DNA Mixture Interpretation

John M. Butler, Ph.D.
National Institute of Standards and Technology
Gaithersburg, Maryland
Acknowledgments and Disclaimers

Funding for research and training on forensic DNA performed by the NIST Applied Genetics Group has come from the National Institute of Justice and the NIST Law Enforcement Standards Office.

Although I chaired the SWGDAM Mixture Committee that produced the 2010 STR Interpretation Guidelines, I cannot speak for or on behalf of the Scientific Working Group on DNA Analysis Methods.

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial 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 National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.
Steps in Forensic DNA Testing

- **Collection/Storage/Characterization**
- **Extraction/Quantitation**
- **Amplification/Marker Sets**
- **Separation/Detection**
- **Interpretation**
- **Report**

Sample Collection & Storage:
- Blood Stain
- Buccal swab

DNA Extraction & Quantitation

- Multiplex PCR Amplification of STR Markers

- CE with LIF Detection

- Data Interpretation, Review & Reporting

- Male: 13, 14-15, 16-...

Equipment:
- GeneAmp 9700 Thermal Cycler
- ABI 3500 Genetic Analyzer
- GeneMapper ID-X software

**capillary electrophoresis**
Current scientific thinking:
• ~99.9% of 6 billion letters (2 x 3 billion bp) are the same between people
• This 0.1% is still ~6 million differences

Father contributes: 22 autosomes (1 of each pair), X or Y
Mother contributes: 22 autosomes (1 of each pair), X and mtDNA
Punnett Square Showing Possible Genotype Combinations (from Genetic Inheritance)

Parental Alleles →
Child Genotypes

<table>
<thead>
<tr>
<th>father</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>AA</td>
<td>AB</td>
</tr>
<tr>
<td>B</td>
<td>AB</td>
<td>BB</td>
</tr>
</tbody>
</table>

Allele Frequencies

<table>
<thead>
<tr>
<th>father</th>
<th>p</th>
<th>q</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p^2</td>
<td>pq</td>
</tr>
<tr>
<td>q</td>
<td>pq</td>
<td>q^2</td>
</tr>
</tbody>
</table>

pq + pq = 2pq

Observed Data

Calculated Statistics
Short Tandem Repeat (STR) Typing

Fluorescent dye-labeled primer

STR Repeat Region

forward primer hybridization region

reverse primer hybridization region

DNA Separation and Detection

RFUs

1000 500

75….80….100….120….140….160….180….200….220….240….260…..

139bp

147bp

1000 500

RFUs
Understanding an STR Electropherogram (E-gram; EPG)

**Peak height** correlates to amount of DNA present (signal detected)
**Peak position** relates to the DNA size, which corresponds to STR allele repeat #
**Peak color** relates to the fluorescent dye label used to copy the specific DNA target

<table>
<thead>
<tr>
<th>Repeat # (allele)</th>
<th>Peak height (relative fluorescence units)</th>
<th>DNA size (nucleotides relative to size std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>2451</td>
<td>106.2</td>
</tr>
</tbody>
</table>

**Alleles** (peaks) are detected - but **Genotypes**, the specific combination of alleles, matter in terms of identifying individuals
Maternal and paternal allele are both 16 so the signal is twice as high.

Possible combinations at D3S1358 include:
- 14, 17 with 16,16
- 14,14 with 16,17
- 14,16 with 17,17

Multiple possible combinations could have given rise to the mixture observed here.
These results are from a DNA test called Identifiler. A single-source (reference) sample displays only 1 or 2 peaks per DNA site.
DNA Mixture Result

Controlled mixture of 4 individuals

Data courtesy of Catherine Grgicak (Boston U.)

