Training for Capital Litigators
Forensic DNA

Michael Coble, PhD
National Institute of Standards and Technology

August 27, 2012
Biological “Artifacts” of STR Markers

- Stutter Products
- Non-template nucleotide addition
- Microvariants
- Tri-allelic patterns
- Null alleles
- Mutations
Stutter Products

• Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis

• Stutter is less pronounced with larger repeat unit sizes (dinucleotides > tri- > tetra- > penta-)

• Longer repeat regions generate more stutter

• Each successive stutter product is less intense (allele > repeat-1 > repeat-2)

• Stutter peaks make mixture analysis more difficult
STR Alleles with Stutter Products

Slipped Strand Mispairing Model

Step 1

Taq DNA Polymerase has extended through 4 repeat units
Step 2

Taq has fallen off allowing the two strands to breathe apart.
When the two strands re-anneal the template (bottom) strand has looped out and the extending strand aligns out-of-register by one repeat unit.
Slipped Strand Mispairing Model

Step 4

The newly completed strand contains only 7 repeat units, while the template strand has the original 8 repeat units.
Stutter Product Formation

Repeat unit bulges out when strand breathing occurs during replication

True allele
(tetranucleotide repeat)

Typically 5-15% of true allele in tetranucleotide repeats STR loci

Deletion caused by slippage on the copied (bottom) strand

Insertion caused by slippage of the copying (top) strand

Occurs less frequently (typically <2%) – often down in the “noise” depending on sensitivity

Typically 5-15% of true allele in tetranucleotide repeats STR loci
Statistical Calculations for Forensic STR Testing
Laws of Heredity

- Gregor Mendel – 1850s developed the laws of heredity by studying pea plants
Laws of Heredity

Parents (P1)

Yellow seeds X Green seeds

1\textsuperscript{st} Filial (F1)

Yellow seeds X Yellow seeds

2\textsuperscript{nd} Filial (F2)

Yellow seeds Green seeds

3:1 ratio
### Population Genetics

**Father Gametes** (sperm)
- Mother Gametes (egg)
- Freq (A) = \( p \)
- Freq (a) = \( q \)
- \( p + q = 1 \)

<table>
<thead>
<tr>
<th>Father gametes (sperm)</th>
<th>A</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>AA</td>
<td>aA</td>
</tr>
<tr>
<td>a</td>
<td>Aa</td>
<td>aa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p^2 )</td>
<td>( 2pq )</td>
<td>( q^2 )</td>
</tr>
</tbody>
</table>

*Punnett square*
Hardy-Weinberg Equilibrium (HWE) – a way to relate allele frequencies to genotype frequencies

HWE assumes: Large population, No Selection, No mutation, No immigration, No emigration, random mating.

None of these assumptions are really true…
To Determine the Rarity of a Type

Hardy-Weinberg Equilibrium

\[ p^2 + 2pq + q^2 \]

Heterozygote = 2pq
Homozygote = \( p^2 \)
Heterozygote = 2pq

Suppose - Suspect is 16, 17

\[ P(16, 17) = 2pq \]
\[ = 2(0.2533)(0.2152) \]
\[ = 0.109 \quad \text{(or 1 in 9.17)} \]

Homozygote = \( p^2 \)

Suppose - Suspect is 15, 15

\[ P(15, 15) = p^2 \]
\[ =(0.2616)^2 \]
\[ = 0.068 \quad \text{(or 1 in 14.6)} \]
Population is: FBI Caucasian
Theta is: 0.01 Default Frequency: 0.0129

<table>
<thead>
<tr>
<th>Locus</th>
<th>allele</th>
<th>value</th>
<th>allele</th>
<th>value</th>
<th>frequency, 1 in</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>17.0</td>
<td>0.2118</td>
<td>18.0</td>
<td>0.1626</td>
<td>14.52</td>
</tr>
<tr>
<td>VWA</td>
<td>16.0</td>
<td>0.2015</td>
<td></td>
<td></td>
<td>23.69</td>
</tr>
<tr>
<td>FGA</td>
<td>18.0</td>
<td>0.0306</td>
<td>22.0</td>
<td>0.1888</td>
<td>86.55</td>
</tr>
<tr>
<td>D8S1179</td>
<td>14.0</td>
<td>0.2015</td>
<td></td>
<td></td>
<td>23.69</td>
</tr>
<tr>
<td>D21S11</td>
<td>30.2</td>
<td>0.0383</td>
<td>32.2</td>
<td>0.1122</td>
<td>116.35</td>
</tr>
<tr>
<td>D18S51</td>
<td>17.0</td>
<td>0.1556</td>
<td>19.0</td>
<td>0.0357</td>
<td>90.01</td>
</tr>
<tr>
<td>D5S818</td>
<td>11.0</td>
<td>0.4103</td>
<td>12.0</td>
<td>0.3539</td>
<td>3.44</td>
</tr>
<tr>
<td>D13S317</td>
<td>10.0</td>
<td>0.0510</td>
<td>14.0</td>
<td>0.0357</td>
<td>274.62</td>
</tr>
<tr>
<td>D7S820</td>
<td>9.0</td>
<td>0.1478</td>
<td>11.0</td>
<td>0.2020</td>
<td>16.75</td>
</tr>
<tr>
<td>D16S539</td>
<td>13.0</td>
<td>0.1634</td>
<td>14.0</td>
<td>0.0322</td>
<td>95.03</td>
</tr>
<tr>
<td>THO1</td>
<td>6.0</td>
<td>0.2266</td>
<td>9.3</td>
<td>0.3054</td>
<td>7.23</td>
</tr>
<tr>
<td>TPOX</td>
<td>8.0</td>
<td>0.5443</td>
<td></td>
<td></td>
<td>3.35</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>10.0</td>
<td>0.2537</td>
<td>13.0</td>
<td>0.0714</td>
<td>27.60</td>
</tr>
</tbody>
</table>

