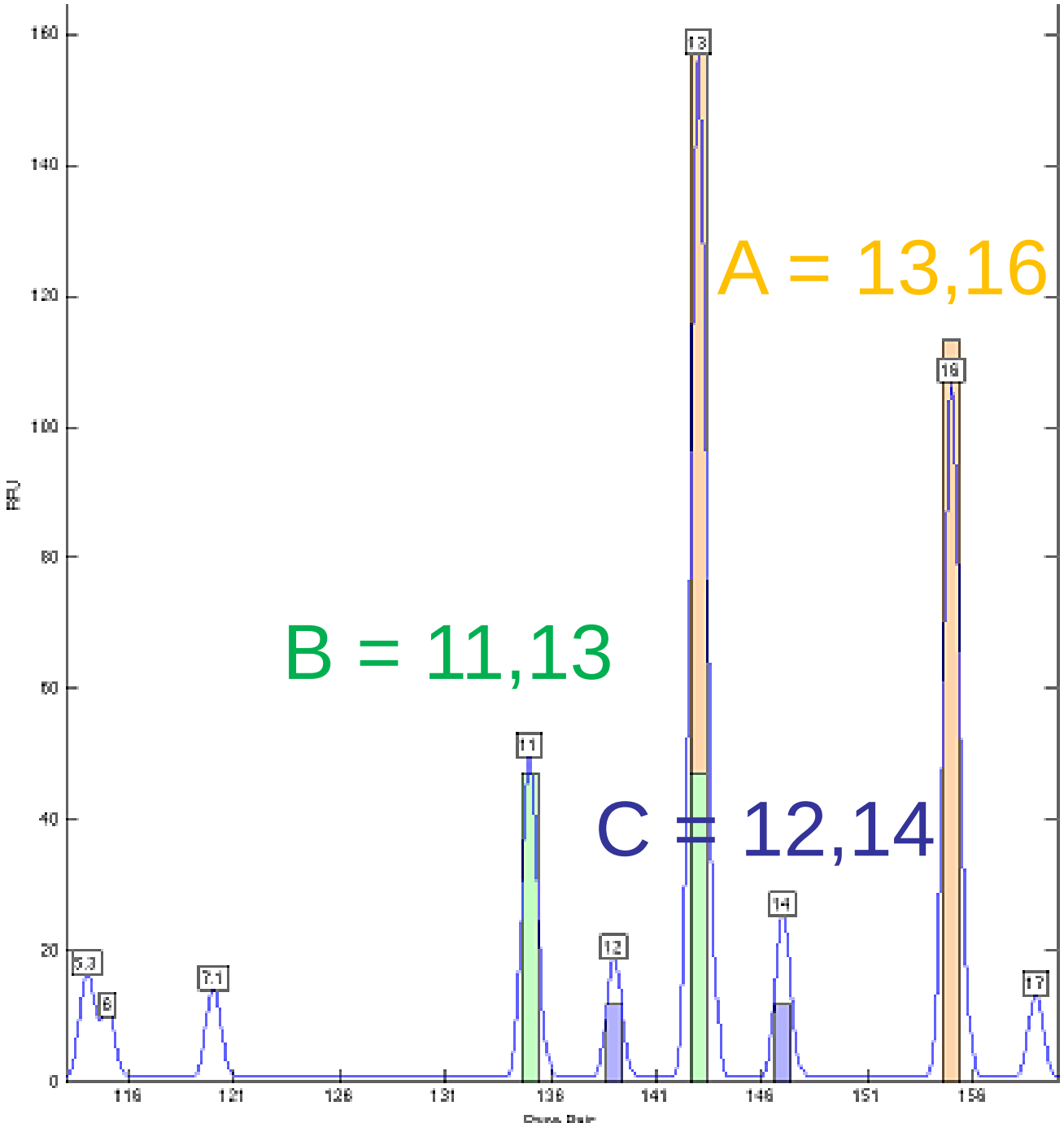
B = 11, 13

C = 14, 15

3 person Mixture – No Conditioning

Major Contributor ≈ 83 pg input DNA

2 Minor Contributors ≈ 21 pg input DNA

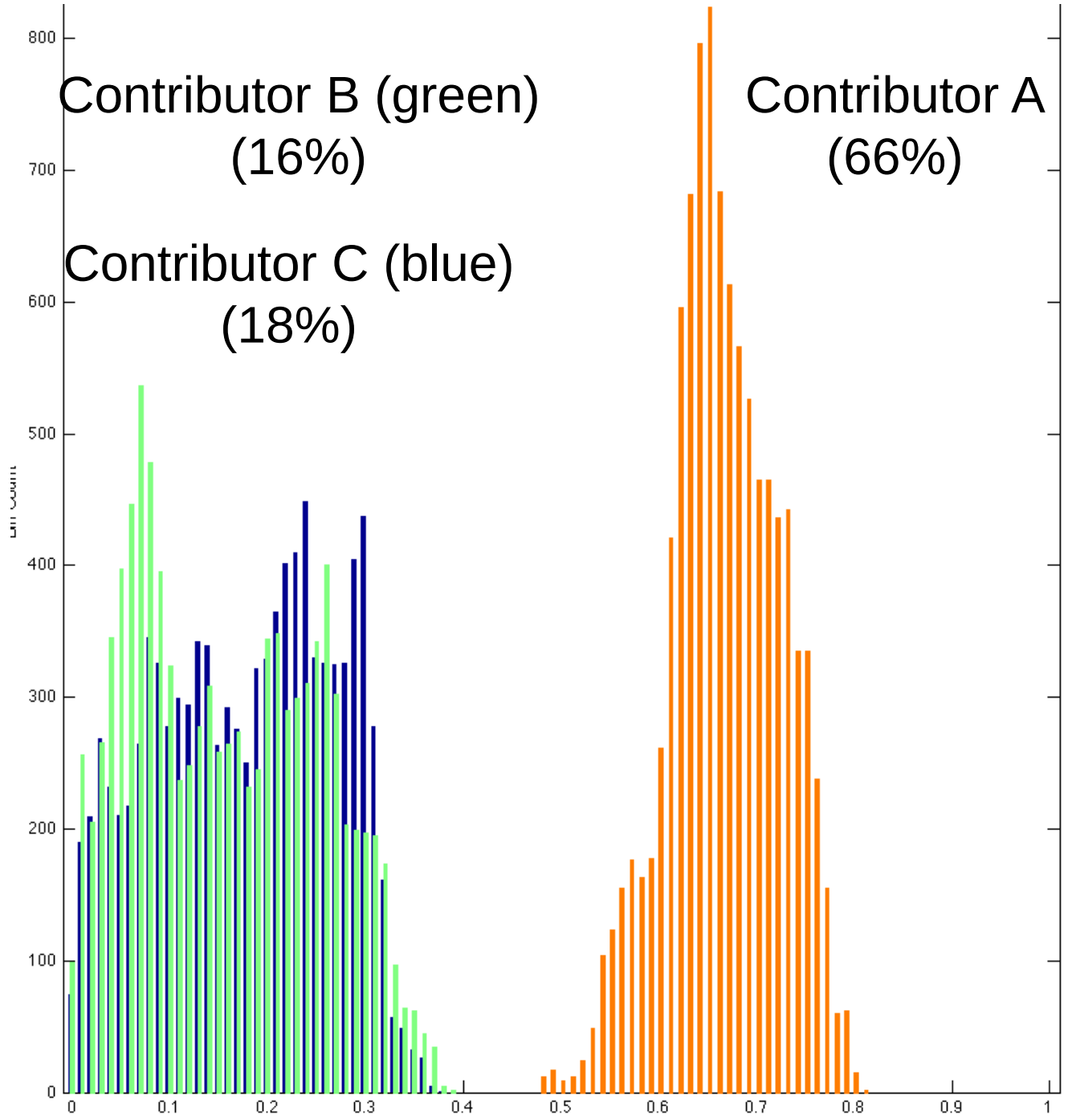


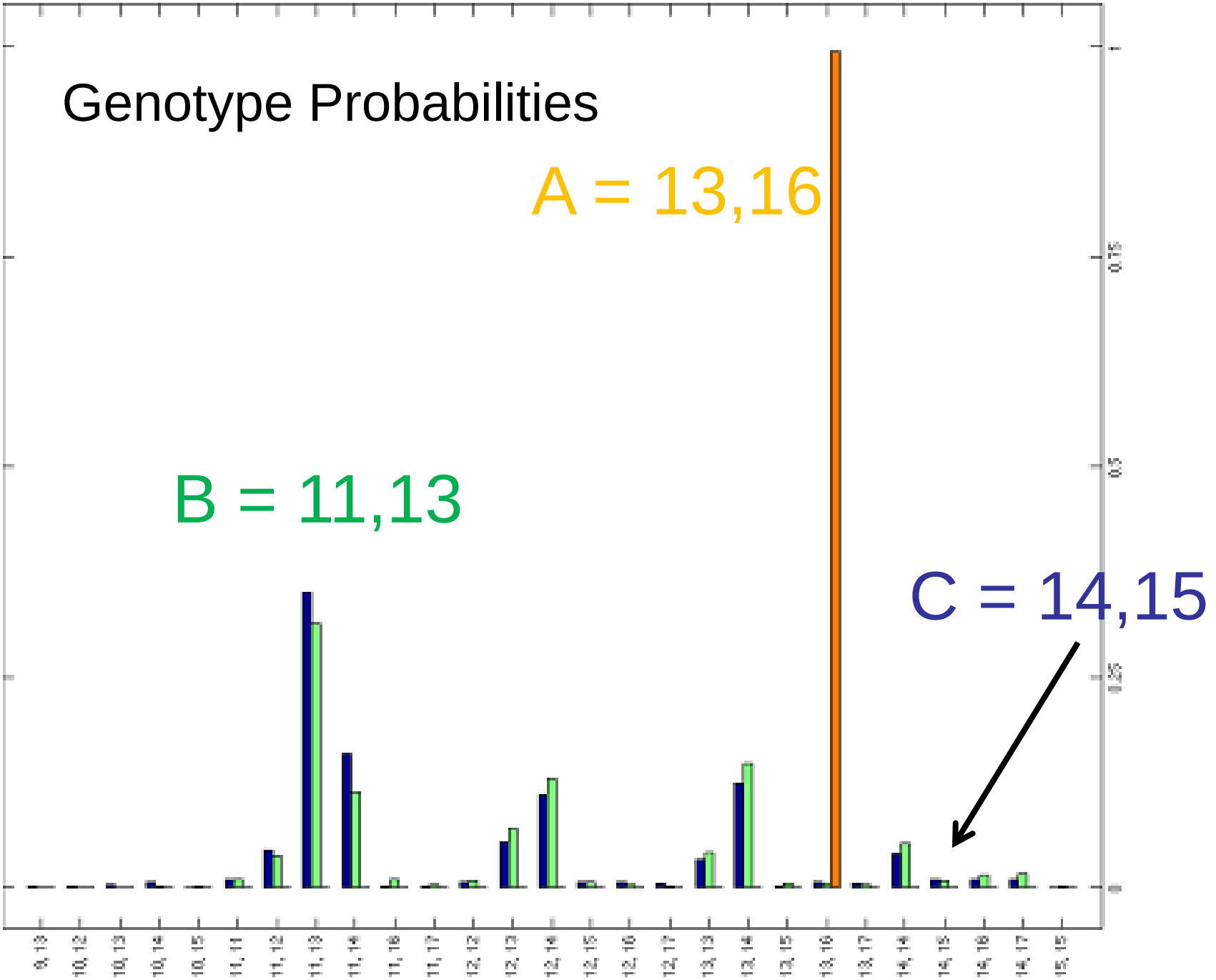
“True Genotypes”

A = 13,16

B = 11,13

C = 14,15


Contributor B (green)
(16%)
Contributor C (blue)
(18%)
Contributor A
(66%)
800
700
600
500
400
300
200
100
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Genotype Probabilities
A = 13,16
B = 11,13
C = 14,15

Results for Contributor A (male)

Locus	Allele Pair	Probability Likelihood	Genotype Frequency	Suspect	H_p Numerator	H_d 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 Likelihood	Genotype Frequency	Suspect	H_p Numerator	H_d Denominator	LR
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)

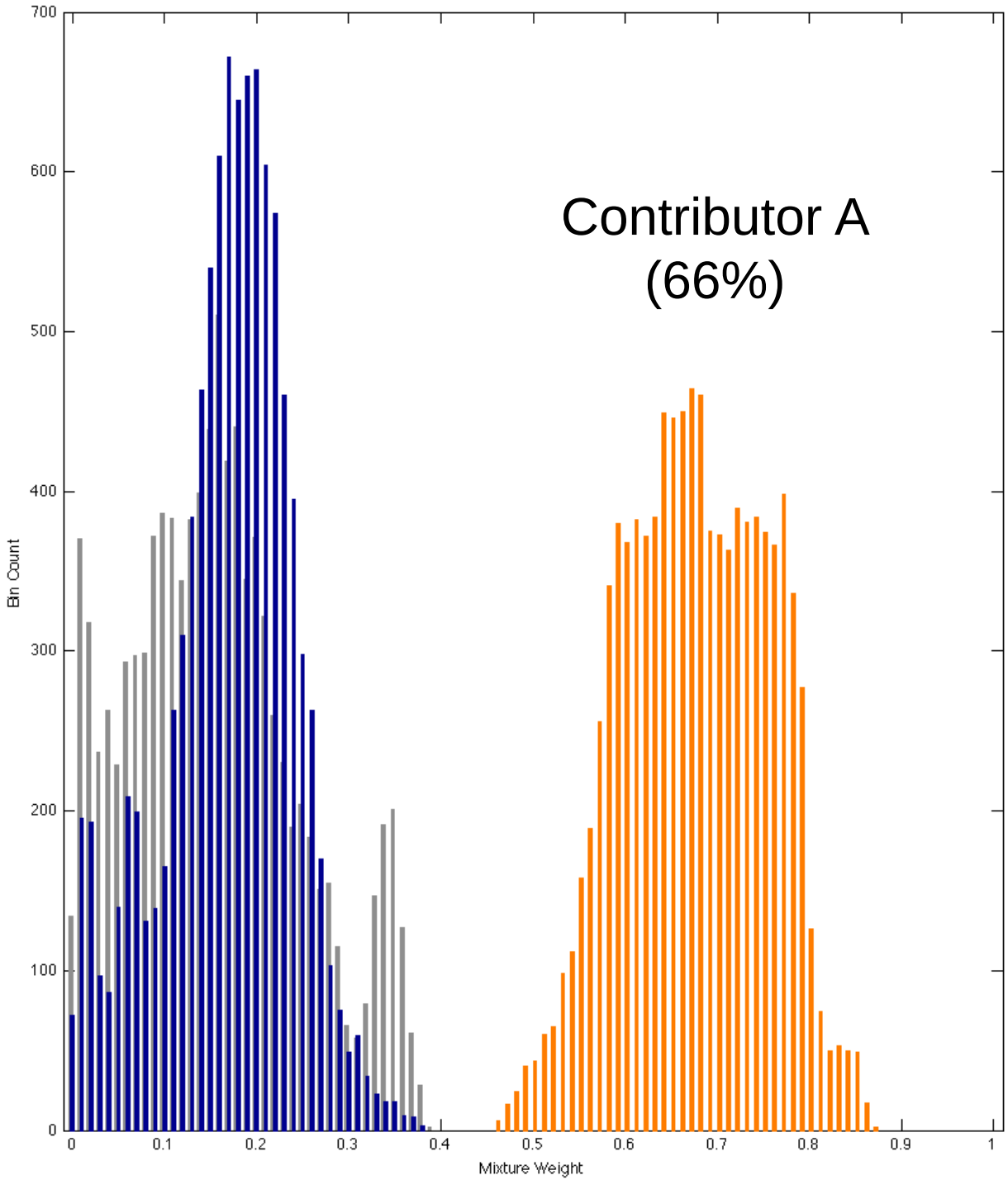
Locus	Allele Pair	Probability Likelihood	Genotype Frequency	Suspect	H_p Numerator	H_d Denominator	LR
D8S1179	11, 13	0.056	0.0498			0.00279	
	13, 14	0.007	0.0996			0.00066	
	12, 14	0.011	0.0606			0.00068	
	11, 14	0.021	0.0271			0.00056	
	12, 13	0.006	0.1115			0.00066	
	14, 14	0.005	0.0271			0.00013	
	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

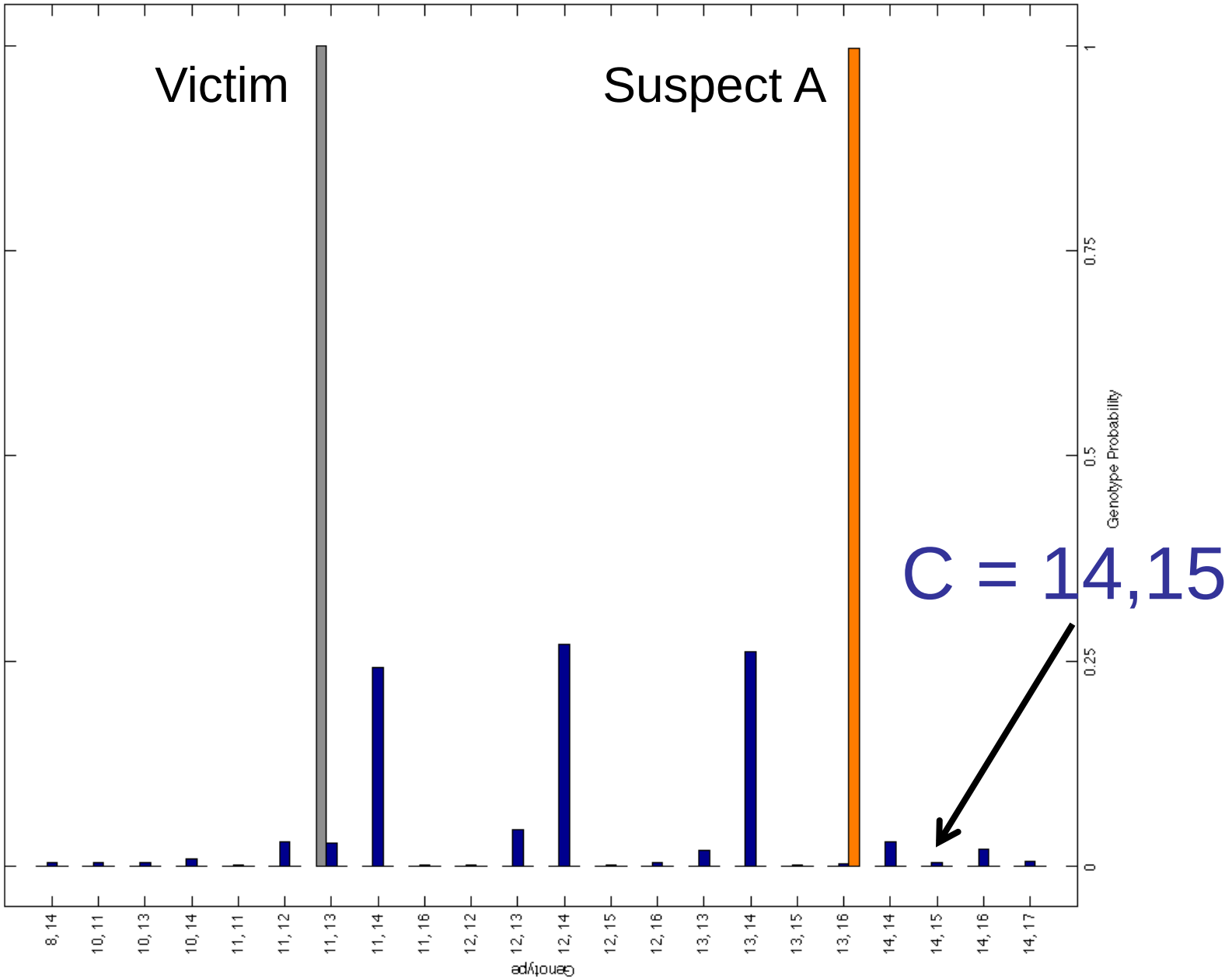
Contributor B (gray)
(16%)

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(LR).

Summary

- We need to move away from the interpretation of mixtures from an “allele-centric” point of view.
- Methods to incorporate probability will be necessary as we make this transition and confront the issues of low-level profiles with drop-out.

- “Just as logic is reasoning applied to truth and falsity, probability is reasoning with uncertainty”

-Dennis Lindley

Summary

- The LR is a method to evaluate evidence that can overcome many of the limitations we are facing today. ISFG Recommendations for incorporating drop-out are in press.

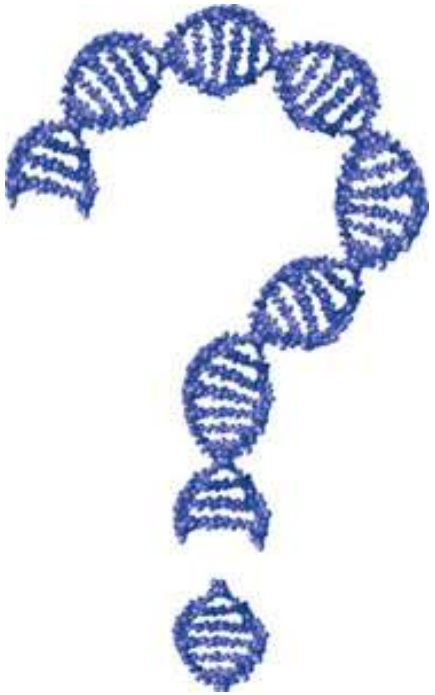
- This will require (obviously) software solutions… however, we need to better understand and be able to explain the statistics as a community.

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