More than two peaks per locus (DNA test site)
# Different DNA Tests from Various STR Kits

<table>
<thead>
<tr>
<th>Kit Name</th>
<th># STR Loci Tested</th>
<th>Manufacturer</th>
<th>Why Used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifiler, Identifiler Plus*</td>
<td>15 autosomal STRs (aSTRs) &amp; amelogenin</td>
<td>Life Technologies (Applied Biosystems)</td>
<td>Covers the 13 core CODIS loci plus 2 extra</td>
</tr>
<tr>
<td>PowerPlex 16 PowerPlex 16 HS*</td>
<td>15 aSTRs &amp; amelogenin</td>
<td>Promega Corporation</td>
<td>Covers the 13 core CODIS loci plus 2 extra</td>
</tr>
<tr>
<td>Profiler Plus &amp; COFiler (2 different kits)</td>
<td>13 aSTRs [9 + 6 with 2 overlapping] &amp; amelogenin</td>
<td>Life Technologies (Applied Biosystems)</td>
<td>Original kits used to provide 13 CODIS STRs</td>
</tr>
<tr>
<td>Yfiler</td>
<td>17 Y-chromosome STRs</td>
<td>Life Technologies (Applied Biosystems)</td>
<td>Male-specific DNA test</td>
</tr>
<tr>
<td>MiniFiler</td>
<td>8 aSTRs &amp; amelogenin</td>
<td>Life Technologies (Applied Biosystems)</td>
<td>Smaller regions examined; helps with degraded DNA samples</td>
</tr>
<tr>
<td>GlobalFiler*</td>
<td>21 aSTRs, DYS391, Y indel, &amp; amelogenin</td>
<td>Life Technologies (Applied Biosystems)</td>
<td>Addresses future US core loci</td>
</tr>
<tr>
<td>PowerPlex Fusion*</td>
<td>22 aSTRs, DYS391, &amp; amelogenin</td>
<td>Promega Corporation</td>
<td>Addresses future US core loci</td>
</tr>
</tbody>
</table>

*Newer kits* that contain improved PCR buffers and DNA polymerases to yield more sensitive results and recover data from difficult samples.
Three Possible Outcomes of Evidence Examination

- **Exclusion** (no match)

- **Inclusion** (match)

- **Inconclusive result**

  - No result (or a complex mixture)

  Unable to make Q → K comparison
DNA Mixture Basics


- Mixtures arise when two or more individuals contribute to the sample being tested.

- Mixtures can be challenging to detect and interpret without extensive experience and careful training.

- Differential extraction can help distinguish male and female components of many sexual assault mixtures.

Even more challenging with poor quality data when degraded DNA is present…

Y-chromosome markers can help here in some cases…
Mixtures: Issues and Challenges


- The probability that a mixture will be detected improves with the use of more loci and genetic markers that have a high incidence of heterozygotes.

- The detectability of multiple DNA sources in a single sample relates to the ratio of DNA present from each source, the specific combinations of genotypes, and the total amount of DNA amplified.

- Some mixtures will not be as easily detectable as other mixtures.
Sources of DNA Mixtures

• **Two (or more) individuals** contribute to the biological evidence examined in a forensic case (e.g., sexual assault with victim and perpetrator or victim, consensual sexual partner, and perp)

  *Victim Reference and Spouse or Boyfriend Reference*

• **Contamination** of a single source sample from
  – evidence collection staff
  – laboratory staff handling the sample
  – Low-level DNA in reagents or PCR tubes or pipet tips

  *Examine Staff Profiles (Elimination Database), etc.*

Reference elimination samples are useful in deciphering both situations due to possibility of intimate sample profile subtraction
Mixture Example
Comparing Alleles Only

Mixed stain

<table>
<thead>
<tr>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Locus 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 16 17 18</td>
<td>12 13 14</td>
<td>10 11 12</td>
</tr>
</tbody>
</table>

Reference (e.g., Defendant)

<table>
<thead>
<tr>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Locus 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 16</td>
<td>12 14</td>
<td>11</td>
</tr>
</tbody>
</table>
Mixture Example
Showing Importance of Using Peak Height Information

Mixed stain

Yes, the reference alleles are present in the evidence mixed stain
BUT the peak height patterns do not fit (for a 2-person mixture)...
Mixture Example
Solving Components Prior to Comparison to Suspect Reference

