DNA Mixture Interpretation & Statistical Analysis

John M. Butler
NIST Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland

Steps Involved in Process of Forensic DNA Typing

1) Data Interpretation
2) Statistical Interpretation

Gathering the Data
Understanding the Data

Advanced Topics: Methodology
Advanced Topics: Interpretation

DNA Mixture Interpretation & Statistical Analysis

John M. Butler
NIST Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland

SWGDAM Website and Resources Available

http://www.swgdam.org/resources.html

Mixture Training Materials
Reviewed by SWGDAM Mixture Committee

SWGDAM Mixture Committee Resource Page

The following information resources have been produced and reviewed by members of the Mixture Committee of the Scientific Working Group on DNA Analysis Methods (SWGDAM) and are available at http://www.cstl.nist.gov/biotech/strbase/mixture/SWGDAM-mixture-info.htm

Mixture Training Examples
- Download "Mixture 6" PowerPoint show (56 Mb)
- Download "Mixture IQAS2014" PowerPoint show (35 Mb)

http://www.cstl.nist.gov/biotech/strbase/mixture/SWGDAM-mixture-info.htm

Mixture Workshop (Promega ISHI 2010)

http://www.cstl.nist.gov/biotech/strbase/mixture.htm

Regional workshops presented in FL, TX, MI, and AZ (October 3, 2011)

Updated Focus Workshop presented at ISHI 2011 (October 3, 2011)

Butler, John, Ph.D.
Coble, Michael, Ph.D.
Cotton, Robin, Ph.D.
Grgicak, Catherine, Ph.D.
Word, Charlotte, Consultant

This workshop is for analysts, technical reviewers and technical leaders performing and interpreting validation studies and/or interpreting and reviewing STR data, particularly more difficult mixtures. Various DNA profiles will be analyzed and interpreted using selected analytical thresholds and stochastic thresholds to demonstrate the impact of those values on the profiles amplified with low-template DNA vs. higher amounts of DNA. Different statistical approaches and conclusions suitable for the profiles will be presented.
Useful Articles on DNA Mixture Interpretation


German Mixture Classification Scheme


(Chowning 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

---

Available for download from the ISFG Website:
http://www.isfg.org/Publication;Gill2006

DNA commission of the International Society of Forensic Genetics:
Recommendations on the interpretation of mixtures
P. Gill1, T.M. Clayton2, J.P. Whittaker3, R. Sparkes4, P. Gill5

1 Forensic Science Service, North Camp, Wattisham, Suffolk IP28 2JE, UK
2 Forensic Science Service, Peterboroug, Southwell, Nottingham NG25 0GG, UK
3 Forensic Science Service, Royaal Huis, Coop Hugh Street, Radcliffe-on-Trent NG3 5HJ, UK
4 Forensic Science Service, Police Force, Manchester, Manchester M3 7JH, UK

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


ISFG Recommendations on Mixture Interpretation

http://www.isfg.org/Publication;Gill2006

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
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
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


Steps in the interpretation of mixtures


<table>
<thead>
<tr>
<th>Step #1</th>
<th>Identify the Presence of a Mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step #2</td>
<td>Designate Allele Peaks</td>
</tr>
<tr>
<td>Step #3</td>
<td>Identify the Number of Potential Contributors</td>
</tr>
<tr>
<td>Step #4</td>
<td>Estimate the Relative Ratio of the Individuals Contributing to the Mixture</td>
</tr>
<tr>
<td>Step #5</td>
<td>Consider All Possible Genotype Combinations</td>
</tr>
<tr>
<td>Step #6</td>
<td>Compare Reference Samples</td>
</tr>
</tbody>
</table>
Steps in DNA Interpretation

<table>
<thead>
<tr>
<th>Peak</th>
<th>Allele</th>
<th>Genotype</th>
<th>Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>(vs. noise)</td>
<td>(vs. artifact)</td>
<td>(allele pairing)</td>
<td>(genotype combining)</td>
</tr>
</tbody>
</table>

Data Interpretation

- Signal observed:
  - Peak
  - Allele
  - Genotype
  - Profile

Data Collection

- Sample Deposited
- Sample Collected
- Extraction
- Quantitation
- PCR
- Amplification
- CE
- Separation/Detection
- Signal observed:
  - Peak
  - Allele
  - Genotype
  - Profile

Comparison to Known(s)

- Weight of Evidence (Stats)

Overview of Two Thresholds

- Called Peak
  - Greater confidence a sister allele has not dropped out
  - Example values (empirically determined based on own internal validation)
- Stochastic Threshold
  - The value above which it is reasonable to assume that allelic dropout of a sister allele has not occurred

Coupling of Statistics and Interpretation

- The CPE/CPI approach for reporting an inclusionary statistic requires that all alleles be observed in the evidence sample
- If allele drop-out is suspected at a locus, then any allele is possible and the probability of inclusion goes to 100% -- in other words, the locus is effectively dropped from consideration
- If alleles are seen below the established stochastic threshold, then the locus is typically eliminated (“INC” -- declared inconclusive) in many current lab SOPs

Can This Locus Be Used for Statistical Calculations?