The Random Match Probability for this profile in the FBI Caucasian population is one in 7.42E18

1 in 7.42 quintillion
DNA Mixture Interpretation
April 14, 2005

“If you show 10 colleagues a mixture, you will probably end up with 10 different answers.”

- Dr. Peter Gill
International Society of Forensic Genetics

http://www.isfg.org/

• An international organization responsible for the promotion of scientific knowledge in the field of genetic markers analyzed with forensic purposes.

• Founded in 1968 and represents more than 1100 members from over 60 countries.

• A DNA Commission regularly offers recommendations on forensic genetic analysis.
Available for download from the ISFG Website: http://www.isfg.org/Publication;Gill2006

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

“…These recommendations have been written to serve two purposes: to define a generally acceptable mathematical approach for typical mixture scenarios and to address open questions where practical and generally accepted solutions do not yet exist. This has been done to stimulate the discussion among scientists in this field. The aim is to invite proposals and criticism in the form of comments and letters to the editors of this journal…We are hoping to continue the process to allow the DNA Commission to critically revise or extend these recommendations in due time…”
German Mixture Classification Scheme


(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

---

**SWGDAM**

“Indistinguishable”

“Distinguishable”

“Uninterpretable”
Overview of the SWGDAM 2010 Interp Guidelines

1. Preliminary evaluation of data – *is something a peak and is the analysis method working properly?*
2. Allele designation – *calling peaks as alleles*
3. Interpretation of DNA typing results – *using the allele information to make a determination about the sample*
   1. Non-allelic peaks
   2. Application of peak height thresholds to allelic peaks
   3. Peak height ratio
   4. Number of contributors to a DNA profile
   5. Interpretation of DNA typing results for mixed samples
   6. Comparison of DNA typing results
4. Statistical analysis of DNA typing results – *assessing the meaning (rarity) of a match*

Other supportive material: statistical formulae, references, and glossary
• “3.6.1. The laboratory must establish guidelines to ensure that, to the extent possible, DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors.”

- While the FBI QAS do not address this issue, this is an example of an issue felt by the committee members to be of such importance that it warranted a “must.”
Stats Required for Inclusions

SWGDAM 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.”

Steps in DNA Interpretation

Data Collection
- Sample Deposited
- Sample Collected
- Extraction
- Quantitation
- PCR
- Amplification
- CE
- Separation/Detection

Signal observed

Data Interpretation
- Peak (vs. noise)
- Allele (vs. artifact)
- Genotype (allele pairing)
- Profile (genotype combining)

All Alleles Detected?
- Genotype(s)
- 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** *(Cannot be confident dropout of a sister allele did not occur)*

**50 RFUs**

- **Peak not considered reliable**

---

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

**Stochastic Threshold**

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

**Analytical Threshold**

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

---

The Analytical Threshold
Stochastic Effects

- **Allele drop-out** is an extension of the amplification disparity that is observed when heterozygous peaks heights are unequal
  - Occurs in single-source samples and mixtures
  - Analyst is unable to distinguish complete allele drop-out in a true heterozygote from a homozygous state

![Diagram showing levels of allele drop-out]

- Slight
- Moderate
- Extreme
- No detectable amplification
What is Allele Drop Out?

• Scientifically
  – **Failure to detect** an allele within a sample or failure to amplify an allele during PCR. *From SWGDAM Guidelines, 2010*

  – Note that: Failure to detect ≠ failure to amplify

• Operationally
  – Setting a threshold(s) or creating a process, based on validation data and information in the literature, which allows assessment of the likelihood of drop-out of an allele or a locus.
Stochastic Effects with Low Levels of DNA When Combined with Higher Sensitivity Techniques

Loss of True Signal (False Negative)  Gain of False Signal (False Positive)

Severe Peak Imbalance  Allelic Drop-out  High Stutter  Allelic Drop-in

Identifiler, 30 pg DNA, 31 cycles  Identifiler, 30 pg DNA, 31 cycles  Identifiler, 10 pg DNA, 31 cycles  Identifiler, 10 pg DNA, 31 cycles

10,11  12,14  12,13  18,19

Correct genotype:

30% peak height ratio  14 allele drop-out  64% stutter  16 allele drop-in

Stochastic Effects with Low Levels of DNA When Combined with Higher Sensitivity Techniques

Loss of True Signal (False Negative)  Gain of False Signal (False Positive)

