Probabilistic Genotyping

Michael D. Coble

October 15, 2012
Nashville, TN
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 a,b,* L. Gusmão c, H. Haned d, W.R. Mayr e, N. Morling f, W. Parson g, L. Prieto h, M. Prinz i, H. Schneider j, P.M. Schneider k, B.S. Weir l
ISFG Recommendations

- \( \Pr(D) = \text{Prob. Drop-out (het)} \)
- \( \Pr(D) = \text{No Prob. Drop-out (het)} \)
- \( \Pr(D_2) = \text{Prob. Drop-out (hom)} \)
- \( \Pr(D_2) = \text{No Prob. Drop-out (hom)} \)
- \( \Pr(C) = \text{Prob. Drop-in} \)
- \( \Pr(C) = \text{No Prob. Drop-in} \)
Prosecutor’s Explanation

No Drop-out of the “A” allele
The “B” allele dropped out
No other Drop-in

Pr(D) Pr(D) Pr(\overline{C})
The LR

\[ LR = \frac{\Pr(D) \Pr(D) \Pr(\overline{C})}{\Pr(\overline{D})} \]
Defense Explanation

4 possibilities

(1) The real culprit is a homozygote

\[ p_a^2 \Pr(\overline{D_2}) \Pr(\overline{C}) \]
Defense Explanation

4 possibilities

(2) Drop out of a heterozygote (not B)
No drop-in of “A”

\[ 2p_a p_Q \overline{Pr(D)} \overline{Pr(D)} \overline{Pr(C)} \]
Defense Explanation

4 possibilities

(3) Drop out of a homozygote (not B)
Drop in of “A”

\[ p_Q^2 \Pr(D_2) \Pr(C)p_a \]
Defense Explanation

4 possibilities

(4) Drop out of a homozygote (not AB)
Drop in of “A”

$2p_Q p_{Q'} \Pr(D)^2 \Pr(C) p_a$
The LR

\[
LR = \frac{\Pr(D) \Pr(D) \Pr(\overline{C})}{p_a^2 \Pr(D^2) \Pr(\overline{C}) + 2p_a p_Q \Pr(D) \Pr(D) \Pr(\overline{C}) + p_Q^2 \Pr(D^2) \Pr(\overline{C}) p_a + 2p_Q p_Q' \Pr(D)^2 \Pr(\overline{C}) p_a}
\]
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
Probabilistic Modeling of TA

Mathematical Modeling of the Data

PHR, Mix Ratio, Stutter etc…

50-100,000 Simulations (MCMC)

Probable Genotypes to explain the mixture

<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>
Uncertainty with D16S539

The 11 allele is at 169 RFU (above 150 ST).

The “12” peak in the stutter position is only slightly below our stutter threshold of 10.4%.

If we assume 8 and 12 are stutter peaks, then the possible genotypes of the minor contributor are - 9,11  11,11  11,13

Should we also include the 8 and 12 alleles in creating our genotype combinations?
Summary – Mixture Weight

100,000 MCMC examinations of the data.

2 unknowns (no conditioning)

Clear separation of the two contributors.
Model doesn’t exactly fit the data

Most of the time (76%), 9,11 is predicted to be the genotype of the minor contributor
### Determining the LR for D16S539 (Hp)

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Probability (Before Conditioning)</th>
<th>Genotype Freq (HWE)</th>
<th>(Prob) x (HWE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,11</td>
<td>0.431</td>
<td>0.0719</td>
<td>0.031</td>
</tr>
<tr>
<td>11,11</td>
<td>0.098</td>
<td>0.1025</td>
<td>0.01</td>
</tr>
<tr>
<td>11,13</td>
<td>0.014</td>
<td>0.093</td>
<td>0.0013</td>
</tr>
<tr>
<td><strong>8,11</strong></td>
<td><strong>0.092</strong></td>
<td>0.0106</td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>11,12</td>
<td>0.008</td>
<td>0.2093</td>
<td>0.0016</td>
</tr>
<tr>
<td>9,9</td>
<td>0.013</td>
<td>0.0126</td>
<td>0.0002</td>
</tr>
<tr>
<td>9,12</td>
<td>0.003</td>
<td>0.0734</td>
<td>0.0002</td>
</tr>
<tr>
<td>10,11</td>
<td>0.003</td>
<td>0.036</td>
<td>0.0001</td>
</tr>
<tr>
<td>8,12</td>
<td>0.014</td>
<td>0.0108</td>
<td>0.0002</td>
</tr>
<tr>
<td><strong>8,9</strong></td>
<td><strong>0.015</strong></td>
<td><strong>0.0037</strong></td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td><strong>(sum)</strong></td>
<td><strong>0.046</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Suspect = 8,11**