- It depends on your assumption as to the number of contributors!
- If you assume a single-source sample, then you can assume that the detection of two alleles fully represents the heterozygous genotype present at this locus.
- If you assume (from examining other loci in the profile as a whole) that the sample is a mixture of two or more contributors, then there may be allele drop-out and all alleles may not be fully represented.

Limitations of Stochastic Thresholds

- The possibility of allele sharing with a complex mixture containing many contributors may make a stochastic threshold meaningless
- “Enhanced interrogation techniques” to increase sensitivity (e.g., increased PCR cycles) may yield false homozygotes with >1000 RFU
- New turbo-charged kits with higher sensitivity will need to be carefully evaluated to avoid allele drop-out and false homozygotes

PowerPlex 16 HS Stochastic Threshold

(ABI 3500 Data – see Poster #42)

- PCR = 30 cycles
- PowerPlex 16 HS
  - AVG
  - AVG + 1SD
  - AVG + 2SD
  - AVG + 3SD
  - MAX
  - PowerPlex 16 HS
  - 6.9

Correct type = 6.9

- AT = 215 RFU

Data from Erica Butts (NIST)
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

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.”

DAB Recommendations on Statistics
February 23, 2000

“The DAB finds either one or both PE or LR calculations acceptable and strongly recommends that one or both calculations be carried out whenever feasible and a mixture is indicated”

- Probability of exclusion (PE)
- Likelihood ratios (LR)

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

Notes from Charles Brenner’s AAFS 2011 talk
The Mythical “Exclusion” Method for Analyzing DNA Mixtures – Does it Make Any Sense at all?

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.
5. “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.”


Forensic inference from genetic markers

B Devlin: Department of Epidemiology and Public Health, Yale University School of Medicine

Section 5.1 Exclusion probability

- Discussion about exclusion probabilities in Paternity cases.
Two types:

(1) Conditional Exclusion Probability - excluding a random man as a possible father, given the mother-child genotypes for a particular case.
(2) Average Exclusion Probability – excluding a random man as a possible father, given a randomly chosen mother-child pair.
Section 5.1 Exclusion probability

“The theoretical concept of exclusion probabilities, however, makes no sense within the framework of normal mixture models.”

“The interpretation of conditional exclusion probability is obvious, which accounts for its value in the legal arena. Unlike [LR], however, it is not fully efficient.”

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.”

Impact of Dropping Loci

• The less data available for comparison purposes, the greater the chance of falsely including someone who is truly innocent

• Are you then being “conservative” (i.e., erring in favor of the defendant)?

Curran and Buckleton (2010)

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.”

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
Some Important Points

• Inclusionary statements (including "cannot exclude") need statistical support to reflect the relevant weight-of-evidence.
• Stochastic thresholds are necessary if using CPI statistics to help identify possible allele dropout.
• CPI is only conservative for guilty suspects as this approach does a poor job of excluding the innocent.
• Uncertainty exists in scientific measurements – this fact needs to be conveyed with the statistical results.
• An increasing number of poor samples are being submitted to labs – labs may benefit from developing a complexity threshold.

Some Mixture Examples Were Provided

• Case 1
  – Evidence (sexual assault victim's underwear bra)
  – Victim
  – Suspect
• Case 2
  – Evidence (sexual assault victim's panties)
• Case 3
  – Evidence (burglary cigarette)

Observations from this Evidence Profile

• The sample is a mixture since there are >2 peaks at multiple loci (at least 2 contributors).
• Two contributors is a reasonable assumption since there are no more than four alleles at a single locus.
• Male and female DNA are present based on amelogenin X/Y ratio.
• A major contributor is not easily discernible so component deconvolution is not an option.
• Results at 4-allele loci (DSS818, FGA, and D16S539) suggest ≈1:1 mixture ratio.
• Overall RFU signals are low especially for larger loci D2S1338 and D6S51 so allele drop-out is a possibility.