Reference (defendant) does not match either component of the mixed stain and therefore could not have contributed to the evidence sample (assuming 2-contributors)
Mixture Example
Different Evidence Sample…

Possibilities include
10,10 with 11,12
11,11 with 10,12
12,12 with 10,11

Here the Reference (defendant) does match solved Component 1 of the mixed stain and therefore could have contributed to the evidence sample
Unrestricted vs. Restricted Genotype Combinations

Use of peak height information to select only certain combinations

Unrestricted

All combinations of alleles are deemed possible (relative peak height differences are not utilized)

$\text{AB} + \text{AC} + \text{AD} + \text{BC} + \text{BD} + \text{CD}$

Restricted

Based on relative peak heights, alleles are paired only where specific combinations of alleles are deemed possible

$\text{AB} + \text{AC} + \text{AD} + \text{BC} + \text{BD} + \text{CD}$

Peak Height Ratios Are Used in Mixture Component Deconvolution (Restricting Possible Genotypes)

Better Explanation of the Data
(assuming 2 contributors)

13,14 and 15,16

13,16 (major) and 14,15 (minor)
Uncertainty with Possible Genotypes

Genotype 9,13 is likely the major contributor (assuming a 2-person mixture)

The 11 allele is at 166 RFU (above a 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 can be 9,11 or 11,11 or 11,13

If we also include the 8 and 12 alleles in creating our genotype combinations, then the minor contributor possible genotypes expands to include 8,11 and 11,12

Slide adapted from Mike Coble (NIST)
Whatever way uncertainty is approached, probability is the *only* sound way to think about it.  
*Understanding Uncertainty*, p. 71

- Dennis Lindley

Wiley (2007)
Approaches to Data Interpretation: Binary vs Probabilistic

**Binary Approach**

0  
Genotype absent

1  
Genotype present

We want our results to be black and white

**Probabilistic Approach**

0  
Genotype absent

1  
Genotype present

Whereas *our reality is 50 shades of grey (a continuum of possibilities)*
Allele Drop-out

• If because of chemistry events sometimes associated with low levels of DNA (termed “stochastic effects”), one of the STR alleles “drop-out” and is not detected, then our sample at that locus looks like a homozygote instead of the heterozygote that it really is.

- True heterozygote (both peaks detected)  
- True homozygote (only a single peak)  
- False homozygote (one peak has “dropped out” and fails to be detected)

\[ p^2 \quad 2pq \quad 2p \]
Likelihood Ratios for Different Possibilities

Evidence

Suspect

Evidence

Suspect

Evidence

Can allele drop-out explain the missing data?

“2p”

Binary LR approach (either 0 or 1)

\[
LR = \frac{1}{2pq}
\]

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

\[
LR = \frac{?}{2pq}
\]
Probabilistic Genotyping Involves Exploring Multiple Possibilities to See Which One Best Fits the Data

Mixture Data (Evidence)

Thousands of computer simulations are performed to see which model is the best fit

Slide adapted from Mike Coble (NIST)
SWGDAM Interpretation Guidelines

SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories

SWGDAM = Scientific Working Group on DNA Analysis Methods (http://www.swgdam.org/)

- Approved January 14, 2010
SWGDAM Mixture Interpretation Guidelines (2010)

• Provide guidance to labs for interpreting single-source and two-person mixtures

• **NOT** intended for Low Template DNA or >2 person mixtures

• Guidelines – **NOT** Standards

• Laboratories are not required to follow, but guidelines are **STRONGLY RECOMMENDED**

• Require statistics when DNA inclusions are made (SWGDAM 2010 section 4.1)
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.**”

“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)

Statistical Approaches with Mixtures


1. **Random Match Probability or RMP (after inferring genotypes of contributors)** – Separate major and minor components into individual profiles and compute the random match probability estimate as if a component was from a single source

2. **Combined Probability of Exclusion/Inclusion** – CPE/CPI (RMNE) – Calculation of the probability that a random (unrelated) person would be excluded/included as a contributor to the observed DNA mixture