\[
LR = \frac{0.0992}{0.046}
\]
## Determining the LR for D16S539 ($H_D$)

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Probability (Before Conditioning)</th>
<th>Genotype Freq (HWE)</th>
<th>(Prob) x (HWE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,11</td>
<td>0.431</td>
<td>0.0719</td>
<td>0.031</td>
</tr>
<tr>
<td>11,11</td>
<td>0.098</td>
<td>0.1025</td>
<td>0.01</td>
</tr>
<tr>
<td>11,13</td>
<td>0.014</td>
<td>0.093</td>
<td>0.0013</td>
</tr>
<tr>
<td>8,11</td>
<td>0.092</td>
<td>0.0106</td>
<td>0.001</td>
</tr>
<tr>
<td>11,12</td>
<td>0.008</td>
<td>0.2093</td>
<td>0.0016</td>
</tr>
<tr>
<td>9,9</td>
<td>0.013</td>
<td>0.0126</td>
<td>0.0002</td>
</tr>
<tr>
<td>9,12</td>
<td>0.003</td>
<td>0.0734</td>
<td>0.0002</td>
</tr>
<tr>
<td>10,11</td>
<td>0.003</td>
<td>0.036</td>
<td>0.0001</td>
</tr>
<tr>
<td>8,12</td>
<td>0.014</td>
<td>0.0108</td>
<td>0.0002</td>
</tr>
<tr>
<td>8,9</td>
<td>0.015</td>
<td>0.0037</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Suspect** = $8,11$

\[
LR = \frac{0.092}{0.046} = 2.0
\]
<table>
<thead>
<tr>
<th></th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assume Stutter @8,12</td>
<td>3.6 (fails to capture 8,11)</td>
</tr>
<tr>
<td>Include 8,12</td>
<td>2.3</td>
</tr>
<tr>
<td>True Allele</td>
<td>2.0</td>
</tr>
</tbody>
</table>
D16S539 Results

Joint LR = 16.7 Billion
(using True Allele, 2unk)

Using “2P” = 26.5 Trillion
Complex Mixture
True Allele Results – 3 person mixture

100K examinations
3 unknowns
(no conditioning)

No clear separation

Mix ratio (green)
10-60%
VERY Poor fit of the data to the model
True Allele Results – 4 person mixture

100K examinations
4 unknowns
(no conditioning)

Better separation,
Still uncertainty.
Still a poor fit of the data to the model
Potential Suspects

- A, B, C and D are the four individuals in the mixture.
- John Butler is also a suspect (The Butler did it).
- “Omni man” is also a possible suspect.
ABCD
14,20
16,18
13,17
13,14

Omni
14,17
Omni Man

![Graph showing data points with values: 0.8820, 4.2619, -1.4630, -3.1251.]
Strategies

• Conditioning will help…

• This may not be possible.

• Multiple replicates will be necessary.

• There is a need to determine an appropriate method for an inclusion log(LR).
Summary of the Issues

- New kits, new instruments can only increase the difficulties of interpreting low-level, challenging samples.

- Probabilistic methods will be necessary to interpret low level samples with drop-out potential (or contaminating alleles) since classical approaches to interpretation such as RMNE or mRMP (even the classic LR) will not suffice.
Thanks to NIJ for Support of BU and NIST

- NIJ Forensic Science Training Development and Delivery Program Grant # 2008-DN-BX-K158, awarded to Biomedical Forensic Science Program at Boston University School of Medicine

- NIJ has an Interagency Agreement (IAA) with the NIST Office of Law Enforcement Standards (OLES)