Case 1 sexual assault victim's underwear (bra)

Taiwan Case 1 Evidence: Full Profile (Identifiler)

Is "13" a stutter of "14"?
(61/453) = 13.5%
SLIDES NOT COMPLETED YET on PROVIDED MIXTURE EXAMPLES

Analytical Threshold (Peaks vs. Noise)

Stutter Threshold (Alelles vs. Artifacts)

Assumptions based upon # of contributors

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%)

Determination of Mixture Ratio

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

Minor component:
= (112 + 152)/1477 = 0.179

Major component:
= (616 + 597)/1477 = 0.821

≈ 4.6 : 1
### Determination of Genotypes (PHR)

**Possible Combinations**

- 13, 14 and 15, 16  
  (36%)  (15%)
- 13, 15 and 14, 16  
  (31%)  (17%)
- 13, 16 and 14, 15  
  (48%)  (85%)

Includes "stutter" from the 14 allele.

### Determination of Mixture Ratio

**Total of all peak heights**

\[= 213 + 589 + 689 + 103 = 1594 	ext{ RFUs} \]

**Minor component**

\[= \frac{(13^+ + 16^+)}{1594} = 0.198 \]

**Major component**

\[= \frac{(14^+ + 15^+)}{1594} = 0.802 \]

\[\approx 4 : 1 \]

### Application of the Mixture Ratio

**Using peak height ratio, all genotypes possible:**

- 12,12 + 13,14  
  1:1.6
- 13,13 + 12,14  
  1:3.3
- 14,14 + 12,13  
  1:1.6
- 12,13 + 12,14  
  1:1.4
- 12,13 + 13,14  
  1:1
- 12,14 + 13,14  
  1:1.4

**Is there a major:minor here?**

**Using MIXTURE RATIO calculations, can eliminate genotype pairs**

### 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)

**“Inferred Genotype” Approach**

- Random Match Probability (RMP)
- Combined Prob. of Exclusion (CPE)
- Likelihood Ratio (LR)

We conclude that the two matters that appear to have real force are:

1. LRs are more difficult to present in court and
2. the RMNE statistic wastes information that should be utilised.
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 where alleles are below stochastic levels cannot be used in the CPI statistic.

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>D2</td>
</tr>
<tr>
<td>D3</td>
<td>D7</td>
</tr>
<tr>
<td>D19</td>
<td>TH01</td>
</tr>
<tr>
<td>TPOX</td>
<td>D13</td>
</tr>
<tr>
<td></td>
<td>D16</td>
</tr>
<tr>
<td></td>
<td>FGA</td>
</tr>
</tbody>
</table>

If CPI/CPE Stats are Used

• CPI statistics using FBI Caucasian Frequencies
  • 1 in 71 Caucasians included
  • 98.59% Caucasians excluded

If RMP/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.

RMP - 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>

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

If Assume 2 Contributors....

Major  Minor
8,8     11,8 OR 11,11

RMP = 8,11 + 11,11
RMP = 2pq + (q^2 + q(1-q)\theta)

RMP = 2(0.5443)(0.2537) + (0.2537)^2 + (0.2537)(0.7463)(0.01)
= 0.3424 or 1 in 2.9

If RMP/LR Stats are Used

Can use
D8
D21
D18
D3
D19
TPOX
FGA
CSF

Loci with potential D-out
D7
D2
TH01
vWA
D13
D5
D16

Challenges with low level, complex mixtures

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.
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!

Assuming Three Contributors…

Some Possible Contributions to This Result

- 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)

\[ \text{CPI} = (0.53 + 0.24)^2 = 0.59 \text{ or } 59\% \]

D5S818 Allele Frequencies (NIST Caucasian, Butler et al. 2003)

\[ \text{CPI} = (0.05 + 0.38)^2 = 0.18 \text{ or } 18\% \]

Combine loci = 0.59 x 0.18 = 0.11 or 11%

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

Impact of Amplifying More DNA

D19S433

Allele 12 is missing

True Contributors

<table>
<thead>
<tr>
<th>15,15 (2x)</th>
<th>14,15 (1x)</th>
<th>12,14 (1x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125 pg total DNA amplified</td>
<td>500 pg total DNA amplified</td>
<td></td>
</tr>
</tbody>
</table>

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 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

ISFG Recommendations on Mixture Interpretation

http://www.isfg.org/Publication/Gill2006

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
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
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."


---

“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 Effect Diagram](image1.png)

---

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.”