   - **RMNE** = Random Man Not Excluded (same as CPI)
   - **CPE** = Combined Probability of Exclusion (CPE = 1 – CPI)
   - **CPI** = Combined Probability of Inclusion (CPI = 1 – CPE)

3. **Likelihood Ratio (LR)** – Compares the probability of observing the mixture data under two alternative hypotheses; in its simplest form

   \[
   LR = \frac{Pr(E \mid H_1)}{Pr(E \mid H_2)}
   \]
Assumptions for CPE/CPI Approach

• **There is no allele dropout** (i.e., all alleles are above stochastic threshold) – low-level mixtures can not reliably be treated with CPE

• All contributors are from the same racial group (i.e., you use the same allele frequencies for the calculations)

• **All contributors are unrelated**

• Peak height differences between various components are irrelevant (i.e., component deconvolution not needed) – this may not convey all information from the available sample data…
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 for statistical purposes

• If alleles are seen below the established stochastic threshold, then the locus is typically eliminated ("INC" – declared inconclusive) in many current lab SOPs
Overview of Two Thresholds

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

- **30 RFUs**
  - Peak not considered reliable

**Analytical Threshold**
Minimum threshold for data comparison and peak detection in the DNA typing process

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

RFU = relative fluorescence units and is the measure of signal detected

Example values (empirically determined based on own internal validation)

Yfiler (Y-STR) Data from a Knife Handle
(1970s post-conviction case)
The lab initially issued a report stating that the results excluded the defendant.

Several months later, the lab changed its assessment and issued a new report:

"The partial Y-STR profile obtained ... is a mixture consistent with originating from two males. No determination can be made as to whether or not [Defendant] is a contributor to this mixture."
Uncertainty in Evidence Result Leads to “Inconclusive” Report

• In my opinion, a high degree of uncertainty in the number of contributors (Y-STR loci with multiple alleles) and the true DNA types (due to extensive allele and locus dropout) makes comparison of this sample to ANY reference sample problematic.

• If evidence cannot be compared due to poor quality data, then the defendant cannot be excluded (and potentially exonerated) based on DNA results…

• Poor quality DNA data (as well as potential, inadvertent contamination) may present challenges with reaching any conclusions on older Innocence Project cases…
New Statistical Tools/Software for Mixtures

- **Lab Retriever** (David Balding → Norah Rudin et al.)
  - Uses likelihood ratios (LRs) and probability of dropout [Pr(D) or P(Do)]

- **FST** – Forensic Statistical Tool (NYC OCME)
  - Uses LRs and empirically determined Pr(D) based on DNA quantity

- **Armed Xpert** (USACIL → Niche Vision)
  - Originally developed by US Army Crime Lab (USACIL)
  - Performs calculations typically manually done by analysts

- **TrueAllele** (Mark Perlin/Cybergenetics)
  - Uses probabilistic genotyping approach with LRs
Lab Retriever Program

Beta-version is available for free download from www.scieg.org

Scientific Article - describes the math and statistical model


David Balding likeLTD – program written in R (computer language)

https://sites.google.com/site/baldingstatisticalgenetics/software/likeltd-r-forensic-dna-r-code

Norah Rudin and colleagues – prepare a GUI for likeLTD to make it more user-friendly

http://www.scieg.org/lab_retriever.html
FST (Forensic Statistical Tool)

Currently undergoing a Frye admissibility hearing in NYC

- “…FST does not deconvolute DNA mixtures, but simply computes a LR for scenarios specified by the user, allowing for mismatches between contributors’ profiles and the DNA alleles labeled in the mixtures. The mismatches are accounted for by incorporating drop-out and drop-in probabilities in the LR calculation. While FST uses empirically determined drop-out and drop-in rates, [other programs] require the user to specify drop-out and drop-in probabilities…”

Armed Xpert

http://www.armedxpert.com/

- Developed by the US Army Crime Lab (USACIL) initially as a Virtual Basic program called “DNA_DataAnalysis”