Severe Peak Imbalance  Allelic Drop-out  High Stutter  Allelic Drop-in

Identifiler, 30 pg DNA, 31 cycles  Identifiler, 30 pg DNA, 31 cycles  Identifiler, 10 pg DNA, 31 cycles  Identifiler, 10 pg DNA, 31 cycles

10,11  12,14  12,13  18,19

Correct genotype:
Heat Map Explanation

Results broken down by locus

Green = full (correct) type
Yellow = allele dropout
Red = locus dropout

This is an easy way to look at a lot of data at once
Stochastic Threshold

Identifiler: 28 cycles

Standard Injection on 3500: 7 sec @ 1.2 kV inj

n=84 Samples

Slide from Erica Butts (NIST) 3500 presentation in Innsbruck, Austria (Sept 5, 2011)
Stochastic Threshold Summary

• A stochastic threshold (ST) may be established for a specific set of conditions to reflect possibility of allele drop-out, which is essential for a CPE/CPI stats approach

• ST should be re-examined with different conditions (e.g., higher injection, sample desalting, increase in PCR cycles)

• ST will be dependent on the analytical threshold set with a method and impacts the lowest expected peak height ratio

• Assumptions of the number of contributors is key to correct application of ST
Two Parts to Mixture Interpretation

• Determination of alleles present in the evidence and deconvolution of mixture components where possible
  – Many times through comparison to victim and suspect profiles

• Providing some kind of statistical answer regarding the weight of the evidence
  – There are multiple approaches and philosophies

Software tools can help with one or both of these…
3.3. Peak Height Ratio

- Intra-locus peak height ratios (PHR) are calculated for a given locus by dividing the peak height of an allele with a lower RFU value by the peak height of an allele with a higher RFU value, and then multiplying this value by 100 to express the PHR as a percentage.

- Typically above 60%
Determination of Genotypes (PHR)

Possible Combinations

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

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

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

D18S51
Determination of Mixture Ratio

Major: 16, 18
Minor: 14, 20

Total of all peak heights
= 112 + 616 + 597 + 152
= 1477 RFUs

Minor component:
(“14”+”20”)/total = (112+152)/1477
= 0.179

Major component:
(“16”+”18”)/ total = (616+597)/1477
= 0.821

≈ 4.6 : 1
Challenges to Interpretation – Stutter
Interpretation of Potential Stutter Peaks in a Mixed Sample

3.5.8.1. For mixtures in which minor contributors are determined to be present, a peak in stutter position (generally n-4) may be determined to be 1) a stutter peak, 2) an allelic peak, or 3) indistinguishable as being either an allelic or stutter peak.
2 Person Mixture - D5S818

Stochastic Threshold
= 200 RFU

Victim – 12,12

Homozygote??

11 122
13 162
12 974
2 Person Mixture - D5S818

Stutter??
12.5%

Victim – 12,12
2 Person Mixture - D5S818

Possible Genotypes

13, 13
12, 13
11, 13
**ISFG Recommendation #6 Example**

**Likely a AA**  
(homozygote)

**Possibly AB**  
(heterozygote)

Could also be AC, AD, AA, or A,? (dropout)
Stutter??

Assumptions will matter!
Steps in DNA Interpretation

Data Collection
- Sample Deposited
- Sample Collected
- Extraction
- Quantitation
- PCR
- Amplification
- CE
- Separation/Detection

Signal observed

Data Interpretation

Peak (vs. noise)
Allele (vs. artifact)
Genotype (allele pairing)
Profile (genotype combining)

Peak
Allele
All Alleles Detected?
Genotype(s)
Contributor profile(s)

Comparison to Known(s)
Weight of Evidence (Stats)
Statistical Analyses with Mixtures

It doesn’t have to be a Shakespearean tragedy!
Statistics
A Tragedy in 400 Quadrillion Acts

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

— William Shakespeare, *Hamlet*
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.

\[
p = f(a) + f(b) + f(c) + f(d)
\]
\[
q = 1 - p
\]
\[
PE = 2pq + q^2
\]
\[
CPE = PE_{M1} \times PE_{M2} \cdots
\]
\[
CPI = 1 - CPE
\]
Statistical Approaches with Mixtures

- **Random Match Probability (RMP)** – 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.

\[
RMP_{\text{major}} = 2pq = 2 \times f(a) \times f(d)
\]
Statistical Approaches with Mixtures

- **Likelihood Ratio** - Comparing the probability of observing the mixture data under two (or more) alternative hypotheses
Conditioning

• **Probabilities are conditional**, which means that the probability of something is based on a hypothesis.

• In math terms, conditioning is denoted by a vertical bar.
  – Hence, $\Pr(a|b)$ means ‘the probability of $a$ given that $b$ is true’.

• The probability of an event $a$ is dependent upon various assumptions—and these assumptions or hypotheses can change…
Probability Example – Will It Rain? (1)

Defining the Event and Assumptions/Hypotheses

• Let’s suppose that \( a \) is the probability of an event (e.g., will it rain?)
• What is the probability that it will rain in the afternoon – \( \Pr(a) \)?

• This probability is dependent upon assumptions
  – We can look at the window in the morning and observe if it is sunny (s) or cloudy (c)
  – \( \Pr(a) \) if it is sunny (s) is less than \( \Pr(a) \) if it is cloudy (c)

• We can write this as \( \Pr(a/s) \) and \( \Pr(a/c) \)
  – Since sunny or cloudy are the only possibilities, \( \Pr(s) + \Pr(c) = 1 \)
  – or \( \Pr(s) = 1 - \Pr(c) \)

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)
Examining Available Data

- $Pr(a|s)$ and $Pr(a|c)$ can be calculated from data

- How often does it rain in the afternoon when it’s sunny in the morning?
  - 10 out of 100 observations so $Pr(a/s) = 0.1$

- How often does it rain in the afternoon when it is cloudy in the morning?
  - 90 out of 100 observations so $Pr(a/c) = 0.9$
Probability Example – Will It Rain? (3)

Formation of the Likelihood Ratio (LR)

• The LR compares two probabilities to find out which of the two probabilities is the most likely

The probability that it will rain in the afternoon when it is cloudy in the morning or $\Pr(a|c)$ is divided by the probability that it will rain in the afternoon when it is sunny in the morning or $\Pr(a|s)$

$$LR = \frac{\Pr(a|c)}{\Pr(a|s)} = \frac{0.9}{0.1} = 9$$
Explanation of the Likelihood Ratio

\[ LR = \frac{\Pr(a \mid c)}{\Pr(a \mid s)} = \frac{0.9}{0.1} = 9 \]

• The probability that it will rain is 9 times more likely \textit{if} it is cloudy in the morning than \textit{if} it is sunny in the morning.

• The word \textit{if} is very important here. It must always be used when explaining a likelihood ratio otherwise the explanation could be misleading.
Likelihood Ratios in Forensic DNA Work

- We evaluate the evidence \( (E) \) relative to alternative pairs of hypotheses

- Usually these hypotheses are formulated as follows:
  - The probability of the evidence if the crime stain originated with the suspect or \( \Pr(E|S) \)
  - The probability of the evidence if the crime stain originated from an unknown, unrelated individual or \( \Pr(E|U) \)

\[
LR = \frac{\Pr(E|S)}{\Pr(E|U)}
\]

The numerator

The denominator

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)
The Likelihood Ratio Must Be Stated Carefully

• The probability of the evidence is $x$ times more likely if the stain came from the suspect Mr. Smith than if it came from an unknown, unrelated individual.

• It is not appropriate to say: “The probability that the stain came from Mr. Smith.” because we must always include the conditioning statement – i.e., always make the hypothesis clear in the statement.

• Always use the word ‘if’ when using a likelihood ratio to avoid this trap.
Likelihood Ratio (LR)

• Provides ability to express and evaluate both the prosecution hypothesis, \( H_p \) (the suspect is the perpetrator) and the defense hypothesis, \( H_d \) (an unknown individual with a matching profile is the perpetrator)

\[
LR = \frac{H_p}{H_d}
\]

• The numerator, \( H_p \), is usually 1 – since in theory the prosecution would only prosecute the suspect if they are 100% certain he/she is the perpetrator

• The denominator, \( H_d \), is typically the profile frequency in a particular population (based on individual allele frequencies and assuming HWE) – i.e., the random match probability
Statistical Approaches with Mixtures

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

\[
\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{1}{P(E \mid H_2)} = \frac{1}{2pq} = 1/RMP
\]

- \( E \) = Evidence
- \( H_1 \) = Prosecutor’s Hypothesis (the suspect did it) = 1
- \( H_2 \) = Defense Hypothesis (the suspect is an unknown, random person)
Advantages and Disadvantages
RMNE and LR

RMNE (CPE/CPI)

Advantages
- Does not require an assumption of the number of contributors to a mixture
- Easier to explain in court

Disadvantages
- Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect’s genotype)
- LR approaches are developed within a consistent logical framework

Likelihood Ratios (LR)

Advantages
- Enables full use of the data including different suspects

Disadvantages
- More difficult to calculate (software programs can assist)
- More difficult to present in court

Summarized from John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223
4.6.3. When using CPE/CPI (with no assumptions of number of contributors) to calculate the probability that a randomly selected person would be excluded/included as a contributor to the mixture, loci with alleles below the stochastic threshold may not be used for statistical purposes to support an inclusion. In these instances, the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the interpreted alleles.
2-person Mixture
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 = 200 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

<table>
<thead>
<tr>
<th>Can use</th>
<th>Cannot use</th>
</tr>
</thead>
<tbody>
<tr>
<td>D21</td>
<td>D8</td>
</tr>
<tr>
<td>CSF</td>
<td>D7</td>
</tr>
<tr>
<td>D3</td>
<td>TH01</td>
</tr>
<tr>
<td>D19</td>
<td>D13</td>
</tr>
<tr>
<td>TPOX</td>
<td>D16</td>
</tr>
</tbody>
</table>

Impact: discarding 2/3 of the data
If CPI/CPE Stats are Used

- CPI statistics using FBI Caucasian Frequencies
- 1 in 71 Caucasians included
- 98.59% Caucasians excluded
If CPI/CPE Stats are Used
(ST = 150 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 - D18S51

If Assume 2 Contributors:

<table>
<thead>
<tr>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>16,18</td>
<td>14,20</td>
</tr>
</tbody>
</table>

\[
mRMP_{\text{minor}} = 2pq = 2 \times f(14) \times f(20) = 2 \times (0.1735) \times (0.0255) = 0.00884 \text{ or } 1 \text{ in } 113
\]
mRMP - TPOX

If Assume 2 Contributors....

<table>
<thead>
<tr>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>8,8</td>
<td>11,8 OR 11,11</td>
</tr>
</tbody>
</table>

mRMP = 8,11 + 11,11
mRMP = 2pq + (q^2 + q(1-q)θ)

mRMP = 2(0.5443)(0.2537) +
(0.2537)^2 + (0.2537)(0.7463)(0.01)
= 0.3424 or 1 in 2.9
Potential for Drop-out
If mRMP/LR Stats are Used

<table>
<thead>
<tr>
<th>Can use</th>
<th>Loci with potential D-out</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8</td>
<td>D7</td>
</tr>
<tr>
<td>D21</td>
<td>D2</td>
</tr>
<tr>
<td>D18</td>
<td>TH01</td>
</tr>
<tr>
<td>D3</td>
<td>vWA</td>
</tr>
<tr>
<td>D19</td>
<td>D13</td>
</tr>
<tr>
<td>TPOX</td>
<td>D5</td>
</tr>
<tr>
<td>FGA</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>D16</td>
</tr>
</tbody>
</table>
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*

http://shakespeareauthorship.com/
Macbeth/Duncan Profile - TH01

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7, 7</td>
<td></td>
</tr>
</tbody>
</table>

Possible Minor Contributors

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7, 9.3</td>
<td>(2pq)</td>
</tr>
<tr>
<td>9.3, 9.3</td>
<td>p²</td>
</tr>
<tr>
<td>9.3, ?</td>
<td>2p</td>
</tr>
</tbody>
</table>
Macbeth/Duncan Profile - TH01

\[
\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{V \& S}{V \& U} = \frac{f_7^2 + f_7 (1-f_7)\theta \& 1}{f_7^2 + f_7 (1-f_7)\theta \& 2p}
\]

\[V = 7, 7\]
\[U = 7, 9.3\]
\[9.3, 9.3\]
\[9.3, ?\]

\[f_{9.3} = 0.3054\]

\[f_{9.3} = 0.3054\]

\[= \frac{1}{2f_{9.3}} = \frac{1}{0.6108} = 1.63\]
Macbeth/Duncan Profile - TH01

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

\[V = 7, 7\]
\[U = 7, 9.3\]
\[9.3, 9.3\]

Let \(ST = 125\) RFU

\[f_{9.3} = 0.3054\]
\[f_7 = 0.1724\]

\[\frac{1}{f_{9.3}^2 + f_{9.3} (1-f_{9.3})\theta + 2f_{9.3}f_7} = \frac{1}{0.2007} = 4.98\]
Macbeth/Duncan Profile - TH01

LR

\[
\begin{array}{l}
\text{ST} = 200 \ (2p \text{ is used}) \quad 1.63 \\
\text{ST} = 125 \ (2pq \text{ is used}) \quad 4.98
\end{array}
\]
CPE/CPI (RMNE) Limitations

• A CPE/CPI approach assumes that all alleles are present (i.e., cannot handle allele drop-out)

• Thus, statistical analysis of low-level DNA CANNOT be correctly performed with a CPE/CPI approach because some alleles may be missing

• Charles Brenner in his AAFS 2011 talk addressed this issue

• Research is on-going to develop allele drop-out models and software to enable appropriate calculations
1. The claim that is requires no assumption about number of contributors is mostly wrong.

2. The supposed ease of understanding by judge or jury is really an illusion.

3. Ease of use is claimed to be an advantage particularly for complicated mixture profiles, those with many peaks of varying heights. The truth is the exact opposite. The exclusion method is completely invalid for complicated mixtures.

4. The exclusion method is only conservative for guilty suspects.

• “Certainly no one has laid out an explicit and rigorous chain of reasoning from first principles to support the exclusion method. It is at best guesswork.”

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)

“The risk of producing apparently strong evidence against an innocent suspect by this approach was not negligible.”

30% of the cases had a CPI < 0.01
48% of the cases had a CPI < 0.05

“It is false to think that omitting a locus is conservative as this is only true if the locus does not have some exclusionary weight.”
Limitations of CPI/CPE

Me = 13, 13
I’m included!

D8S1179

2 person mixture
CPI labs should include the stutter allele as part of their calculation.

mRPM/LR labs can treat this as stutter since they assume 2 contributors.

Hurts the innocent
Helps the guilty
Challenges with low level, complex mixtures
D8S1179  D21S11  D7S820  CSF1PO

D3S1358  TH01  D13S317  D16S539  D2S1338

D19S433  vWA  TPOX  D18S51

Amelogenin  D5S818  FGA

Identifiler
125 pg total DNA

AT = 30 RFU
ST = 150 RFU
Stutter filter off
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.
**Complex Mixture**

**Identifier**  
125 pg total DNA

AT = 30 RFU  
ST = 150 RFU  
Stutter filter off

Peaks below stochastic threshold

**TPOX**  
5 alleles

**D18S51**

**D5S818**
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

- 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?

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!

Stochastic threshold = 150 RFU
Assuming Three Contributors…

Some Possible Contributions to This Result

1:1:1

3:1:1

Stochastic alert!

Stochastic alert!

Stochastic alert!
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%**

**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%**

Combine loci = 0.59 x 0.18 = 0.11 or **11%**

*Approximately 1 in every 9 Caucasians could be included in this mixture*
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
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.


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."

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.

“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
“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.”
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
3 person mixture – 1 major, 2 minor

D19S433
Review of One Replicate (of 50K)

3P mixture, 2 Unknowns, Conditioned on the Victim (major)

Good fit of the data to the model

D19S433

150 RFU
Review of 3 person mixture

≈75% major

≈12% minor “A”

≈13% minor “B”

Width of the spread is related to determining the uncertainty of the mix ratios.
Victim

Suspect B

Suspect A

Genotype Probability

D19S433

Genotypes

94.8%
1.0%
2.4%
1.7%

Genotypes:
- 13,14
- 13,14,2
- 13,16,2
- 14,14
- 14,16,2
Determining the LR for D19S433

Suspect A = 14, 16.2  

\[
H_p = 0.967
\]

<table>
<thead>
<tr>
<th>Allele Pair</th>
<th>Before Conditioning</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>14, 16.2</td>
<td></td>
<td>0.967</td>
</tr>
<tr>
<td>14, 14</td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>13, 16.2</td>
<td></td>
<td>0.026</td>
</tr>
<tr>
<td>13, 14</td>
<td></td>
<td>0.001</td>
</tr>
</tbody>
</table>

\[
LR = \frac{0.967}{0.967}
\]
Determining the LR for D19S433

Suspect A = 14, 16.2

\( H_P = 0.967 \)

\[
\text{LR} = \frac{0.967}{0.0122} = 79.26
\]

\( H_D \)
Combined LR = 5.6 Quintillion

<table>
<thead>
<tr>
<th>locus</th>
<th>allele pair</th>
<th>Likelihood</th>
<th>Genotype Probability Distribution</th>
<th>Weighted Likelihood</th>
<th>Likelihood Ratio</th>
<th>log(LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>x</td>
<td><img src="image" alt="CSF1PO" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF1PO</td>
<td>11, 12</td>
<td>0.686</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D13S317</td>
<td>9, 12</td>
<td>1</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D16S539</td>
<td>9, 11</td>
<td>0.985</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D18S51</td>
<td>13, 17</td>
<td>0.999</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D19S433</td>
<td>14, 16.2</td>
<td>0.967</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D21S11</td>
<td>28, 30</td>
<td>0.968</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2S1338</td>
<td>23, 24</td>
<td>0.998</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3S1358</td>
<td>15, 17</td>
<td>0.988</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D5S818</td>
<td>11, 11</td>
<td>0.451</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7S820</td>
<td>11, 12</td>
<td>0.984</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S1179</td>
<td>13, 12</td>
<td>0.203</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGA</td>
<td>21, 25</td>
<td>0.32</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH01</td>
<td>7, 7</td>
<td>0.887</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPOX</td>
<td>8, 8</td>
<td>1</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vWA</td>
<td>15, 20</td>
<td>0.998</td>
<td><img src="image" alt="Combined LR" /></td>
<td><img src="image" alt="Combined LR" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Review of One Replicate (of 50K)

D19S433

3P mixture,
3 Unknowns

Poor fit of the data to the model

150 RFU
D19S433

Major contributor ≈ 75%
(13, 14)
Pr = 1
No Conditioning (3 Unknowns)

D19S433

8.1% chances for the two contributors
<table>
<thead>
<tr>
<th>locus</th>
<th>allele pair</th>
<th>L</th>
<th>Q</th>
<th>R</th>
<th>S</th>
<th>L*S</th>
<th>L*R</th>
<th>LR</th>
<th>log(LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D19S433</td>
<td>13, 14</td>
<td>0.002</td>
<td>0.146</td>
<td>0.1082</td>
<td></td>
<td>0.00020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 14</td>
<td>0.270</td>
<td>0.109</td>
<td>0.0044</td>
<td></td>
<td>0.00118</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 14.2</td>
<td>0.002</td>
<td>0.093</td>
<td>0.0498</td>
<td></td>
<td>0.00008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 14.2</td>
<td>0.017</td>
<td>0.088</td>
<td>0.0392</td>
<td></td>
<td>0.00068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 16.2</td>
<td>0.013</td>
<td>0.081</td>
<td>0.0120</td>
<td>1</td>
<td>0.01295</td>
<td>0.00016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 16.2</td>
<td>0.018</td>
<td>0.074</td>
<td>0.0131</td>
<td></td>
<td>0.00023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 14.2</td>
<td>0.009</td>
<td>0.067</td>
<td>0.0361</td>
<td></td>
<td>0.00031</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 14</td>
<td>0.002</td>
<td>0.059</td>
<td>0.0498</td>
<td></td>
<td>0.0012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 15</td>
<td>0.001</td>
<td>0.038</td>
<td>0.0343</td>
<td></td>
<td>0.00002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 13</td>
<td>0.001</td>
<td>0.034</td>
<td>0.0587</td>
<td></td>
<td>0.00007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 13</td>
<td>0.002</td>
<td>0.029</td>
<td>0.0541</td>
<td></td>
<td>0.00019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 15</td>
<td>0.001</td>
<td>0.024</td>
<td>0.0373</td>
<td></td>
<td>0.00002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 16.2</td>
<td>0.017</td>
<td>0.021</td>
<td>0.0060</td>
<td></td>
<td>0.00010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 14.2</td>
<td>0.013</td>
<td>0.020</td>
<td>0.0180</td>
<td></td>
<td>0.00023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 15.2</td>
<td>0.001</td>
<td>0.018</td>
<td>0.0275</td>
<td></td>
<td>0.00003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15, 16</td>
<td>0.002</td>
<td>0.015</td>
<td>0.0006</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 15.2</td>
<td>0.001</td>
<td>0.009</td>
<td>0.0299</td>
<td></td>
<td>0.00003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 15.2</td>
<td>0.003</td>
<td>0.009</td>
<td>0.0137</td>
<td></td>
<td>0.00004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 16</td>
<td>0.000</td>
<td>0.009</td>
<td>0.0017</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 12</td>
<td>0.004</td>
<td>0.009</td>
<td>0.0125</td>
<td></td>
<td>0.00004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 15</td>
<td>0.001</td>
<td>0.006</td>
<td>0.0172</td>
<td></td>
<td>0.00001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 16</td>
<td>0.000</td>
<td>0.006</td>
<td>0.0019</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 13.2</td>
<td>0.001</td>
<td>0.004</td>
<td>0.0261</td>
<td></td>
<td>0.00003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.2, 14</td>
<td>0.001</td>
<td>0.003</td>
<td>0.0240</td>
<td></td>
<td>0.00002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.2, 15</td>
<td>0.001</td>
<td>0.002</td>
<td>0.0083</td>
<td></td>
<td>0.00001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 18.2</td>
<td>0.002</td>
<td>0.002</td>
<td>0.0017</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 19.1</td>
<td>0.019</td>
<td>0.002</td>
<td>0.0000</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 13.2</td>
<td>0.002</td>
<td>0.002</td>
<td>0.0120</td>
<td></td>
<td>0.00003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.2, 16</td>
<td>0.001</td>
<td>0.002</td>
<td>0.0006</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.2, 13</td>
<td>0.001</td>
<td>0.002</td>
<td>0.0168</td>
<td></td>
<td>0.00002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 18.2</td>
<td>0.002</td>
<td>0.001</td>
<td>0.0019</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.2, 14</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0155</td>
<td></td>
<td>0.00001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.2, 14.2</td>
<td>0.004</td>
<td>0.001</td>
<td>0.0065</td>
<td></td>
<td>0.00003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15, 15</td>
<td>0.000</td>
<td>0.001</td>
<td>0.0059</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15, 15.2</td>
<td>0.000</td>
<td>0.001</td>
<td>0.0035</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 17</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0000</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15, 16.2</td>
<td>0.000</td>
<td>0.001</td>
<td>0.0042</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.2, 15.2</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0038</td>
<td></td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1, 14.2</td>
<td>0.072</td>
<td>0.001</td>
<td>0.0097</td>
<td></td>
<td>0.00069</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D19S433

Suspect “A”

Genotype

39 probable genotypes
<table>
<thead>
<tr>
<th>Allele Pair</th>
<th>Probability</th>
<th>Frequency</th>
<th>Genotype</th>
<th>Prob *</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,14</td>
<td>0.002</td>
<td>0.1082</td>
<td></td>
<td>0.00020</td>
</tr>
<tr>
<td>14.2, 16.2</td>
<td>0.270</td>
<td>0.0044</td>
<td></td>
<td>0.00118</td>
</tr>
<tr>
<td>14, 14</td>
<td>0.002</td>
<td>0.0498</td>
<td></td>
<td>0.00008</td>
</tr>
<tr>
<td>13, 14.2</td>
<td>0.017</td>
<td>0.0392</td>
<td></td>
<td>0.00068</td>
</tr>
<tr>
<td>14, 16.2</td>
<td>0.013</td>
<td>0.0120</td>
<td></td>
<td>0.00016</td>
</tr>
<tr>
<td>13, 16.2</td>
<td>0.018</td>
<td>0.0131</td>
<td></td>
<td>0.00023</td>
</tr>
<tr>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td></td>
<td>etc...</td>
</tr>
</tbody>
</table>

\[
\text{LR} = \frac{0.013}{0.00385} = 3.38
\]

\[
\text{HP} = 0.013
\]

\[
\text{LR} = \frac{0.013}{0.00385} = 3.38
\]

\[
\text{HD} = 0.00385
\]
No Conditioning

<table>
<thead>
<tr>
<th>Profile</th>
<th>Combined log(LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D19S433</td>
<td>LR = 3.38</td>
</tr>
<tr>
<td>Suspect A log(LR) = 8.03</td>
<td></td>
</tr>
<tr>
<td>Suspect B log(LR) = 7.84</td>
<td></td>
</tr>
</tbody>
</table>

Conditioned on Victim

<table>
<thead>
<tr>
<th>Profile</th>
<th>Combined log(LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D19S433</td>
<td>LR = 79.26</td>
</tr>
<tr>
<td>Suspect A log(LR) = 18.72</td>
<td></td>
</tr>
<tr>
<td>Suspect B log(LR) = 19.45</td>
<td></td>
</tr>
</tbody>
</table>
LR with Pr(Drop-out)

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

V = 13, 14
CP = 13, 14.2
S = 15, 16.2

\[
\frac{P(E \mid H_1)}{P(E \mid H_2)}
\]
$V = 13, 14$
$CP = 13, 14.2$
$S = 15, 16.2$

$Pr(\text{Drop-out}) = 10\%$
$Pr(\text{Drop-in}) = 1\%$

$P(E|H_1) = \frac{Pr(\text{No Drop-out at 16.2}) \cdot Pr(\text{Drop-out at 15}) \cdot Pr(\text{No Drop-in})}{0.90 \cdot 0.10 \cdot 0.99} = 0.0891$
3 Person Mixture

\[
\begin{align*}
V &= 13, 14 \\
CP &= 13, 14.2 \\
S &= 15, 16.2 \\
\end{align*}
\]

\[
\frac{P(E | H_1)}{P(E | H_2)} = 0.0891
\]

Keith Inman, Norah Rudin and Kirk Lohmueller have modified the Balding program to incorporate your own data for estimating Pr(Drop-out).
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
D5S818

5 alleles
“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
Genotype Probabilities

A = 13,16

B = 11,13

C = 14,15
## Results for Contributor A (male)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele Pair</th>
<th>Probability</th>
<th>Genotype</th>
<th>Suspect</th>
<th>$H_p$ Numerator</th>
<th>$H_d$ Denominator</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF1PO</td>
<td>10, 11</td>
<td>0.572</td>
<td>0.1292</td>
<td></td>
<td>0.30563</td>
<td>0.07395</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, 12</td>
<td>0.306</td>
<td>0.2133</td>
<td>1</td>
<td>0.30563</td>
<td>0.0652</td>
<td>1.935</td>
</tr>
<tr>
<td></td>
<td>10, 12</td>
<td>0.12</td>
<td>0.1547</td>
<td></td>
<td>0.30563</td>
<td>0.01861</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.30563</td>
<td>0.15791</td>
<td>1.935</td>
</tr>
<tr>
<td>D13S317</td>
<td>11, 11</td>
<td>1</td>
<td>0.1149</td>
<td>1</td>
<td>1</td>
<td>0.11488</td>
<td>8.704</td>
</tr>
<tr>
<td>D8S1179</td>
<td>13, 16</td>
<td>0.998</td>
<td>0.0199</td>
<td>1</td>
<td>0.99786</td>
<td>0.0199</td>
<td>49.668</td>
</tr>
</tbody>
</table>

The match rarity between the evidence and suspect is 1.21 quintillion
Results for Contributor B (female)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele Pair</th>
<th>Probability</th>
<th>Genotype Frequency</th>
<th>Suspect</th>
<th>$H_p$</th>
<th>$H_d$</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>11, 13</td>
<td>0.073</td>
<td>0.0498</td>
<td>1</td>
<td>0.07338</td>
<td>0.00366</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, 14</td>
<td>0.034</td>
<td>0.0271</td>
<td></td>
<td></td>
<td>0.00092</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 14</td>
<td>0.006</td>
<td>0.0996</td>
<td></td>
<td></td>
<td>0.00065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 14</td>
<td>0.011</td>
<td>0.0606</td>
<td></td>
<td></td>
<td>0.00068</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 13</td>
<td>0.005</td>
<td>0.1115</td>
<td></td>
<td></td>
<td>0.0006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, 12</td>
<td>0.018</td>
<td>0.0303</td>
<td></td>
<td></td>
<td>0.00054</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 14</td>
<td>0.004</td>
<td>0.0271</td>
<td></td>
<td></td>
<td>0.00012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 13</td>
<td>0.003</td>
<td>0.0916</td>
<td></td>
<td></td>
<td>0.00031</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 16</td>
<td>0.003</td>
<td>0.0108</td>
<td></td>
<td></td>
<td>0.00003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 15</td>
<td>0.001</td>
<td>0.0379</td>
<td></td>
<td></td>
<td>0.00003</td>
<td></td>
</tr>
</tbody>
</table>

etc...

The match rarity between the evidence and suspect is 1.43 million
Results for Contributor C (male)

The match rarity between the evidence and suspect is 9.16 thousand
Contributor B (gray) (16%)

Contributor C (blue) (18%)

Conditioned on the Victim

Contributor A (66%)
The Power of Conditioning

Victim

Suspect A

\[ C = 14,15 \]
## The Power of Conditioning

<table>
<thead>
<tr>
<th>Contributor A</th>
<th>LR (no conditioning, 3unk)</th>
<th>Contributor B (victim)</th>
<th>LR (conditioned on victim + 2unk)</th>
<th>Contributor C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.21 Quintillion</td>
<td>1.43 Million</td>
<td>1.32 Quintillion</td>
<td>2.19 Million</td>
<td>9.16 Thousand</td>
</tr>
</tbody>
</table>

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.
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?

michael.coble@nist.gov
301-975-4330