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

Analyze
Data
Server

D19S433
All Peaks above 10 RFU are considered

True Allele Casework Workflow
5 Modules
- Analyze
- Data
- Request
- State Assumptions
  2, 3, 4 unknowns
  1 Unk with Victim?
- Set Parameters
  MCMC modeling
  (e.g., 50K)
- Degradation?
- Computation

Analyze
Data
Request
Review

Computation

Server
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

Review of 3 person mixture

≈12% minor “A”

≈13% minor “B”

Width of the spread is Related to determining the Uncertainty of the mix ratios

Mixture Weight

Determining the LR for D19S433

Suspect A = 14, 16.2

$H_P = 0.967$

$\begin{array}{c|c}
\text{Allele Pair} & \text{Probability Before Conditioning} \\
14, 16.2 & 0.967 \\
14, 14 & 0.003 \\
13, 16.2 & 0.026 \\
13, 14 & 0.001 \\
\end{array}$

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

$H_D$
### Combined LR = 5.6 Quintillion

<table>
<thead>
<tr>
<th>Locus</th>
<th>allele</th>
<th>Genotype Probability Distribution</th>
<th>Weighted Likelihood</th>
<th>Likelihood Ratio</th>
<th>log(LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF1PO</td>
<td>12, 12</td>
<td>0.686</td>
<td>1</td>
<td>0.68615</td>
<td>0.68615</td>
</tr>
<tr>
<td>D19S433</td>
<td>13, 14</td>
<td>0.988</td>
<td>1</td>
<td>0.99952</td>
<td>0.99952</td>
</tr>
<tr>
<td>D19S433</td>
<td>14, 16</td>
<td>0.967</td>
<td>0.012</td>
<td>0.96715</td>
<td>0.96715</td>
</tr>
<tr>
<td>D21S11</td>
<td>28, 30</td>
<td>0.998</td>
<td>0.027</td>
<td>0.99809</td>
<td>0.99809</td>
</tr>
<tr>
<td>D21S11</td>
<td>30, 30</td>
<td>0.984</td>
<td>0.016</td>
<td>0.98383</td>
<td>0.98383</td>
</tr>
<tr>
<td>D2S1338</td>
<td>23, 24</td>
<td>0.988</td>
<td>0.012</td>
<td>0.98799</td>
<td>0.98799</td>
</tr>
<tr>
<td>D5S818</td>
<td>11, 12</td>
<td>0.915</td>
<td>0.085</td>
<td>0.91237</td>
<td>0.91237</td>
</tr>
<tr>
<td>D5S818</td>
<td>12, 12</td>
<td>0.984</td>
<td>0.016</td>
<td>0.98383</td>
<td>0.98383</td>
</tr>
<tr>
<td>D7S820</td>
<td>11, 12</td>
<td>0.998</td>
<td>0.027</td>
<td>0.99809</td>
<td>0.99809</td>
</tr>
<tr>
<td>D8S1179</td>
<td>13, 14</td>
<td>0.203</td>
<td>0.9</td>
<td>0.20267</td>
<td>0.20267</td>
</tr>
<tr>
<td>FGA</td>
<td>21, 25</td>
<td>0.32</td>
<td>0.35</td>
<td>0.31398</td>
<td>0.31398</td>
</tr>
<tr>
<td>TH01</td>
<td>7, 7</td>
<td>0.887</td>
<td>0.139</td>
<td>0.88861</td>
<td>0.88861</td>
</tr>
<tr>
<td>TPOX</td>
<td>8, 8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>vWA</td>
<td>15, 20</td>
<td>0.998</td>
<td>0.005</td>
<td>0.99808</td>
<td>0.99808</td>
</tr>
</tbody>
</table>

**Results**

- Results are expressed as logLR values

\[
LR = 1,000,000 = 10^6
\]

\[
\log(LR) = \log 10^6 = 6 \cdot \log(10) = 6
\]

- Combined LR = 5.6 Quintillion

---

### Review of One Replicate (of 50K)

- **D19S433**
  - 3P mixture,
  - 3 Unknowns
  - 150 RFU

- **No Conditioning (3 Unknowns)**
  - Major contributor = 75%
  - Pr = 1

### No Conditioning (3 Unknowns)

- **D19S433**
  - 8.1% for the two contributors

- **Suspect “A” Genotype**
  - 39 probable genotypes
Suspect A = 14, 16.2  \[ H_P = 0.013 \]