- Enables RMP, CPI, and LR calculations for 2-person and 3-person mixtures

- Plan to incorporate probability of drop-out models developed by John Buckleton (New Zealand)
True Allele Casework

http://www.cybgen.com/systems/casework.shtml

- Performs thousands of simulations to model mixture data
- Calculates a combined likelihood ratio
- A commercial product so not all of the mathematical details have been published
- Has been admitted in several states including PA and CA
- Validation work published with NYSP (JFS Nov 2011)
Probabilistic Modeling of TrueAllele

Mathematical Modeling of the Data

Typically 50,000 or 100,000 Simulations Performed (MCMC)

Probable Genotypes to explain the mixture

PHR, Mix Ratio, Stutter etc…

D16S539

- Quantitative computer interpretation using numerous Markov Chain Monte Carlo (MCMC) simulations
- Models peak uncertainty and infers possible genotypes
- Results are presented as the Combined LR

<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>8,9</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>…</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>
DNA Case Example

- Portions of **redacted results and lab report** were kindly provided by Olga Akselrod (Innocence Project)

- **Three pieces of evidence (mixtures)** plus victim and defendant DNA profiles to enable Q→K comparisons
  - Fingernail clippings (right hand & left hand) and jeans

- Testing was performed using **MiniFiler**
  - 8 STR loci + amelogenin (sex-typing marker)
  - MiniFiler is a miniSTR test that aids recovery of results from damaged DNA because it examines smaller portions of the DNA molecules than other STR typing kits

- Statistical analysis of mixtures were performed using **CPI** (combined probability of inclusion) and **FBI Popstats** computer program
Right hand fingernail clipping: **1 in 1965** (8 loci used)
Left hand fingernail clipping: **1 in 358** (7 loci used)
Jeans: **1 in 16** (5 loci used)
The Minifiler DNA profile obtained from the right hand fingernail clippings (Item #2.16) is a mixture of DNA from at least three individuals, including at least one male and one female individual.

The DNA profile of DEFENDANT (Item #5) cannot be excluded from the DNA in the mixture.

For the [MiniFiler] loci D21S11, …, the probability of randomly selecting an unrelated individual as a possible contributor to the DNA profile of the mixture at the [MiniFiler loci] is at least 1 in 1,965 for U.S. individuals. …
MiniFiler Green Channel
Right Hand Fingernail Clippings (Item #2.16)

Amelogenin
Victim: X,X
Defendant: X,Y

D2S1338
Victim: 19,21
Defendant: 21,22

D21S11
Victim: 30,37
Defendant: 29,31
<table>
<thead>
<tr>
<th>ITEM</th>
<th>Report #1</th>
<th>Report #1</th>
<th>Report #1</th>
<th>Report #1</th>
<th>Report #1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Item #2.16</td>
<td>Item #2.16</td>
<td>Item #2.16</td>
<td>Item #2.16</td>
<td>Item #2.36</td>
</tr>
<tr>
<td></td>
<td>Right Hand</td>
<td>Left Hand</td>
<td>Jeans</td>
<td>Fingernail Clippings</td>
<td>Dried Red Stain</td>
</tr>
<tr>
<td></td>
<td>Minifiler</td>
<td>Minifiler</td>
<td>Minifiler</td>
<td>Minifiler</td>
<td>on Bra From</td>
</tr>
<tr>
<td>D21S11</td>
<td>29,30,31,33,2,37</td>
<td>30,31,37</td>
<td>NT</td>
<td>NT</td>
<td>14,15</td>
</tr>
<tr>
<td>D7S820</td>
<td>8,9,10,11</td>
<td>8,10,11</td>
<td>NT</td>
<td>NT</td>
<td>10,11</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>7,8,10,11</td>
<td>7,10,11,12</td>
<td>NT</td>
<td>NT</td>
<td>10,11</td>
</tr>
<tr>
<td>D3S1358</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>7,10</td>
</tr>
<tr>
<td>TH01</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>16,17</td>
</tr>
<tr>
<td>D13S317</td>
<td>11,12,13</td>
<td>11,12,13</td>
<td>NT</td>
<td>NT</td>
<td>15,15</td>
</tr>
<tr>
<td>D16S539</td>
<td>9,11,12,13</td>
<td>9,11,12,13</td>
<td>9,11,12,13</td>
<td>9,11,12,13</td>
<td>11,12</td>
</tr>
<tr>
<td>D2S1338</td>
<td>19,21,22</td>
<td>17,19,21,22</td>
<td>16,19,21,23,24</td>
<td>19,21</td>
<td></td>
</tr>
<tr>
<td>D19S433</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>14,15</td>
</tr>
<tr>
<td>vWA</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>16,17</td>
</tr>
<tr>
<td>TPOX</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>8,11</td>
</tr>
<tr>
<td>D18S51</td>
<td>10,15,16,20</td>
<td>10,15,16,17,20</td>
<td>10,12,14,15,16,17,19</td>
<td>15,16</td>
<td></td>
</tr>
<tr>
<td>Amelos</td>
<td>X,Y</td>
<td>X,Y</td>
<td>X,Y</td>
<td>X,Y</td>
<td>X,Y</td>
</tr>
<tr>
<td>D5S818</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>11,12</td>
</tr>
<tr>
<td>FGA</td>
<td>20,21,23,24,25</td>
<td>20,23,24</td>
<td>20,21,23,24</td>
<td>20,24</td>
<td>12,12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Victim**

