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

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>📈 📈 📈 📈</td>
<td>📈 📈 📈</td>
</tr>
</tbody>
</table>

**4 alleles**
All heterozygotes and non-overlapping alleles

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>📈 📈 📈</td>
<td>📈 📈 📈</td>
</tr>
</tbody>
</table>

**3 alleles**
Heterozygote + heterozygote, one overlapping allele
Heterozygote + homozygote, no overlapping alleles

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>📈 📈 📈</td>
<td>📈 📈 📈</td>
</tr>
</tbody>
</table>

**2 alleles**
Heterozygote + heterozygote, two overlapping alleles
Heterozygote + homozygote, one overlapping allele
Homozygote + homozygote, no overlapping alleles

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>📈 📈</td>
<td>📈 📈</td>
</tr>
</tbody>
</table>

**1 allele**
Homozygote + homozygote, overlapping allele
### 3-Person Mixtures

<table>
<thead>
<tr>
<th>Observed Profile</th>
<th>6 alleles</th>
<th>5 alleles</th>
<th>4 alleles</th>
<th>3 alleles</th>
<th>2 alleles</th>
<th>1 allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>All heterozygotes and non-overlapping alleles</td>
<td>Two heterozygotes and one homozygote</td>
<td>Six combinations of heterozygotes, homozygotes, and overlapping alleles</td>
<td>Eight combinations of heterozygotes, homozygotes, and overlapping alleles</td>
<td>Five combinations of heterozygotes, homozygotes, and overlapping alleles</td>
<td></td>
<td>All homozygotes, overlapping allele</td>
</tr>
</tbody>
</table>

**150 total combinations**
Observed profile

4-Person Mixtures

8 alleles
All heterozygotes and non-overlapping alleles

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

“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.”
Psychedelic Mixtures

Turn On…

Tune In…

(Talk about) Drop Out
Next Issue of FSI-Genetics

Editorial

Focus issue—Analysis and biostatistical interpretation of complex and low template DNA samples
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\textsuperscript{a, b,*}, L. Gusm\~{a}o\textsuperscript{c}, H. Haned\textsuperscript{d}, W.R. Mayr\textsuperscript{e}, N. Morling\textsuperscript{f}, W. Parson\textsuperscript{g}, L. Prieto\textsuperscript{h}, M. Prinz\textsuperscript{i}, H. Schneider\textsuperscript{j}, P.M. Schneider\textsuperscript{k}, B.S. Weir\textsuperscript{l}
\[ LR = \frac{1}{2pq} \]

\[ LR = \frac{0}{2pq} \]

\[ LR = \frac{?}{2pq} \]

\[ p^2 + 2p(1-p) \text{ “2p”} \]
Exploratory data analysis for the interpretation of low template DNA mixtures

H. Haned\textsuperscript{a,*}, K. Slooten\textsuperscript{a,b}, P. Gill\textsuperscript{c,d}

\textsuperscript{a} Netherlands Forensic Institute, Department of Human Biological traces, The Hague, The Netherlands
\textsuperscript{b} VU University Amsterdam, Amsterdam, The Netherlands
\textsuperscript{c} Norwegian Institute of Public Health, Oslo, Norway
\textsuperscript{d} University of Oslo, Norway
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 interpretation of low level DNA mixtures

Hannah Kelly\textsuperscript{a,*}, Jo-Anne Bright\textsuperscript{a}, James Curran\textsuperscript{b}, John Buckleton\textsuperscript{a}

\textsuperscript{a} ESR, PB 92021 Auckland, New Zealand
\textsuperscript{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
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

\[
\begin{align*}
4! & = 1!2!1! \\
4 \times 3 \times 2 \times 1 & = 1 \times 2 \times 1 \\
& = 12
\end{align*}
\]
Let’s Prove It!

Allelic Vectors

<table>
<thead>
<tr>
<th>Allele</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>13, 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14, 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14, 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13, 15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
13, 14 & \text{ and } 14, 15 = 2ab \times 2bc = 4ab^2c \\
13, 15 & \text{ and } 14, 14 = 2ac \times b^2 = 2ab^2c \\
14, 15 & \text{ and } 13, 14 = 2bc \times 2ab = 4ab^2c \\
14, 14 & \text{ and } 13, 15 = b^2 \times 2bc = 2ab^2c
\end{align*}
\]

= 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

Allelic Vector
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

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

$$= 24\Pr(13, 14, 15 \mid X)$$
Apply the Sampling Formula (Balding and Nichols 1994)

\[
x \frac{\theta + (1 - \theta)p_A}{1 + (n-1) \theta}
\]

\(x = \) value calculated from the F-model.
\(p_a = \) frequency of the “a” allele.
\(\Theta = \) coancestry coefficient \((F_{ST})\).
\(n = \) number of alleles.
A Worked Example

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

POI = 28, 30
2 peaks – 4 alleles

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)

\[
\text{Pr}(A|X) = \frac{x\theta + (1 - \theta)p_a}{1 + (n-1)\theta}
\]

\[
\text{Pr}(E|Hp) = 1
\]

\[
\text{Pr}(E|Hd) = 12\text{Pr}(28,30|28,30)
\]

\[
12(\theta(1 - \theta)p_{28})(\theta + (1 - \theta)p_{30})
\]

\[
(1 + \theta)(1 + 2\theta)
\]

\[
LR = 1.86
\]
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
4Pr(28, 28, 28, 30|28, 30) + 6Pr(28, 28, 30, 30|28, 30) + 4Pr(28, 30, 30, 30|28, 30) + 12Pr(28, 28, 30, Q|28, 30)
+ 12Pr(28, 30, 30, Q|28, 30)
+ 12Pr(28, 30, Q, Q|28, 30)

Pr(E|Hd) = 2Pr(28, 30|28, 30) × \[6 - 6Pr(28|28, 28, 30, 30) - 6Pr(30|28, 28, 30, 30) + 2Pr(28, 28|28, 28, 30, 30)
+ 2Pr(30, 30|28, 28, 30, 30)
+ 3Pr(28, 30|28, 28, 30, 30)\]
\[\frac{2(\theta(1 - \theta)p_{28})(\theta + (1 - \theta)p_{30})}{(1 + \theta)(1 + 2\theta)(1 + 3\theta)(1 + 4\theta)} \times \]
\[6 - \frac{6(2\theta + (1 - \theta)p_{28})}{(1 + 3\theta)} - \frac{6(2\theta + (1 - \theta)p_{30})}{(1 + 3\theta)} + \frac{2(2\theta + (1 - \theta)p_{28})(3\theta + (1 - \theta)p_{28})}{(1 + 3\theta)(1 + 4\theta)} + \frac{2(2\theta + (1 - \theta)p_{30})(3\theta(1 - \theta)p_{30})}{(1 + 3\theta)(1 + 4\theta)} \]
\[+ \frac{3(2\theta + (1 - \theta)p_{28})(2\theta + (1 - \theta)p_{30})}{(1 + 3\theta)(1 + 4\theta)} \]
LR with Pr(Drop-out)
3 person mixture – 1 major, 2 minor
3 Person Mixture

\[ V = 13, 14 \]
\[ CP = 13, 14.2 \]
\[ S = 15, 16.2 \]

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

Pr(Drop-out) = 10%
Pr(Drop-in) = 1%

\[
P(E \mid H_1) = \frac{\text{Pr(No Drop-out at 16.2)}}{\text{Pr(Drop-out at 15)}} \cdot \text{Pr(No Drop-in)}
\]

\[
= \frac{0.90}{0.10} \cdot 0.99
\]

\[
= 0.0891
\]
Keith Inman, Norah Rudin and Kirk Lohmueller have modified the Balding program to incorporate your own data for estimating \( \Pr(\text{Drop-out}) \).
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/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

After the 75. roll
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

Yes - White
No - Black

Posterior Prob = 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.
- 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.

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.

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…

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,11</td>
<td>76%</td>
</tr>
<tr>
<td>11,11</td>
<td>15%</td>
</tr>
<tr>
<td>11,13</td>
<td>2%</td>
</tr>
<tr>
<td>8,11</td>
<td>2%</td>
</tr>
<tr>
<td>11,12</td>
<td>2%</td>
</tr>
<tr>
<td>9,9</td>
<td>1%</td>
</tr>
<tr>
<td>9,12</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>10,11</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>8,12</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>8,9</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>
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

D19S433
True Allele Casework Workflow
5 Modules

- Analyze
- Data
- Request

Server

State Assumptions
2, 3, 4 unknowns
1 Unk with Victim?

Set Parameters
MCMC modeling (e.g. 50K)
Degradation?
True Allele Casework Workflow
5 Modules

Analyze → Data → Request → Review

Server

Computation
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
True Allele Casework Workflow
5 Modules

Analyze → Data → Request → Review

Server

Computation

Report
## Determining the LR for D19S433

Suspect A = 14, 16.2  \quad H_p = 0.967

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

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

Suspect A = 14, 16.2

<table>
<thead>
<tr>
<th>Allele Pair</th>
<th>Probability Before Conditioning</th>
<th>Probability * Genotype Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>14, 16.2</td>
<td>0.967</td>
<td>0.01164</td>
</tr>
<tr>
<td>14, 14</td>
<td>0.003</td>
<td>0.00013</td>
</tr>
<tr>
<td>13, 16.2</td>
<td>0.026</td>
<td>0.00034</td>
</tr>
<tr>
<td>13, 14</td>
<td>0.001</td>
<td>0.00009</td>
</tr>
</tbody>
</table>

$\text{H}_P = 0.967$

$$\text{LR} = \frac{0.967}{0.0122} = 79.26$$

$\text{H}_D$
# Combined LR = 5.6 Quintillion

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

• Results are expressed as logLR values

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

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

\[
\log(\text{LR}) = 6 \times \log 10 \ (1)
\]

\[
\log(\text{LR}) = 6
\]
Review of One Replicate (of 50K)

D19S433

3P mixture,
3 Unknowns

Poor fit of the data to the model

150 RFU
No Conditioning
(3 Unknowns)

D19S433

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

D19S433

8.1% Genotype Probability

Genotype Probability

14, 15.2
14, 16
14, 16.2
14, 17
14, 18.2

14, 15.2
14, 16.2
14, 17
15, 15.2
15, 16.2
15, 17
<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>0.00020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.2, 16.2</td>
<td>0.270</td>
<td>0.109</td>
<td>0.0044</td>
<td>0.00118</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 14</td>
<td>0.002</td>
<td>0.093</td>
<td>0.0498</td>
<td>0.00008</td>
<td></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>0.00058</td>
<td></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>0.01295</td>
<td>0.0016</td>
<td></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>0.00023</td>
<td></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>0.00031</td>
<td></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>0.00012</td>
<td></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>0.00002</td>
<td></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>0.00007</td>
<td></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>0.00010</td>
<td></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>0.00002</td>
<td></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>0.00010</td>
<td></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>0.00023</td>
<td></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>0.00033</td>
<td></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>0.00000</td>
<td></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>0.00003</td>
<td></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>0.00004</td>
<td></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>0.00000</td>
<td></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>0.00004</td>
<td></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>0.00001</td>
<td></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>0.00000</td>
<td></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>0.00003</td>
<td></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>0.00002</td>
<td></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>0.00001</td>
<td></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>0.00000</td>
<td></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>0.00000</td>
<td></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>0.00003</td>
<td></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>0.00000</td>
<td></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>0.00002</td>
<td></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>0.00000</td>
<td></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>0.00001</td>
<td></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>0.00003</td>
<td></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>0.00000</td>
<td></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>0.00000</td>
<td></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>0.00000</td>
<td></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>0.00000</td>
<td></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>0.00000</td>
<td></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>0.00009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01295</td>
<td>0.00385</td>
<td>3.367</td>
<td>0.527</td>
</tr>
</tbody>
</table>

Suspect “A”
Genotype

39 probable genotypes

D19S433
<table>
<thead>
<tr>
<th>Allele Pair</th>
<th>Probability</th>
<th>Genotype</th>
<th>Probability</th>
<th>Frequency</th>
<th>GenFreq</th>
<th>Prob *</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,14</td>
<td>0.002</td>
<td></td>
<td>0.00020</td>
<td>0.1082</td>
<td>0.00020</td>
<td></td>
</tr>
<tr>
<td>14.2, 16.2</td>
<td>0.270</td>
<td></td>
<td>0.00118</td>
<td>0.0044</td>
<td>0.00044</td>
<td></td>
</tr>
<tr>
<td>14, 14</td>
<td>0.002</td>
<td></td>
<td>0.00008</td>
<td>0.0498</td>
<td>0.00049</td>
<td></td>
</tr>
<tr>
<td>13, 14.2</td>
<td>0.017</td>
<td></td>
<td>0.00068</td>
<td>0.0392</td>
<td>0.00039</td>
<td></td>
</tr>
<tr>
<td>14, 16.2</td>
<td>0.013</td>
<td></td>
<td>0.00016</td>
<td>0.0120</td>
<td>0.00012</td>
<td></td>
</tr>
<tr>
<td>13, 16.2</td>
<td>0.018</td>
<td></td>
<td>0.00023</td>
<td>0.0131</td>
<td>0.00013</td>
<td></td>
</tr>
<tr>
<td>etc...</td>
<td>etc...</td>
<td></td>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>0.00385</td>
<td></td>
<td>0.00385</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

\[ H_P = 0.013 \]

\[ H_D = 0.00385 \]

Suspect A = 14, 16.2

No Conditioning (3 Unknowns)
No Conditioning

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

Conditioned on Victim

D19S433
LR = 3.38

D19S433
LR = 79.26
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
4 alleles – 10 loci
3 alleles – 5 loci
2 alleles – 0 loci
1 allele – 0 loci

“Medium” Sharing
4 alleles – 3 loci
3 alleles – 8 loci
2 alleles – 4 loci
1 allele – 0 loci

“High” Sharing
4 alleles – 0 loci
3 alleles – 6 loci
2 alleles – 8 loci
1 allele – 1 loci
```

Match Score in Duplicate Runs

Match Rarity (log(LR))

RMP

Minor Component

Major Component

“Easy” for Deconvolution
Match Score in Duplicate Runs

Match Rarity (log(LR))

10:90  20:80  30:70  50:50  60:40  70:30  80:20  90:10

Minor Component

Major Component

“Challenging” for Deconvolution
Match Score in Duplicate Runs

Match Rarity (log(LR))

Minor Component

Major Component

"Difficult" for Deconvolution
Match Rarity log(LR)

<table>
<thead>
<tr>
<th>RMNE</th>
<th>LR (Classic)</th>
<th>LR (True Allele)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

TPOX
D5S818
D18S51

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
A = 13,16
B = 11,13
C = 12,14

“A = 13,16”
“B = 11,13”
“C = 14,15”
Contributor B (green)  
(16%)

Contributor C (blue)  
(18%)

Contributor A  
(66%)
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>1</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></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>$H_p$ Numerator</th>
<th>$H_d$ Denominator</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)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele Pair</th>
<th>Probability</th>
<th>Genotype</th>
<th>Frequency</th>
<th>Suspect</th>
<th>Numerator</th>
<th>Denominator</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>11, 13</td>
<td>0.056</td>
<td></td>
<td>0.0498</td>
<td></td>
<td></td>
<td></td>
<td>0.00279</td>
</tr>
<tr>
<td></td>
<td>13, 14</td>
<td>0.007</td>
<td></td>
<td>0.0996</td>
<td></td>
<td></td>
<td></td>
<td>0.00066</td>
</tr>
<tr>
<td></td>
<td>12, 14</td>
<td>0.011</td>
<td></td>
<td>0.0606</td>
<td></td>
<td></td>
<td></td>
<td>0.00068</td>
</tr>
<tr>
<td></td>
<td>11, 14</td>
<td>0.021</td>
<td></td>
<td>0.0271</td>
<td></td>
<td></td>
<td></td>
<td>0.00056</td>
</tr>
<tr>
<td></td>
<td>12, 13</td>
<td>0.006</td>
<td></td>
<td>0.1115</td>
<td></td>
<td></td>
<td></td>
<td>0.00066</td>
</tr>
<tr>
<td></td>
<td>14, 14</td>
<td>0.005</td>
<td></td>
<td>0.0271</td>
<td></td>
<td></td>
<td></td>
<td>0.00013</td>
</tr>
</tbody>
</table>

... 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

Victim

Suspect A

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

<table>
<thead>
<tr>
<th>Contributor</th>
<th>LR (no conditioning, 3unk)</th>
<th>LR (conditioned on victim + 2unk)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributor A</td>
<td>1.21 Quintillion</td>
<td>1.32 Quintillion</td>
</tr>
<tr>
<td>Contributor B (victim)</td>
<td>1.43 Million</td>
<td>2.19 Million</td>
</tr>
<tr>
<td>Contributor C</td>
<td>9.16 Thousand</td>
<td>59.8 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.
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

Questions?

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

john.butler@nist.gov
301-975-4049

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