Basic STR Interpretation Workshop

John M. Butler, Ph.D.
Simone N. Gittelson, Ph.D.

National Institute of Standards and Technology
Gaithersburg, Maryland, USA
31 August 2015
Basic STR Interpretation Workshop

John M. Butler, Ph.D.
Simone N. Gittelson, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Intended Audience and Objectives

• **Intended audience:** students and beginning forensic DNA scientists (less than 5 years of experience)

• **Objectives:** To provide easy-to-follow, basic-to-intermediate level information and to introduce key concepts and fundamental literature in STR data and statistical interpretation.

• Participants should expect to come away with an understanding of key concepts related to interpreting single-source samples and simple two-person DNA mixtures and foundational literature to support work with STR data and statistical interpretation.

Instructor: **John M. Butler**

NIST Fellow and Special Assistant to the Director for Forensic Science at the U.S. National Institute of Standards and Technology (NIST) where he has worked for the past two decades to advance use and understanding of STR typing methods. His Ph.D. research, conducted at the FBI Laboratory, involved developing capillary electrophoresis for forensic DNA analysis. The most recent of his five textbooks forms the basis for this workshop.

Phone: +1-301-975-4049
Email: john.butler@nist.gov

Instructor: **Simon N. Gittelson**

Forensic statistician in the NIST Statistical Engineering Division. She conducted her Ph.D. research at the University of Lausanne (Switzerland) in applying probability and decision theory to inference and decision problems in forensic science. She then specialized in the interpretation of DNA evidence during her postdoc at NIST and the University of Washington.

Phone: +1-301-975-4892
Email: simone.gittelson@nist.gov

Resources

  - All figures available on STRBase: [http://www.cstl.nist.gov/strbase/training.htm](http://www.cstl.nist.gov/strbase/training.htm)
- Boston University DNA Mixture Training: [http://www.bu.edu/dnamixtures/](http://www.bu.edu/dnamixtures/)
- STRBase DNA Mixture Information: [http://www.cstl.nist.gov/strbase/mixture.htm](http://www.cstl.nist.gov/strbase/mixture.htm)

Slides Available for Use from Forensic DNA Typing Books

[http://www.cstl.nist.gov/strbase/training.htm](http://www.cstl.nist.gov/strbase/training.htm)
### Workshop Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Module (Instructor)</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>0900-0930</td>
<td>Welcome &amp; Introductions</td>
<td>Review expectations and questions from participants</td>
</tr>
<tr>
<td>0930-1100</td>
<td>Data Interpretation 1 (John)</td>
<td>STR kits, loci, alleles, genotypes, profiles Data interpretation thresholds and models Simple PCR and CE troubleshooting</td>
</tr>
<tr>
<td>1100-1130</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1130-1300</td>
<td>Statistical Interpretation 1 (Simone)</td>
<td>Introduction to probability and statistics STR population data collection, calculations, and use Approaches to calculating match probabilities</td>
</tr>
<tr>
<td>1300-1430</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1430-1600</td>
<td>Data Interpretation 2 (John)</td>
<td>Mixture interpretation: Clayton rules, # contributors Stochastic effects and low-template DNA challenges Worked examples</td>
</tr>
<tr>
<td>1600-1630</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1630-1800</td>
<td>Statistical Interpretation 2 (Simone)</td>
<td>Approaches to calculating mixture statistics Likelihood ratios and formulating propositions Worked examples</td>
</tr>
</tbody>
</table>

### What We Will Not Cover…

- Handling low-template DNA information
- Probabilistic genotyping for DNA mixtures
- Complicated mixtures with >2 contributors
- Y-STRs or other lineage markers
- Kinship analysis and dealing with relatives

Other ISFG 2015 workshops, being held tomorrow, cover more advanced aspects of DNA interpretation:

- **a)** The interpretation of complex DNA profiles using open-source software LRmix Studio and EuroForMix (EFM) - Peter Gill, Hinda Haned, Corina Benschop, Oskar Hansson, Oyvind Bleka
- **b)** Interpretation of complex DNA profiles using a continuous model – an introduction to STRmix™ - John Buckleton, Jo-Anne Bright, Catherine McGovern, Duncan Taylor
- **c)** Kinship analysis - Thore Egeland, Klaas Slooten
Where are you from?