<table>
<thead>
<tr>
<th>Allele Pair</th>
<th>Probability</th>
<th>Frequency</th>
<th>Genotype</th>
<th>Prob *</th>
<th>GenFreq</th>
</tr>
</thead>
<tbody>
<tr>
<td>13, 14</td>
<td>0.002</td>
<td>0.1082</td>
<td>0.00020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14, 14</td>
<td>0.002</td>
<td>0.0498</td>
<td>0.00008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13, 14, 16.2</td>
<td>0.017</td>
<td>0.0392</td>
<td>0.00068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14, 16.2</td>
<td>0.013</td>
<td>0.0120</td>
<td>0.00016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13, 16.2</td>
<td>0.018</td>
<td>0.0131</td>
<td>0.00023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>0.00385</td>
<td></td>
<td>H_D</td>
<td>0.013</td>
<td>3.38</td>
</tr>
</tbody>
</table>

No Conditioning (3 Unknowns)

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

\[ D19S433 \]

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/\mu L using Identifiler (standard conditions).
- Each sample was amplified twice.

Mixture Data Set

- Three different combinations:
  - "Low" Sharing
  - "Medium" Sharing
  - "High" Sharing


Match Score in Duplicate Runs

"Easy" for Deconvolution
Exploring the Capabilities

- Degree of Allele Sharing
- Mixture Ratios
- DNA Quantity
"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

Contributor B (green) (16%)
Contributor A (66%)
Contributor C (blue) (18%)

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 Frequency</th>
<th>Suspect</th>
<th>Numerator</th>
<th>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>0.0795</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, 12</td>
<td>0.306</td>
<td>0.2133</td>
<td>1</td>
<td>0.0612</td>
<td>0.03669</td>
<td>0.0795</td>
</tr>
<tr>
<td></td>
<td>10, 12</td>
<td>0.12</td>
<td>0.1547</td>
<td></td>
<td>0.01861</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10563</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>0.11488</td>
<td>8.704</td>
<td></td>
</tr>
<tr>
<td>D8S1179</td>
<td>11, 16</td>
<td>0.098</td>
<td>0.0109</td>
<td>1</td>
<td>0.99766</td>
<td>0.0059</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>Numerator</th>
<th>Denominator</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>11, 13</td>
<td>0.073</td>
<td>0.0088</td>
<td>1</td>
<td>0.07318</td>
<td>0.0086</td>
<td>0.00596</td>
</tr>
<tr>
<td></td>
<td>11, 14</td>
<td>0.014</td>
<td>0.0271</td>
<td>1</td>
<td>0.0271</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 14</td>
<td>0.006</td>
<td>0.0996</td>
<td>1</td>
<td>0.0996</td>
<td>0.00065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 14</td>
<td>0.011</td>
<td>0.0506</td>
<td>1</td>
<td>0.0506</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 13</td>
<td>0.005</td>
<td>0.1115</td>
<td>1</td>
<td>0.1115</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, 12</td>
<td>0.018</td>
<td>0.0303</td>
<td>1</td>
<td>0.0303</td>
<td>0.0054</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 14</td>
<td>0.004</td>
<td>0.0271</td>
<td>1</td>
<td>0.0271</td>
<td>0.00012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 13</td>
<td>0.003</td>
<td>0.0916</td>
<td>1</td>
<td>0.0916</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 15</td>
<td>0.005</td>
<td>0.0296</td>
<td>1</td>
<td>0.0296</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 15</td>
<td>0.001</td>
<td>0.0319</td>
<td>1</td>
<td>0.0319</td>
<td>0.0003</td>
<td></td>
</tr>
</tbody>
</table>

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>HP</th>
<th>HD</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>0.0498</td>
<td>0.00279</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 14</td>
<td>0.007</td>
<td>0.0596</td>
<td>0.00066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 14</td>
<td>0.011</td>
<td>0.0606</td>
<td>0.00058</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, 14</td>
<td>0.021</td>
<td>0.0721</td>
<td>0.00056</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 13</td>
<td>0.006</td>
<td>0.1115</td>
<td>0.00066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 14</td>
<td>0.005</td>
<td>0.0771</td>
<td>0.00013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 15</td>
<td>0.001</td>
<td>0.0379</td>
<td>0.00002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 15</td>
<td>0.001</td>
<td>0.0424</td>
<td>0.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td>etc...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10, 15</td>
<td>0</td>
<td>0.0227</td>
<td>0.00001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The match rarity between the evidence and suspect is 9.16 thousand

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).

Contact Information

John M. Butler  
NIST Fellow & Applied Genetics  
Group Leader  
john.butler@nist.gov  
301-975-4049

Thank you for your attention

Acknowledgments: NIJ & FBI Funding  
Many slides from Mike Coble (NIST)

Our team publications and presentations are available at:  
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm