

2012 Mixture Interpretation Workshop:

Mixtures Using *SOUND* Statistics, Interpretation, & Conclusions



Impact of Changing Thresholds on Data Interpretation - Statistics

Michael D. Coble

October 15, 2012

Nashville, TN



Statistics

A Tragedy in 400 Quadrillion Acts



“Though this be madness,
yet there is method in't.”

— William Shakespeare, *Hamlet*

Stats Required for Inclusions

SWGDM Interpretation Guideline 4.1:

“The laboratory **must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis.”**

Buckleton & Curran (2008): “There is a considerable aura to DNA evidence. Because of this aura **it is vital that weak evidence is correctly represented as weak or not presented at all.**”

Buckleton, J. and Curran, J. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2: 343-348.

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)

“Allele-centric”

“Inferred Genotype” Approach

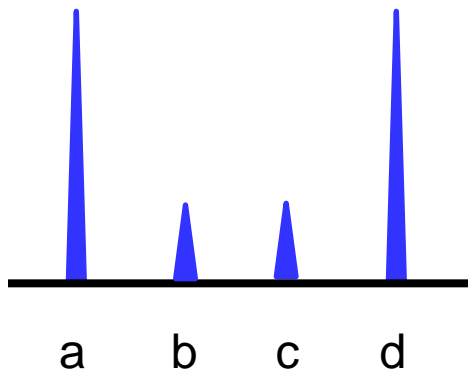
Random Match Probability
[modified]
(mRMP)

Likelihood Ratio
(LR)

“Genotype-centric”

Statistical Approaches with Mixtures

- **Random Man Not Excluded (CPE/CPI)** - The probability that a random person (unrelated individual) would be excluded as a contributor to the observed DNA mixture.



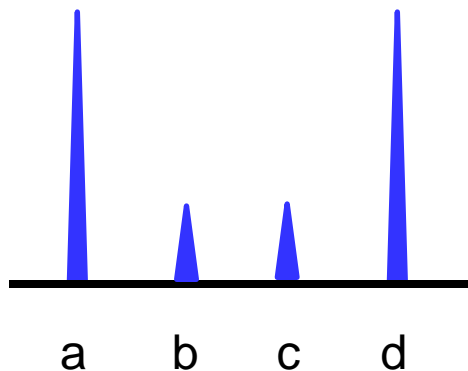
$$PI = (f(a) + f(b) + f(c) + f(d))^2$$

$$CPI = PI_{M1} \times PI_{M2} \dots$$

$$CPE = 1 - CPI$$

Statistical Approaches with Mixtures

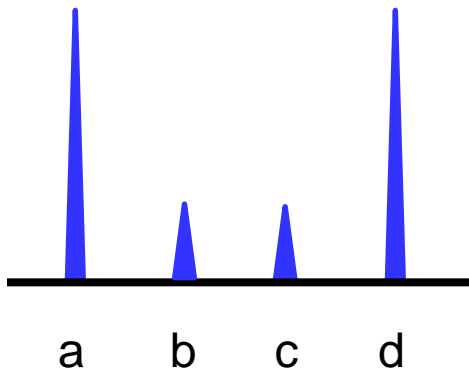
- **modified Random Match Probability (mRMP)**
 - The major and minor components can be successfully separated into individual profiles. A random match probability is calculated on the evidence as if the component was from a single source sample.



$$\begin{aligned} \text{mRMP}_{\text{major}} &= 2pq \\ &= 2f(a)f(d) \end{aligned}$$

Statistical Approaches with Mixtures

- **Likelihood Ratio** - Comparing the probability of observing the mixture data under two (or more) alternative hypotheses; in its simplest form $LR = 1/RMP$



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

E = Evidence

H_1 = Prosecutor's Hypothesis
(the suspect did it) = 1

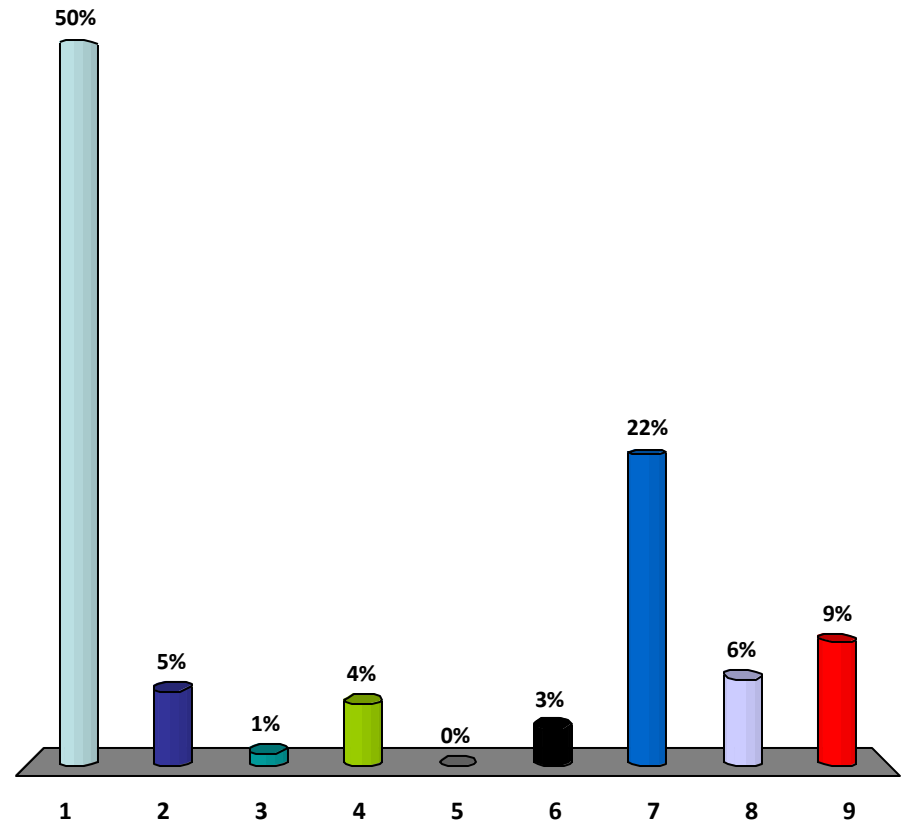
H_2 = Defense Hypothesis
(the suspect is an unknown,
random person)

$$= \frac{1}{P(E | H_2)} = \frac{1}{2pq} = 1/RMP$$

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

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

Data from 115 responses
ISHI Mixture Workshop (Oct 2012)



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.

Curran and Buckleton (2010)



PAPER

CRIMINALISTICS; GENERAL

James M. Curran,¹ M.Sc.(Hons.), Ph.D. and John Buckleton,² Ph.D.

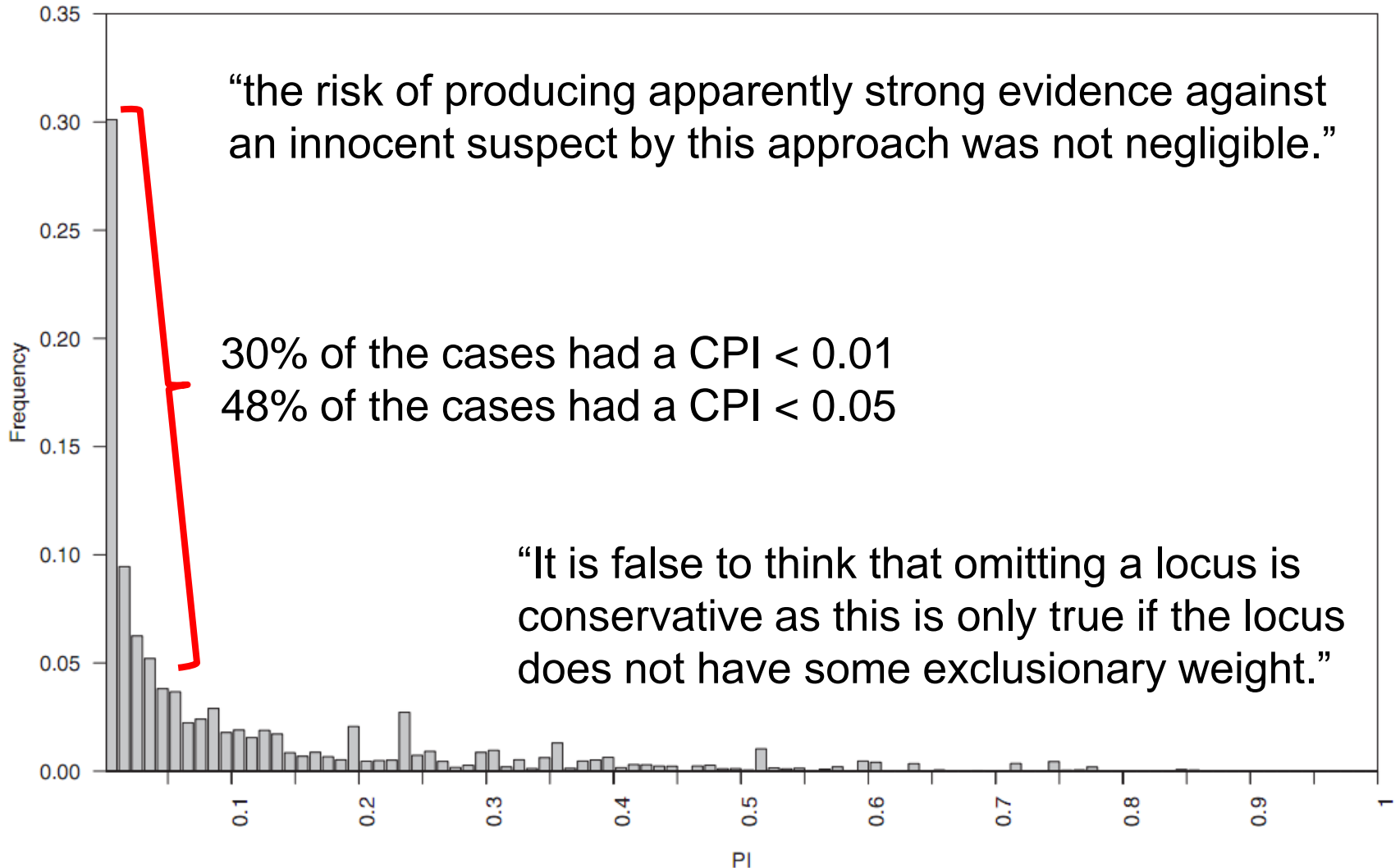
Inclusion Probabilities and Dropout

Created 1000 Two-person Mixtures (Budowle *et al.* 1999 AfAm freq.).

Created 10,000 “third person” genotypes.

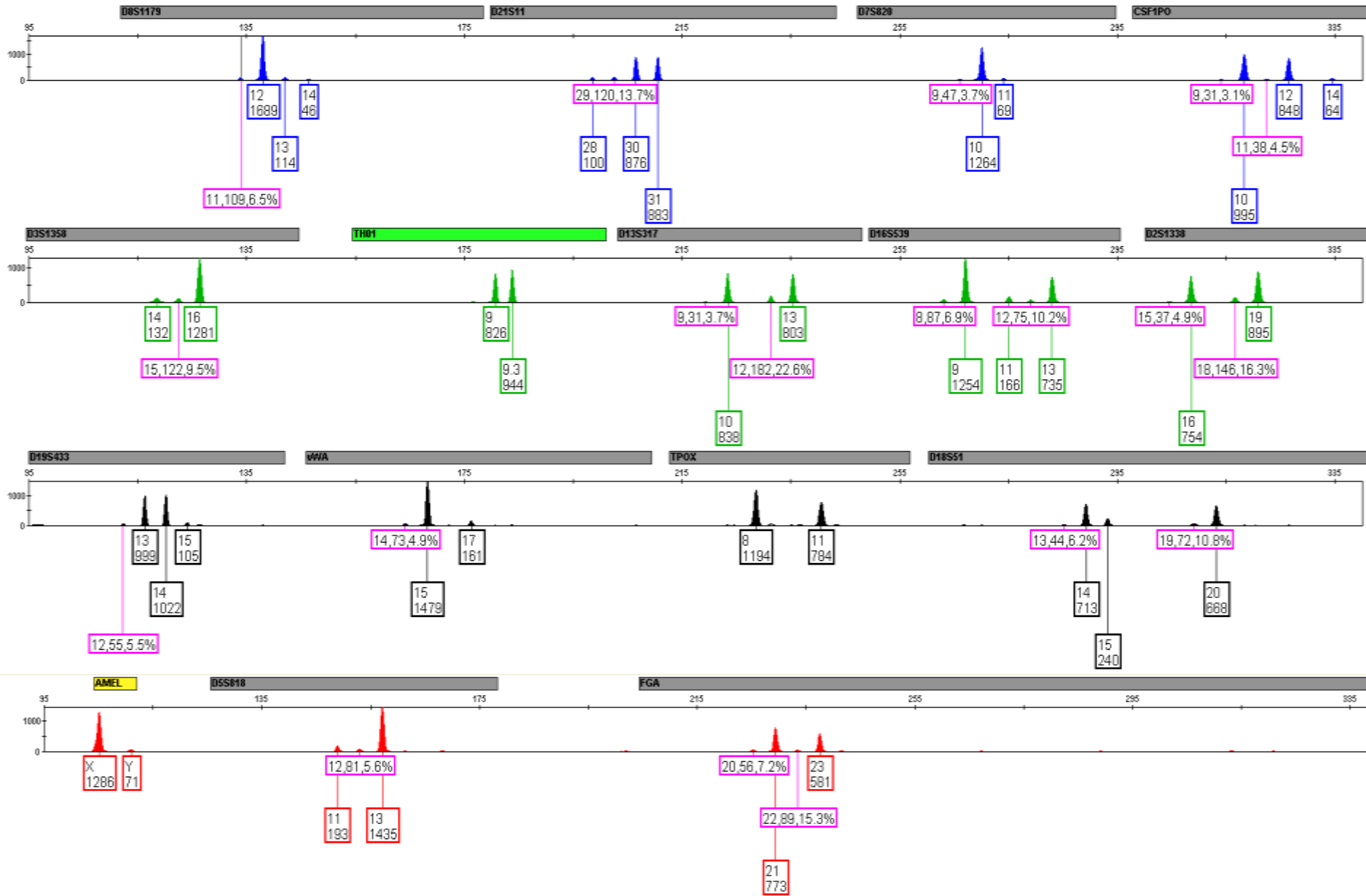
Compared “third person” to mixture data, calculated PI for included loci, ignored discordant alleles.

Curran and Buckleton (2010)



2-person Mixture





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 with alleles below the stochastic threshold cannot be used in the CPI statistic.

If CPI/CPE Stats are Used (ST = 150 RFU)

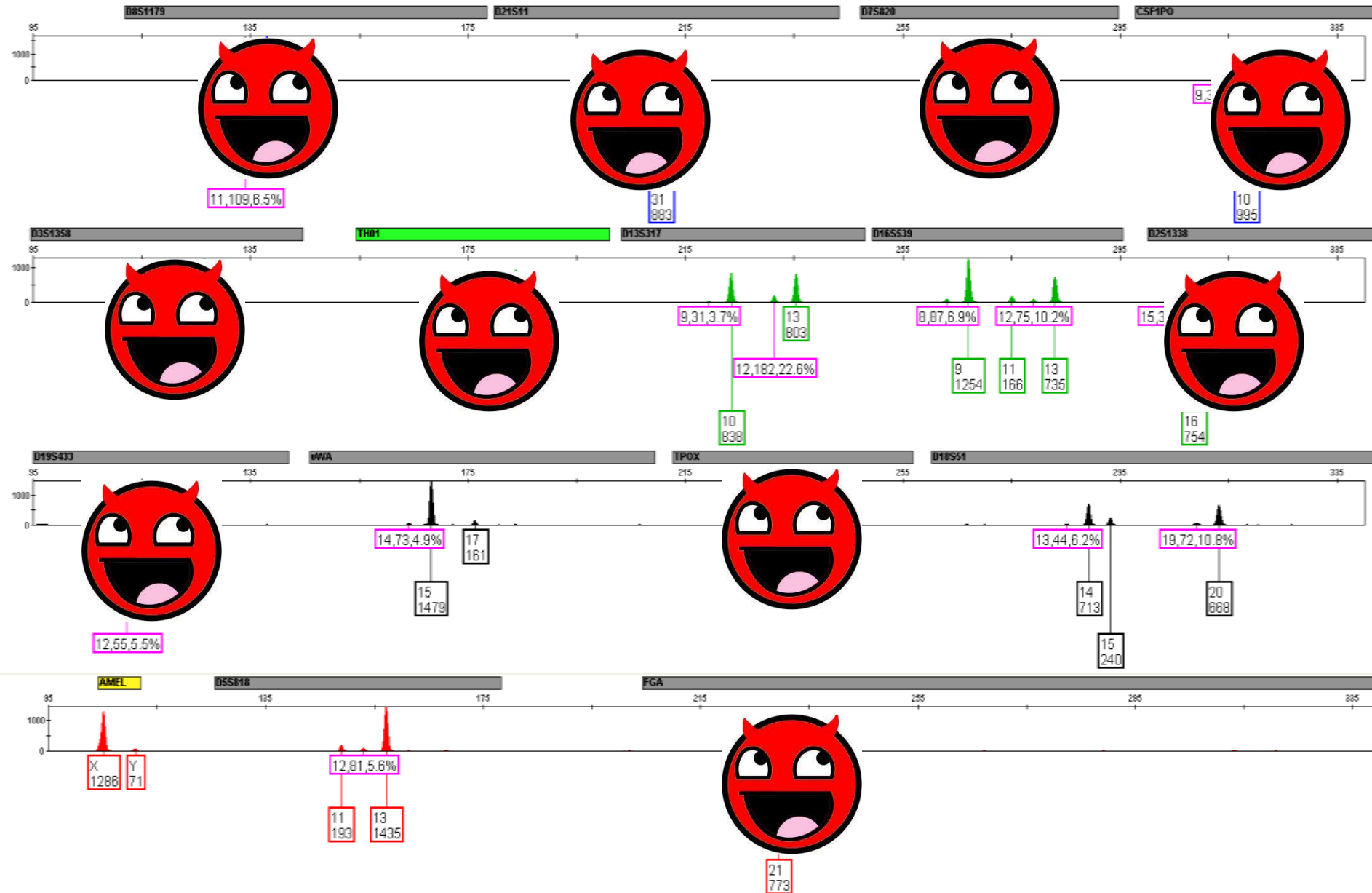


Shakespeare on Allelic Drop-Out

“Hell is empty and all the devils are here.”
— William Shakespeare, *The Tempest*



If CPI/CPE Stats are Used



If CPI/CPE Stats are Used

Can use

D13

D16

D18

vWA

D5

Cannot use

D8

D2

D7

D21

CSF

D3

TH01

D19

TPOX

FGA

Impact: discarding 2/3 of the data

If CPI/CPE Stats are Used

- CPI statistics using Caucasian Allele Frequencies
- 1 in 109 Caucasians included
- 99.09% Caucasians excluded

If CPI/CPE Stats are Used (ST = 120 RFU)

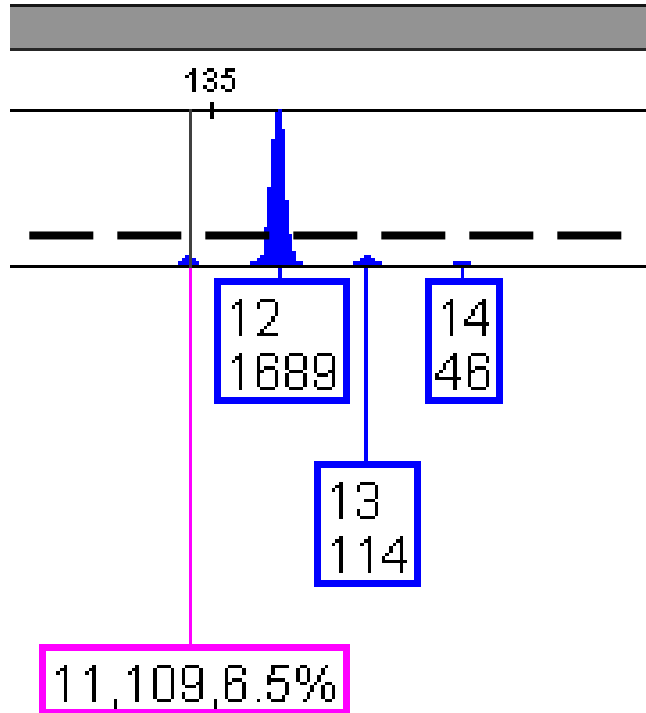


The impact of changing thresholds

If mRMP/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.

mRMP – D8S1179



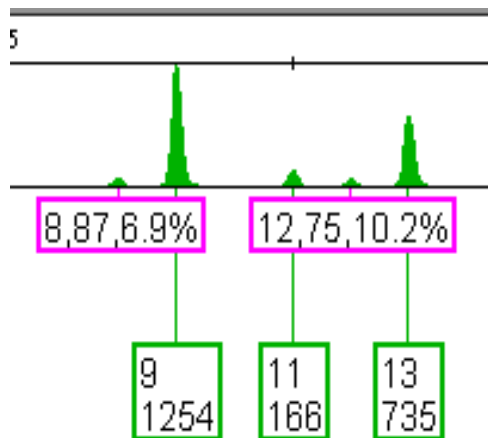
$$\begin{aligned}
 \text{mRMP}_{\text{minor}} &= 2pq \\
 &= 2f(13)f(14) \\
 &= 0.117 \quad \text{or 1 in 8.6}
 \end{aligned}$$

(LR = 8.6)

Possible genotype combinations if 11 stutter, no DO (12,13,14)

Person 1	Person 2
12,12	13,14
12,13	12,14 or 13,14 or 14,14
12,14	12,13 or 13,13 or 13,14
13,13	12,14
13,14	12,12 or 12,13 or 12,14
14,14	12,13

mRMP – D16S539



Possible genotype combinations if 8 contained allele (8,9,11,13)

Person 1	Person 2
8,9	11,13
8,11	9,13
8,13	9,11
9,11	8,13
9,13	8,11
11,13	8,9

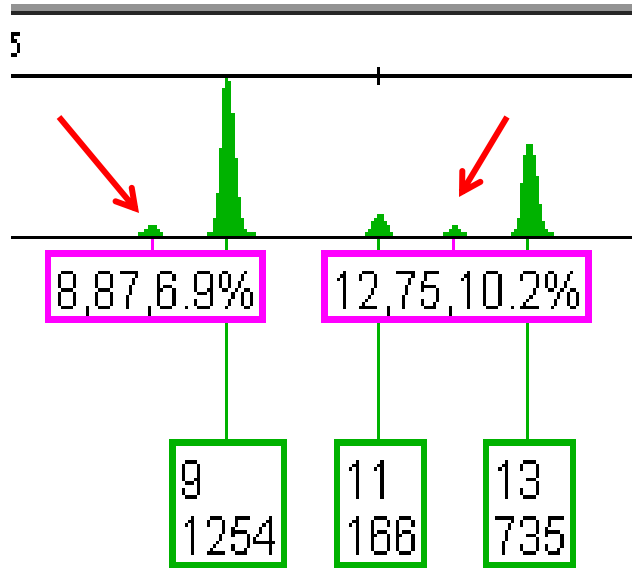
Possible genotype combinations if 12 contained allele (9,11,12,13)

Person 1	Person 2
9,11	12,13
9,12	11,13
9,13	11,12
11,12	9,13
11,13	9,12
12,13	9,11

Possible genotype combinations if 8 and 12 were stutter (9,11,13)

Person 1	Person 2
9,9	11,13
9,11	9,13 or 11,13 or 13,13
9,13	9,11 or 11,11 or 11,13
11,11	9,13
11,13	9,9 or 9,11 or 9,13
13,13	9,11

mRMP – D16S539



Description	D16S539
Inferred Genotypes of Minor	8,11 or 9,11 or 11,11 or 11,12 or 11,13

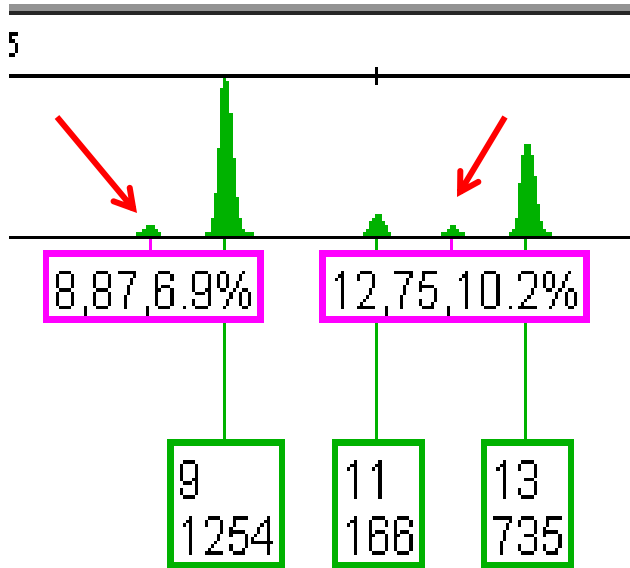
$$\text{mRMP}_{\text{minor}} = 2pq + 2pq + p^2 + 2pq + 2pq$$

$$= 2f(8)f(11) + 2f(9)f(11) + f(11)^2 + 2f(11)f(12) + 2f(11)f(13)$$

$$= 0.426 \quad \text{or 1 in 2.34}$$

$$(\text{LR} = 2.34)$$

mRMP – D16S539

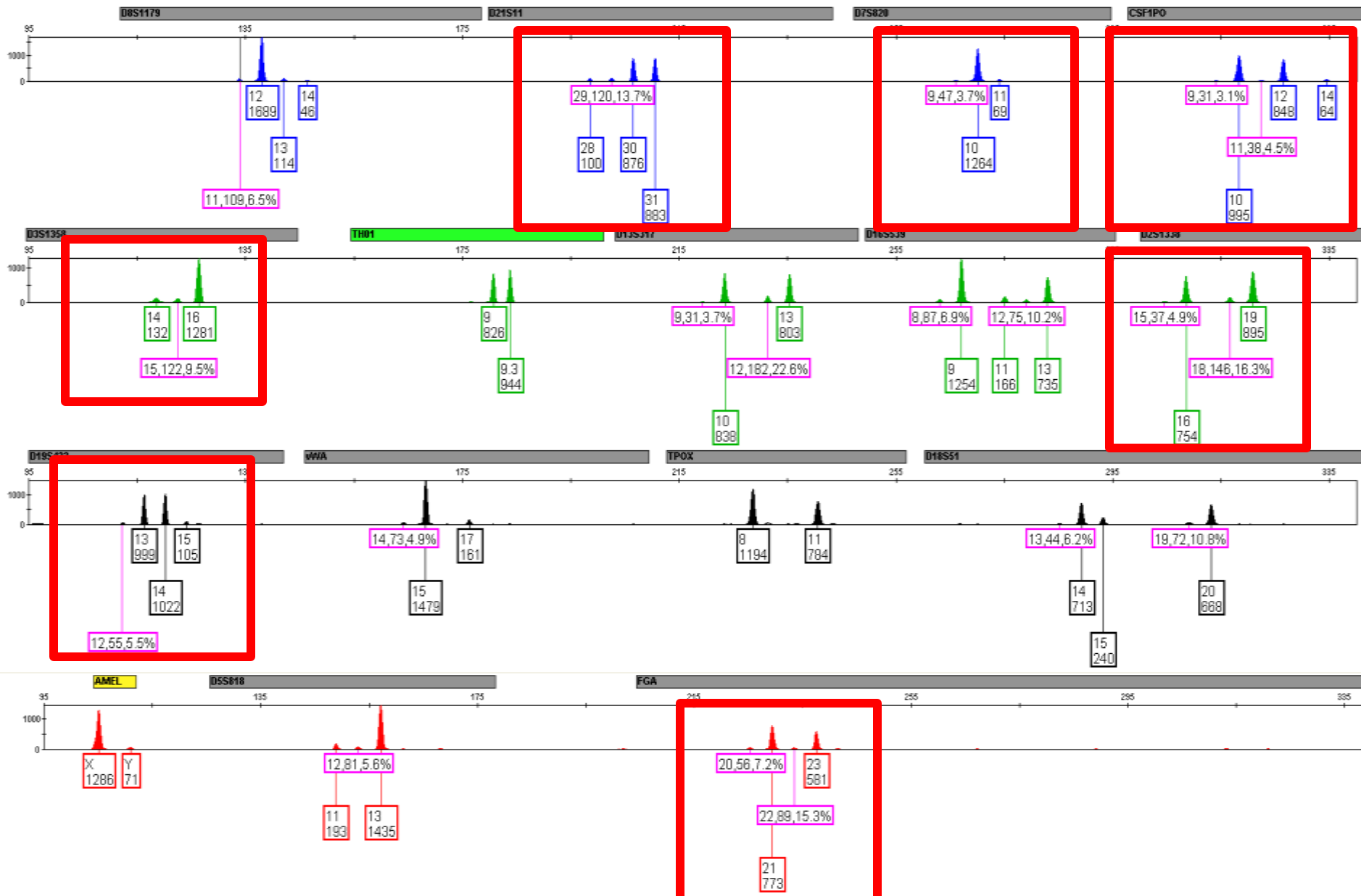


IF – we assume the 8 and 12 alleles are stutter – then we have 3 possible genotype combinations...

9,11 or 11,11 or 11,13

$$\begin{aligned} \text{mRMP}_{\text{minor}} &= 2pq + p^2 + 2pq \\ &= 2f(9)f(11) + f(11)^2 + 2f(11)f(13) \\ &= 0.276 \quad \text{or } 1 \text{ in } 3.62 \quad (\text{LR} = 3.62) \end{aligned}$$

Potential for Drop-out



If mRMP/LR Stats are Used

Can use

D8

D18

vWA

D13

D5

D16

Loci with potential D-out

D21

D7

CSF

D3

D2

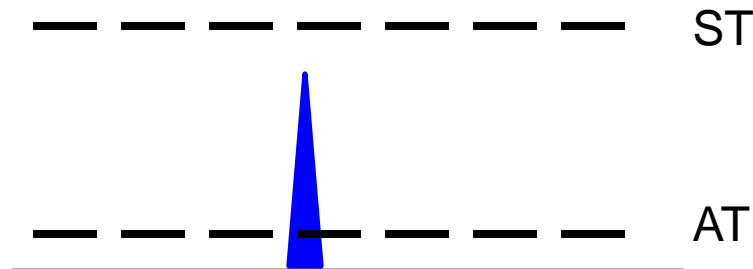
D19

FGA

TPOX and TH01 – used for exclusionary comparisons

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?



“2p” or not “2p”... That is the question.

Shakespeare on “2p”



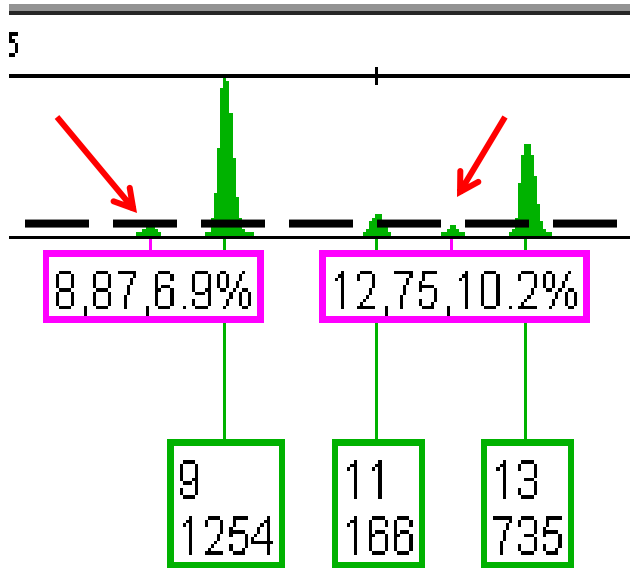
“Drink sir, is a great provoker of three things....
nose painting, sleep and urine.”

— William Shakespeare, *Macbeth*

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.

Macbeth/Duncan Profile – D16



ST = 150

(we did this)

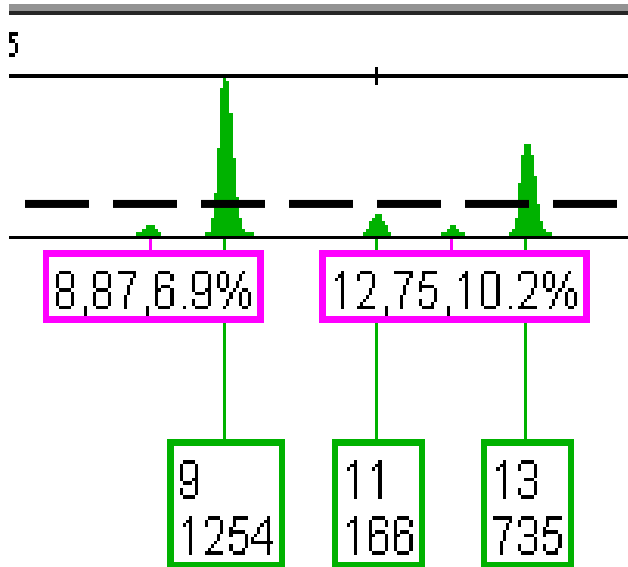
$$\text{mRMP}_{\text{minor}} = 2pq + 2pq + p^2 + 2pq + 2pq$$

$$= 2f(8)f(11) + 2f(9)f(11) + f(11)^2 + 2f(11)f(12) + 2f(11)f(13)$$

$$= 0.426 \quad \text{or 1 in 2.34}$$

$$(\text{LR} = 2.34)$$

Macbeth/Duncan Profile – D16



ST = 200

(2p invoked)

$$\text{mRMP}_{\text{minor}} = p^2 + 2p(1-p)$$

$$= f(11)^2 + 2f(11)x(1 - f(11))$$

$$= 0.500 \quad \text{or} \quad 1 \text{ in } 2$$

(LR = 2.0)

Macbeth/Duncan Profile – D16

	<u>LR</u>
ST = 200 (2p is used)	2.0
ST = 150 (2pq is used)	2.3

2p is conservative...

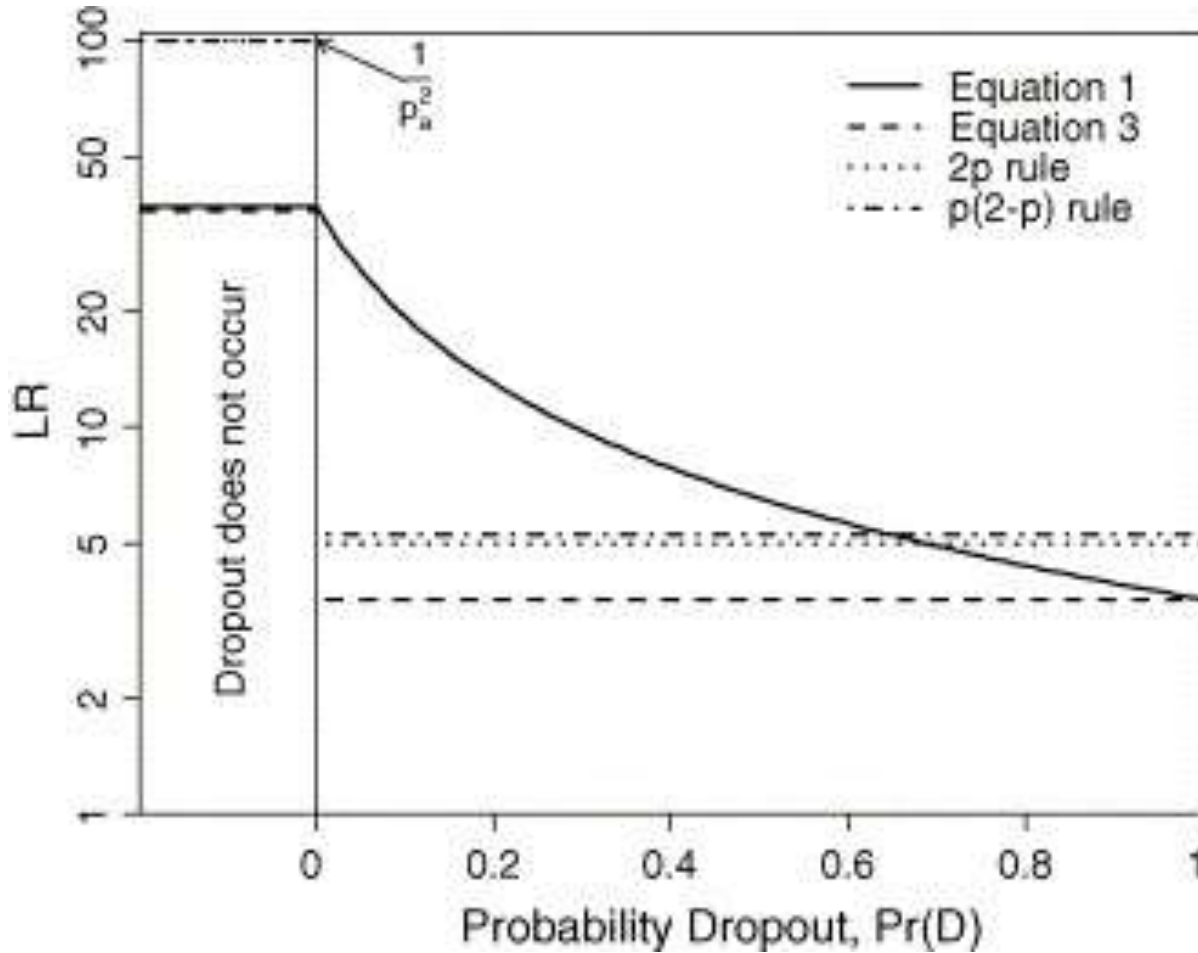
The “2p” Rule

- “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)

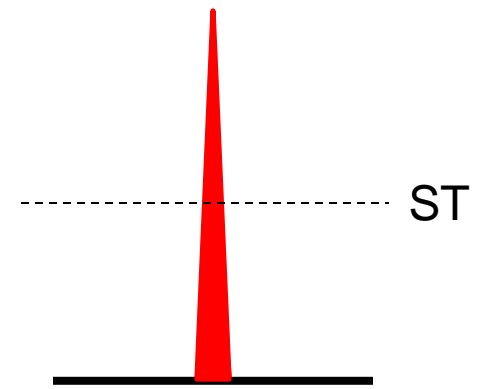
Is the $2p$ rule always conservative?”

The “2p” Rule



Stain = AA

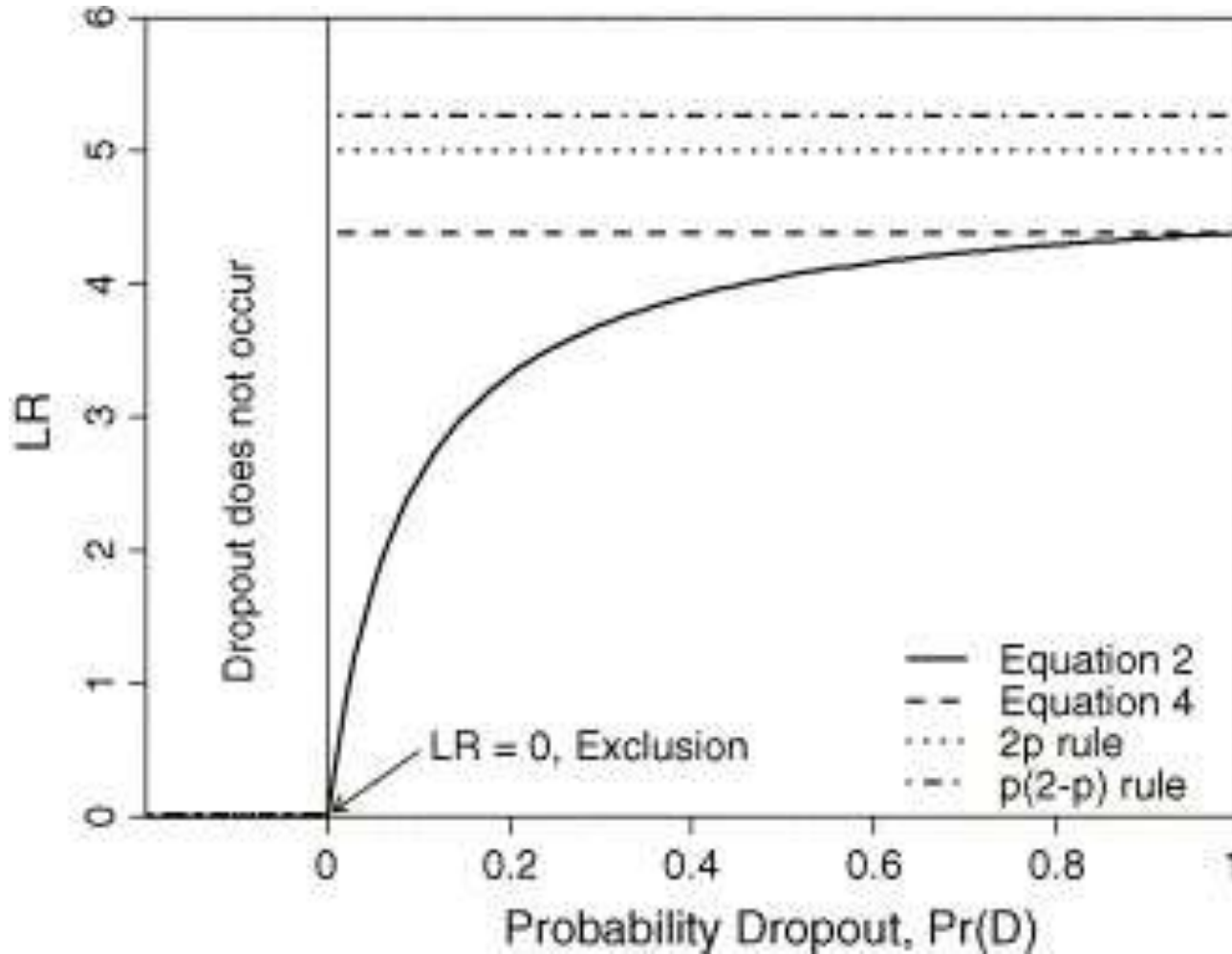
Suspect = AA



LR = 100

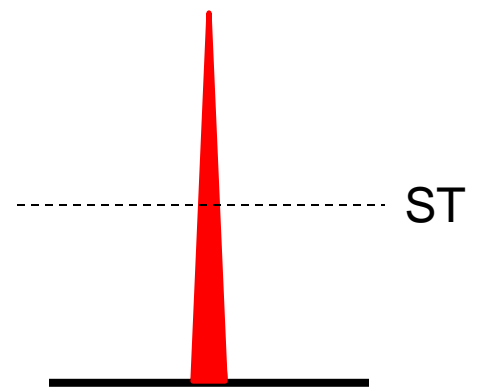
$$f(a) = 0.10 \quad 1/p^2 = 100 \quad 1/2p = 5$$

The "2p" Rule



Stain = AA

Suspect = AB



Exclusion

$$f(a) = 0.10 \quad 1/2p = 5$$

Is there a way forward?

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.”

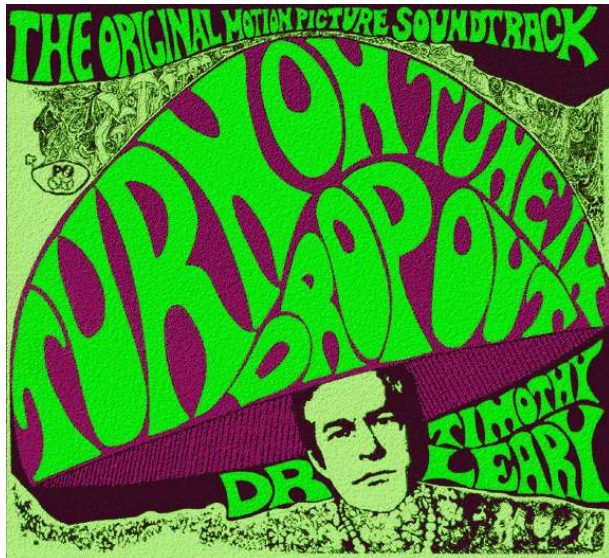
Psychedelic Mixtures

Turn On...



Tune In...

(Talk about) Drop Out



Article in press...



Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

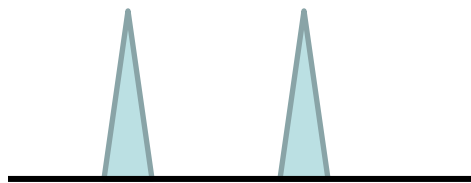
journal homepage: www.elsevier.com/locate/fsig



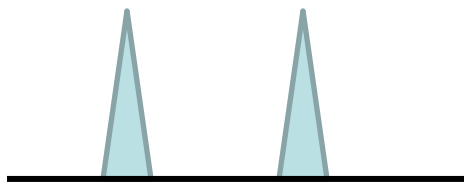
DNA commission of the International Society of Forensic Genetics:
Recommendations on the evaluation of STR typing results that may
include drop-out and/or drop-in using probabilistic methods

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

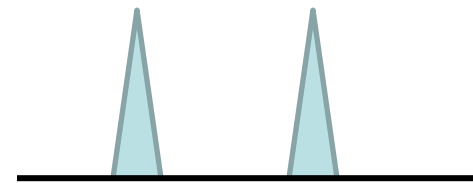
Suspect



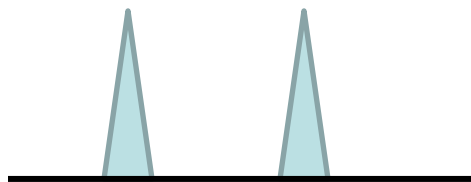
Suspect



Suspect



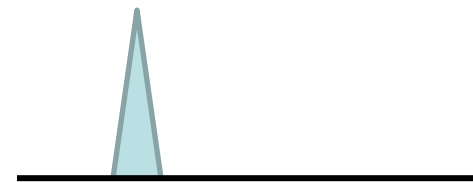
Evidence



Evidence



Evidence



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

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

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

“2p”

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

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



-Dennis Lindley

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



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.”



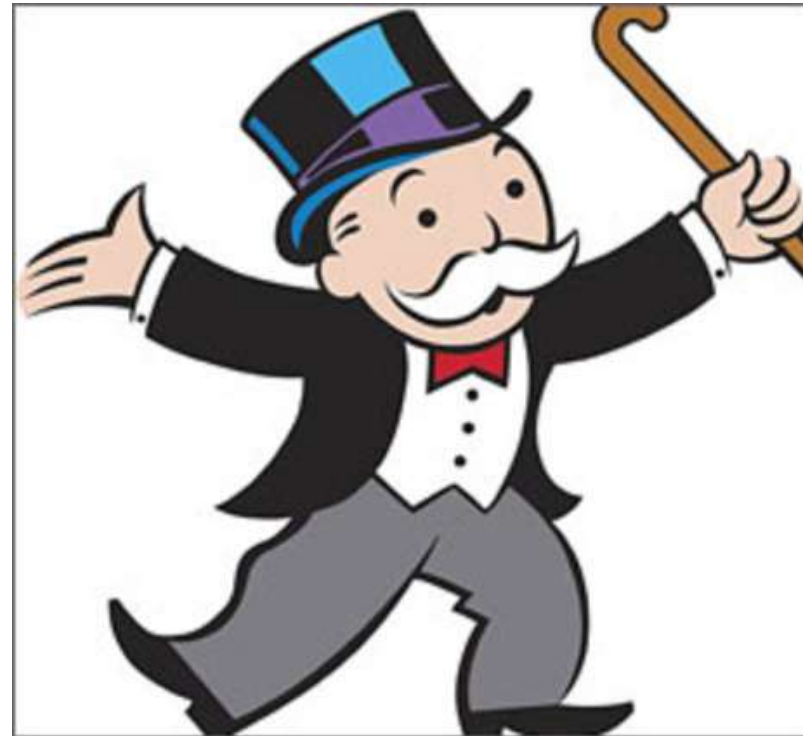
Andrey Markov

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

Is Blackjack a Markov Chain?



Monopoly is a Markov Chain

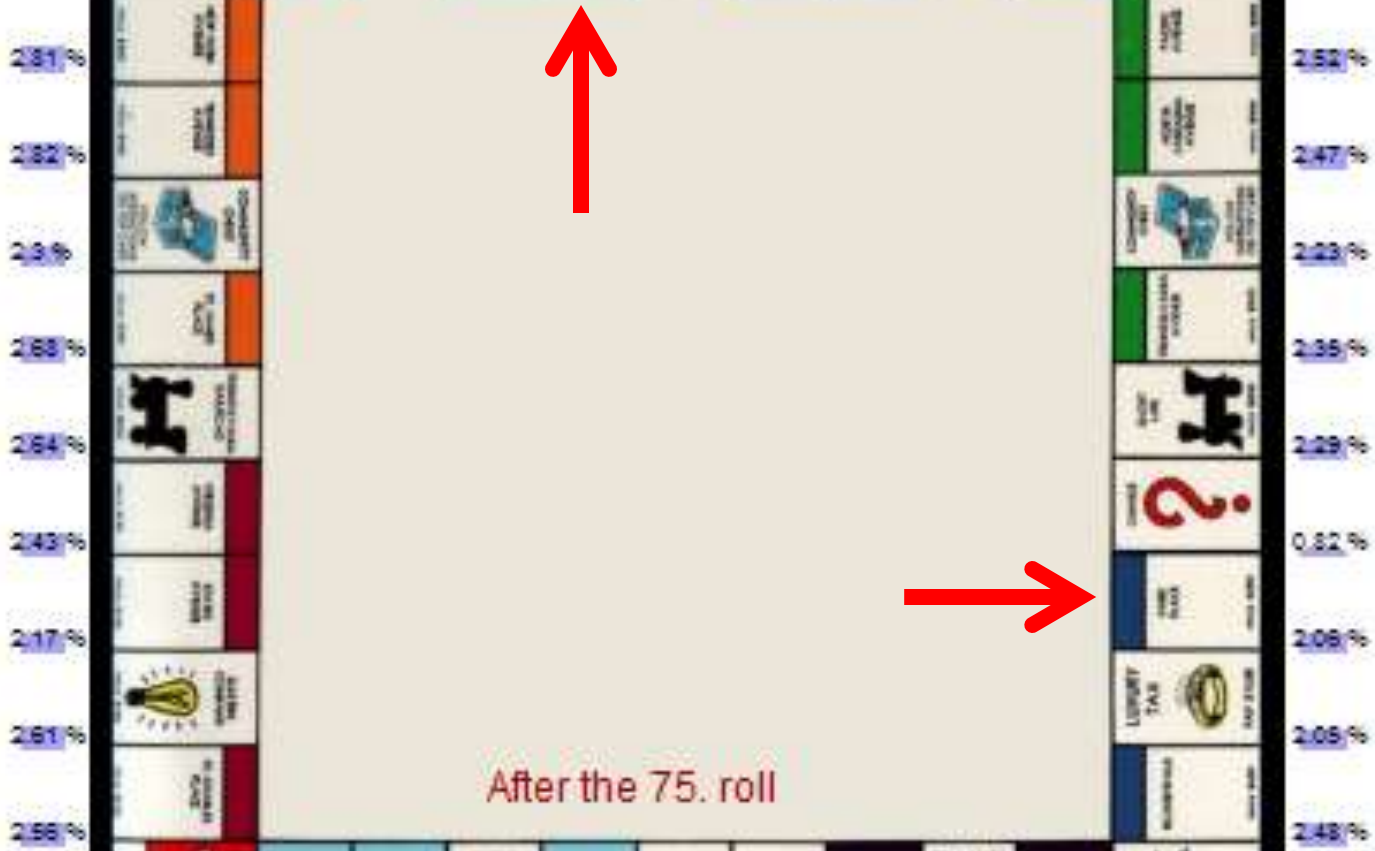


Monopoly simulation

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

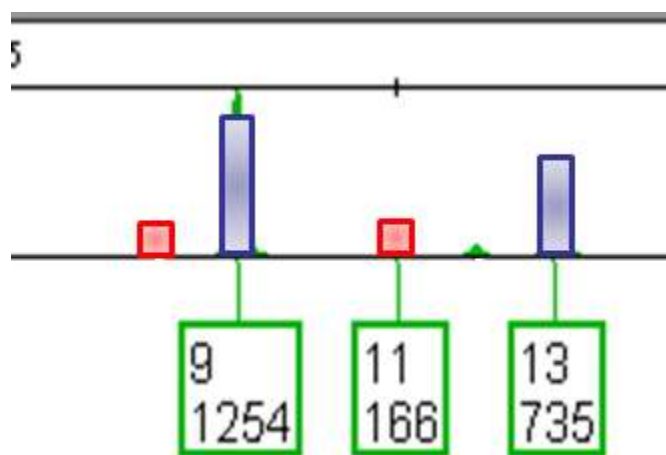
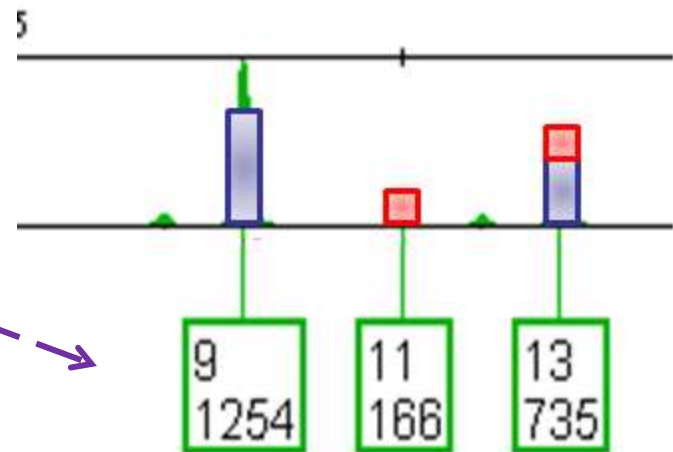
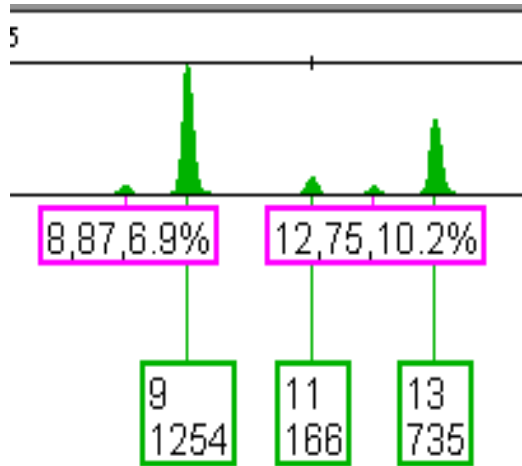
2.83% 2.61% 1.04% 2.57% 2.99% 2.89% 2.54% 2.52% 2.65% 2.44%

Higher Prob.
of being in jail



9.46%
in jail

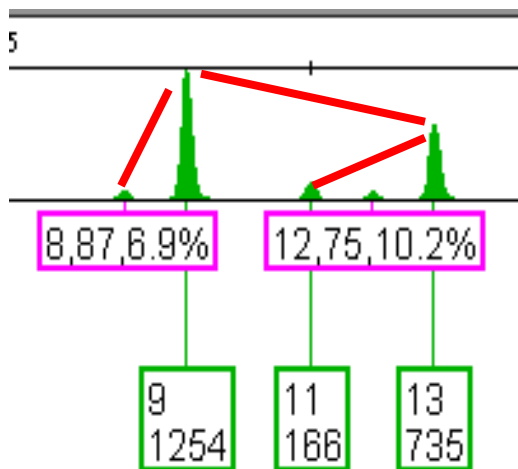
2.14% just vis. 2.17% 2.19% 0.82% 2.19% 2.8% 2.19% 2.04% 1.77% 2.01% 2.91%



Probabilistic Modeling of TA

Mathematical Modeling
of the Data

PHR, Mix Ratio, Stutter etc...



50-100,000
Simulations

(MCMC)

Probable **Genotypes**
to explain the mixture

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%

Summary of the Issues

- 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 of the Issues

- The LR is a method to evaluate evidence that can overcome many of the limitations we are facing today. ISFG Recommendations 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.
- “But, for my own part, it was Greek to me”
— William Shakespeare, *Julius Caesar*
- “We know what we are, but know not what we may be.” — William Shakespeare, *Hamlet*

Summary of the Issues

- Extensive training will be necessary – and a single 8 hour workshop will once a year will not suffice.

- “Do, or do not. There is no try.”
— Yoda



Thank You

- “I can no other answer make but thanks, and thanks.” - William Shakespeare, *Twelfth Night*



http://es.wikipedia.org/wiki/William_Shakespeare

Shakespeare and Forensics

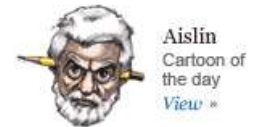
Shakespeare's Remains to be Tested for Marijuana

Tuesday, 28 Jun 2011 01:56 PM

Share: [f](#) [t](#) [g+](#) More...

[A A](#) | [Email Us](#) | [Print](#) | [Forward Article](#)

The Gazette
montrealgazette.com



[HOME](#) [NEWS](#) [OPINION](#) [BUSINESS](#) [SPORTS](#) [ARTS](#) [LIFE](#) [HEALTH](#) [TECHNOLOGY](#) [T](#)

‘Stunning’ find in the search for grave of king Richard III; Canadian descendant to have DNA tested for proof

LONDON DAILY TELEGRAPH SEPTEMBER 11, 2012

Thanks to NIJ for Support of BU and NIST



- **NIJ** Forensic Science Training Development and Delivery Program Grant # 2008-DN-BX-K158, awarded to Biomedical Forensic Science Program at **Boston University** School of Medicine
- NIJ has an Interagency Agreement (IAA) with the NIST Office of Law Enforcement Standards (OLES)