A. Europe  
B. North America  
C. South America  
D. Africa  
E. Asia  
F. Australia/NZ  
G. Other

Your experience with forensic DNA?

A. Student  
B. 0-1 years  
C. 1-2 years  
D. 2-3 years  
E. 3-4 years  
F. 4-5 years  
G. >5 years

Background of Participants…

Without Your Clicker…

1) Your name

2) Where you are from (your organization)

3) What you hope to learn from this workshop

Ask Questions!

• If you feel uncomfortable asking questions in front of everyone, please write your question down and give it to us at a break

• We will read the question and attempt to answer it in front of the group

• Or raise your hand and ask a question at any time!

Greg Matheson on Forensic Science Philosophy

The CAC News – 2nd Quarter 2012 – p. 6  
“Generalist vs. Specialist: a Philosophical Approach”  

• If you want to be a technician, performing tests on requests, then just focus on the policies and procedures of your laboratory. If you want to be a scientist and a professional, learn the policies and procedures, but go much further and learn the philosophy of your profession. Understand the importance of why things are done the way they are done, the scientific method, the viewpoint of the critiques, the issues of bias and the importance of ethics.

D.N.A. Approach to Understanding

• Doctrine or Dogma (why?)
  – A fundamental law of genetics, physics, or chemistry
    • Offspring receive one allele from each parent
    • Stochastic variation leads to uneven selection of alleles during PCR amplification from low amounts of DNA templates
    • Signal from fluorescent dyes is based on …

• Notable Principles (what?)
  – The amount of signal from heterozygous alleles in single-source samples should be similar

• Applications (how?)
  – Peak height ratio measurements can associate alleles into possible genotypes

Adapted from David A. Bednar, Increase in Learning (Deseret Book, 2011)
ISFG 2015: Basic STR Interpretation Workshop
(J.M. Butler & S.N. Gittelson)

Data Interpretation 1:
STR kits, loci, alleles, genotypes, profiles
Data interpretation thresholds and models
Simple PCR and CE troubleshooting

John M. Butler, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Acknowledgment and Disclaimers
I will quote from my recent book entitled “Advanced Topics in Forensic DNA Typing: Interpretation” (Elsevier, 2015). I do not receive any royalties for this book. Completing this book was part of my job last year at NIST.

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.

I have been fortunate to have had discussions with numerous scientists on interpretation issues including Mike Coble, Bruce Heidebrecht, Robin Cotton, Charlotte Word, Catherine Grigcak, Peter Gill, Ian Evett ...

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 Analysis

Gathering the Data
Understanding Results Obtained & Sharing Them

Interpretation

Presentation Outline

1. Data interpretation overview
2. Data collection with ABI Genetic Analyzers
3. STR alleles and PCR amplification artifacts
4. STR genotypes and heterozygote balance
5. STR profiles and tri-allelic patterns
6. ...
7. ...
8. Troubleshooting data collection

These points correspond to chapter numbers in Advanced Topics in Forensic DNA Typing: Interpretation (2015)

http://www.cstl.nist.gov/strbase/training.htm
How Book Chapters Map to Data Interpretation Process

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Input Information</th>
<th>Decision to be made</th>
<th>How decision is made</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Data file</td>
<td>Peak or Noise</td>
<td>Analytical threshold</td>
</tr>
<tr>
<td>3</td>
<td>Peak, Allele or Artifact</td>
<td>Stutter threshold; precision sizing bin</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Allele</td>
<td>Heterozygote or Homozygote or Allele(s) missing</td>
<td>Peak heights and peak height ratios; stochastic threshold</td>
</tr>
<tr>
<td>5</td>
<td>Genotype/full profile</td>
<td>Single-source or Mixture</td>
<td>Numbers of peaks per locus</td>
</tr>
<tr>
<td>6</td>
<td>Mixture</td>
<td>Deconvolution or not</td>
<td>Major/minor mixture ratio</td>
</tr>
<tr>
<td>7</td>
<td>Low level DNA</td>
<td>Interpret or not</td>
<td>Complexity threshold</td>
</tr>
<tr>
<td>8</td>
<td>Poor quality data</td>
<td>Replace CE components (buffer, polymer, array) or call service engineer</td>
<td>Review size standard data quality with understanding of CE principles</td>
</tr>
</tbody>
</table>

Our Backgrounds Influence Our Interpretation

We see the world, not as it is, but as we are – or, as we are conditioned to see it.

- Stephen R. Covey (*The 7 Habits of Highly Effective People*, p. 28)

Challenges in Real-World Data

- **Stochastic (random) variation** in sampling each allele during the PCR amplification process
  - This is highly affected by DNA quantity and quality
  - Imbalance in allele sampling gets worse with low amounts of DNA template and higher numbers of contributors
- **Degraded DNA** template may make some allele targets unavailable
- **PCR inhibitors** present in the sample may reduce PCR amplification efficiency for some alleles and/or loci
- **Overlap of alleles** from contributors in DNA mixtures
  - Stutter products can mask true alleles from a minor contributor
  - Allele stacking may not be fully proportional to contributor contribution

Using Ideal Data to Discuss Principles

Using **Ideal Data** to Discuss Principles

<table>
<thead>
<tr>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Locus 3</th>
<th>Locus 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>14</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>8,8</td>
<td></td>
</tr>
</tbody>
</table>

(1) 100% PHR (Hb) between heterozygous alleles

(2) Homozygotes are exactly twice heterozygotes due to allele sharing

(3) No peak height differences exist due to size spread in alleles (any combination of resolvable alleles produces 100% PHR)

(4) No stutter artifacts enabling mixture detection at low contributor amounts

(5) Perfect inter-locus balance

(6) Completely repeatable peak heights from injection to injection on the same or other CE instruments in the lab or other labs

(7) Genetic markers that are so polymorphic all profiles are fully heterozygous with distinguishable alleles enabling better mixture detection and interpretation

Steps in DNA Interpretation

**Question sample**

**Known sample**

**Weight of Evidence**

**Report Written & Reviewed**

http://www.cstl.nist.gov/strbase/training.htm
Overview of the SWGDAM 2010 Interp Guidelines

See http://www.swgdam.org/

1. Preliminary evaluation of data – is something a peak and is the analysis method working properly?
2. Allele designation – calling peaks as alleles
3. Interpretation of DNA typing results – using the allele information to make a determination about the sample
   1. Non-allelic peaks
   2. Application of peak height thresholds to allelic peaks
   3. Peak height ratio
   4. Number of contributors to a DNA profile
   5. Interpretation of DNA typing results for mixed samples
   6. Comparison of DNA typing results
4. Statistical analysis of DNA typing results – assessing the meaning (rarity) of a match

Other supportive material: statistical formulae, references, and glossary

Have you read the 2010 SWGDAM STR Interpretation Guidelines?

1. Yes
2. No
3. Never heard of them before!
4. What’s SWGDAM?

Overview of Data Interpretation Process

Questions for Workshop Participants

• STR kits in your lab?
  – Examples: Identifiler, NGM SElect, PP16, PP21

• CE instrument(s)?
  – Examples: ABI 310, ABI 3130xl, ABI 3500

• Analysis software?
  – Examples: GeneMapperID, GMID-X, GeneMarkerHD

Autosomal STR Kit(s) in Your Laboratory?

1. Identifiler or ID Plus
2. NGM or NGM Select
3. GlobalFiler
4. PowerPlex 16 or 16HS
5. PowerPlex ESX or ESI 16/17
6. PowerPlex Fusion
7. Qiagen
8. More than one

CE Instrumentation in Your Laboratory?

A. ABI 3130 or 3130xl
B. ABI 3100
C. ABI 3500 or 3500xl
D. ABI 310
E. Other
Analysis Software in Your Laboratory?

A. GeneMapper ID
B. GeneMapper ID-X
C. GeneMarker HID
D. Qualilty GeneProof
E. Other

Types of STR Repeat Units

Requires size based DNA separation to resolve different alleles from one another

- Dinucleotide: (CA)(CA)(CA)(CA)
- Trinucleotide: (GCC)(GCC)(GCC)
- Tetranucleotide: (AATG)(AATG)(AATG)
- Pentanucleotide: (AGAAA)(AGAAA)
- Hexanucleotide: (AGTACA)(AGTACA)

Short tandem repeat (STR) = microsatellite = simple sequence repeat (SSR)