**Defendant**

**Known Identifier**

| D8S1179 | 14,115 | 14,14 | 29,31 |
| D7S820  | 10,11  | 8,11  | 7,10  |
| CSF1PO  | 16,17  | 15,15 | 15,15 |
| D3S1358 | 7,9,3  | 8,9   | 8,9   |
| TH01    | 12,13  | 11,11 | 11,11 |
| D13S317 | 11,12  | 9,13  | 9,13  |
| D16S539 | 19,21  | 21,22 | 21,22 |
| D2S1338 | 14,15  | 12,2,13| 12,2,13|
| D19S433 | 16,17  | 14,16 | 14,16 |
| vWA     | 8,11   | 9,9   | 9,9   |
| TPOX    | 15,16  | 10,15 | 10,15 |
| D18S51  | X,Y    | X,Y   | X,Y   |
| Amelos  | 11,12  | 12,12 | 12,12 |
| D5S818  | 20,24  | 20,23 | 20,23 |
Item #2.16 (Right Hand Fingernail Clippings)

Information in Report Table
D21S11: 29, 30, 31, 33.2, 37*

Victim: 30, 37
Defendant: 29, 31

Electropherogram (mixture data observed)

Graphical representation of DNA data at D21S11

Based on results at this single locus, we can assume at least three individuals contributed to the DNA results because there are more than 4 alleles.
Source of the Numbers for Right Hand Fingernail Clippings

Using D21S11 alleles 29, 30, 31, 33.2, and 37 in statistical calculations

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>CAU</th>
<th>BLK</th>
<th>SWH</th>
</tr>
</thead>
<tbody>
<tr>
<td>D21S11</td>
<td>29</td>
<td>0.1811</td>
<td>0.1899</td>
<td>0.2044</td>
</tr>
<tr>
<td>D21S11</td>
<td>30</td>
<td>0.2321</td>
<td>0.1788</td>
<td>0.3301</td>
</tr>
<tr>
<td>D21S11</td>
<td>31</td>
<td>0.0714</td>
<td>0.0922</td>
<td>0.069</td>
</tr>
<tr>
<td>D21S11</td>
<td>33.2</td>
<td>0.0306</td>
<td>0.0335</td>
<td>0.0419</td>
</tr>
<tr>
<td>D21S11</td>
<td>37</td>
<td>0.0128</td>
<td>0.014</td>
<td>0.0123</td>
</tr>
</tbody>
</table>

Population Group | CPI Stats Calculated
--- | ---
Caucasian (CAU) | 1 in 26,080
Black (BLK) | 1 in 1,965
Southwest Hispanic (SWH) | 1 in 18,440
MiniFiler Blue Loci for Jeans (Item #2.7)

Allele below ST would invalidate D7S820 locus from use in CPI statistics

Stochastic threshold = 200 RFU?
**CPI Stats Calculations for the Jeans**

D7S820 appears to have been used in CPI statistical calculation.

Only 1 in 16.9 Caucasians
Where Can Potential Errors Occur in DNA Interpretation?

• Incorrect inclusion of an innocent person using allele drop-out as a reason for mismatch between evidence and suspect with a CPI approach

• Inclusion of loci in CPI calculations with alleles below stochastic threshold (CPI requires all alleles to be detected) could lead to an inflation of the match statistic

• Setting thresholds too high and thus losing relevant data that could be used to exclude

• Use of $p^2$ with single peaks (assuming genotype is a homozygote) instead of $2p$ (allowing for allele drop-out) will falsely inflate statistics

• Failure to exclude when alleles are present but genotypes do not fit
Is the Known Individual Included or Excluded?

Known: 13,14

Known: 28,30

Assumptions:
1) 2 contributors and all data are present →
2) 1 major and 1 minor contributor →
3) Major must have 13,16 and 28,28 genotypes and
4) Minor must have 14,15 and 30,32.2 genotypes

Based on these assumptions, the individual is excluded

Genotypes are excluded even if alleles are included
Different Experts $\rightarrow$ Different Opinions

- Are the experts asking/answering the same question?
- Are they using the same information and data?
- Are they using the same interpretation methods?
- Are they using good scientific practices?
- Any possibility of bias?
- Are the differences meaningful or trivial?
Some Thoughts on the Future…

• **PCR amplification**
  – Faster enzymes to enable rapid PCR
  – More robust enzymes and master mixes that work better

• **Instrumentation**
  – More dye colors to aid in analyzing more loci simultaneously
  – Rapid, integrated devices
  – Alternatives to capillary electrophoresis: next-generation sequencing

• **Marker systems**
  – Expanding sets of STR loci for growing DNA databases
  – Other marker systems: SNPs, InDels, X-STRs, RM Y-STRs
  – Body fluid identification using other molecules such as RNA
  – Phenotyping for external visible characteristics
  – Privacy challenges with additional genome information

• **Data interpretation**
  – Probabilistic genotyping for low-level DNA and mixture interpretation
  – Probability of dropout incorporated into DNA data interpretation
DNA Mixture Detected with PowerPlex Fusion (24plex STR kit)

22 autosomal STR loci need to be interpreted...(+50% over current 15 STRs)

Data courtesy of Becky Hill (NIST)
New Efforts to Improve DNA Interpretation (especially low-level DNA and mixtures)

December 2012 – Forensic Science International: Genetics, volume 6, issue 6

Approaches to mixture data interpretation is in a state of change throughout the forensic DNA community
April 12, 2013 Webcast

http://www.nist.gov/oles/forensics/dna-analyst-
training-on-mixture-interpretation.cfm

• 8-hours of DNA mixture interpretation training
• 11 presentations from five different presenters
  – John Butler, Mike Coble, Robin Cotton, Bruce Heidebrecht, Charlotte Word
• 20 poll questions asked via SurveyMonkey (>600 participated)
  – Addressed additional questions sent via email or Twitter
• >1000 participants (almost entire U.S. represented and >10 countries)
• Will be available for viewing or download (by early May) for at least six months (storage costs may limit longer-term storage)
Acknowledgments

Case Examples and Input on This Presentation
• Olga Akselrod (Innocence Project)
• Jennifer Friedman (Los Angeles Public Defender’s Office)

Slides and Discussions on DNA Mixtures
• Mike Coble (NIST Applied Genetics Group)
• Robin Cotton & Catherine Grgicak (Boston U.)
• Bruce Heidebrecht (Maryland State Police)
• Charlotte Word (consultant)

Contact info:
john.butler@nist.gov
301-975-4049