Exploring the Capabilities of Mixture Interpretation Using True Allele Software

Michael D. Coble and John M. Butler
NIST Applied Genetics Group

September 3 2011
Official Disclaimer

The opinions and assertions contained herein are solely those of the author and are not to be construed as official or as views of the U.S. Department of Commerce, the U.S. Department of Justice, or the National Institute of Justice.

Commercial software, equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the U.S. Department of Commerce, the U.S. Department of Justice, or the National Institute of Justice nor does it imply that any of the software, materials, instruments or equipment identified are necessarily the best available for the purpose.
Fallible DNA evidence can mean prison or freedom

Q: Lab staff need more training on how to deal with complex profiles such as mixtures and very small samples of DNA

Responses from Australia, Canada, India, New Zealand, UK, and US.
“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.”
Quantitative computer interpretation using Markov Chain Monte Carlo testing
- Models peak uncertainty and infers possible genotypes
- Results are presented as the Combined LR
True Allele Software (Cybergenetics)

• We purchased the software in September 2010.
• Three day training at Cybergenetics (Pittsburgh, PA) in October.
• Software runs on a Linux Server with a Mac interface.
True Allele Casework Workflow
5 Modules

Analyze

.fsa files imported
Size Standard check
Allelic Ladder check
Alleles are called
True Allele Casework Workflow
5 Modules

All Peaks above 10 RFU are considered
True Allele Casework Workflow
5 Modules

- Analyze
- Data
- Request

Server

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

Set Parameters
-Degradation?
-MCMC modeling (e.g. 50K)
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

150 RFU

D19S433
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

13,14
13,14,2
13,16,2
14,14
14,16,2

94.8%
1.0%
2.4%
1.7%

True Allele Casework Workflow
5 Modules

Analyze → Data → Request → Review

Server → Computation → Review → Report
Determining the LR for D19S433

Suspect A = 14, 16.2

\[ H_P = 1 \times 0.967 \]

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

\[ H_D \]

<table>
<thead>
<tr>
<th>Allele Pair</th>
<th>Probability Before Conditioning</th>
<th>Genotype Frequency</th>
<th>Probability * Genotype Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>14, 16.2</td>
<td>0.967</td>
<td>0.0120</td>
<td>0.01164</td>
</tr>
<tr>
<td>14, 14</td>
<td>0.003</td>
<td>0.0498</td>
<td>0.00013</td>
</tr>
<tr>
<td>13, 16.2</td>
<td>0.026</td>
<td>0.0131</td>
<td>0.00034</td>
</tr>
<tr>
<td>13, 14</td>
<td>0.001</td>
<td>0.1082</td>
<td>0.00009</td>
</tr>
</tbody>
</table>

sum: 0.0122
<table>
<thead>
<tr>
<th>locus</th>
<th>allele pair</th>
<th>Likelihood</th>
<th>Genotype Probability Distribution</th>
<th>Questioned</th>
<th>Reference</th>
<th>Suspect</th>
<th>Numerator</th>
<th>Denominator</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></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></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></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></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></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></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></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></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></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></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></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></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></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></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></td>
<td>0.99808</td>
<td>0.00569</td>
<td>174.834</td>
<td>2.243</td>
</tr>
</tbody>
</table>

Combined LR = 5.6 Quintillion
Results

- Results are expressed as logLR values

\[
\begin{align*}
LR &= 1,000,000 = 10^6 \\
\log(LR) &= \log10^6 \\
\log(LR) &= 6 \times \log10 \ (1) \\
\log(LR) &= 6
\end{align*}
\]
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

Uncertainty remains for the two minor contributors

8.1%
### Suspect “A” Genotype

<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>14.2, 16.2</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>14, 16.2</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>14, 14.2</td>
<td>0.017</td>
<td>0.088</td>
<td>0.0332</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>0.00012</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>0.00010</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>0.00009</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**D19S433**

39 probable genotypes
Genera 1 = 14, 16.2

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

\[
\text{Sum} = 0.00385 \\
\text{LR} = \frac{0.013}{0.00385} = 3.38 \\
\text{HD} = 1 * 0.013
\]

No Conditioning (3 Unknowns)
No Conditioning

Profile - Combined \( \log(\text{LR}) \)
- Suspect A \( \log(\text{LR}) = 8.03 \)
- Suspect B \( \log(\text{LR}) = 7.84 \)

Conditioned on Victim

Profile - Combined \( \log(\text{LR}) \)
- D19S433 \( \log(\text{LR}) = 3.38 \)
- Suspect A \( \log(\text{LR}) = 18.72 \)
- Suspect B \( \log(\text{LR}) = 19.45 \)
Exploring the Capabilities

- Degree of Allele Sharing
- Mixture Ratios
- DNA Quantity
Mixture Data Set

- Mixtures of pristine male and female DNA amplified at a total concentration of 1.0 ng/μL using Identifiler (standard conditions).
- 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))

Minor Component

Major Component

“Easy” for Deconvolution
Match Score in Duplicate Runs

Match Rarity (log(LR))

Minor Component

Major Component

“Challenging” for Deconvolution
**Match Score in Duplicate Runs**

- **Match Rarity (log(LR))**
- **Minor Component**
  - 10:90
  - 20:80
  - 30:70
  - 50:50
- **Major Component**
  - 60:40
  - 70:30
  - 80:20
  - 90:10

- **RMP**

- **“Difficult” for Deconvolution**
Match Rarity log(LR)

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

5 alleles

TPOX
D18S51
D5S818
“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
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 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).
Future Work

• More work will be performed with low level, complex (3 and 4 person) mixtures.
Thank You!

Forensic DNA Team

John Butler  Mike Coble  Becky Hill  Margaret Kline

Data Analysis Support

Dave Duewer

DNA Biometrics Team

Pete Vallone  Erica Butts  Kristen Lewis O’Connor

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

Funding from the FBI S&T Branch through NIST Information Access Division

Cybergenetics – Dr. Mark Perlin, Dr. Ria David, Bill Allan, Matt Legler, and Erin Turo