Categories for STR Markers

<table>
<thead>
<tr>
<th>Category</th>
<th>Example Repeat Structure</th>
<th>13 CODIS Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple repeats – contain units of identical length and sequence</td>
<td>(GATA)(GATA)(GATA)</td>
<td>TPOX, CSF1PO, D5S818, D13S317, D16S539</td>
</tr>
<tr>
<td>Simple repeats with non-consensus alleles (e.g., TH01 9.3)</td>
<td>(GATA)(CAT)(GATA)</td>
<td>TH01, D18S51, D7S820</td>
</tr>
<tr>
<td>Compound repeats – comprise two or more adjacent simple repeats</td>
<td>(GATA)(GATA)(GACA)</td>
<td>VWA, FGA, D3S1358, D8S1179</td>
</tr>
<tr>
<td>Complex repeats – contain several repeat blocks of variable unit length</td>
<td>(GATA)(GACA)(CA)(CAT)</td>
<td>D2IS11</td>
</tr>
</tbody>
</table>
The CODIS Core Loci Working Group selected a consortium of 11 CODIS laboratories...these laboratories performed validation experiments...

With the assistance of the National Institute of Standards and Technology (NIST), the data generated through these validation studies were compiled, reviewed, and analyzed.

**Value of STR Kits**

**Advantages**
- Quality control of materials is in the hands of the manufacturer (saves time for the end-user)
- Improves consistency in results across laboratories – same allelic ladders used
- Common loci and PCR conditions used – aids DNA databasing efforts
- Simpler for the user to obtain results

**Disadvantages**
- Contents may not be completely known to the user (e.g., primer sequences)
- Higher cost to obtain results

**Different DNA Tests from Various STR Kits**

- **Identifiler**, **Identifiler Plus**: 15 autosomal STRs (amplifiers) + amelogenin
  - ThermoFisher
  - Covers the 13 core CODIS loci plus 2 extra

- **PowerPlex 16**, **PowerPlex 16 HS**: 15 autosomal STRs & amelogenin
  - Promega Corporation
  - Covers the 13 core CODIS loci plus 2 extra

- **Profiler Plus & COFiler** (2 different kits): 13 autosomal STRs + 2 extra
  - ThermoFisher
  - Original kits used to provide 13 CODIS STRs

- **Yfiler**: 17 Y-chromosome STRs
  - ThermoFisher
  - Male-specific DNA test

- **MiniFiler**: 8 autosomal STRs & amelogenin
  - ThermoFisher
  - Male-specific DNA test for degraded DNA recovery

- **GlobalFiler**
  - 21 autosomal STRs, DYS391, Y-STRs, amelogenin
  - Covers expanded U.S. core loci; quality sensors

- **PowerPlex Fusion**
  - 22 autosomal STRs, DYS391, & amelogenin
  - Promega Corporation
  - Covers expanded U.S. core loci; quality sensors

- **Investigator 24plex**
  - 21 autosomal STRs, DYS391, quality sensor, & amelogenin
  - Qiagen
  - Covers expanded U.S. core loci; quality sensors

Newer kits that contain improved PCR buffers and DNA polymerases to yield more sensitive results and recover data from difficult samples.

**U.S. Core Loci Expansion Efforts**

- **U.S. Core Loci Announced**: 1990: 4 VNTRs (RFLP), 1997: 13 STRs (PCR), 2011: 20+ STRs
- **Implementation to be required**: 2 years after announcement
- **PowerPlex Fusion** and GlobalFiler
  - 24plex kits available

**Position of Forensic STR Markers on Human Chromosomes**

- **13 Core U.S. STR Loci**
- **15 STR loci overlap between U.S. and Europe**

**Different DNA Tests from Various STR Kits**

- **ThermoFisher ABI STR Kits**
  - Internal Size Standard: LIZ 500 – 5-plex; LIZ 600 – 5-plex
  - **PowerPlex Fusion**: 24plex

http://www.cstl.nist.gov/strbase/training.htm
**Promega STR Kits** (Internal Size Standard: CXR ILS600 – 4-dye; CXR ILS550 – 5-dye)

- **100 bp**
- **400 bp**
- **300 bp**
- **200 bp**

**Promega STR Kits** (Internal Size Standard: CXR ILS600 – 4-dye; CXR ILS550 – 5-dye)

- **100 bp**
- **400 bp**
- **300 bp**
- **200 bp**

**PowerPlex 16**

- vWA
- TPOX
- D18S51
- Penta E

**PowerPlex ESI 17**

- vWA
- TPOX
- D18S51
- Penta E

**PowerPlex Fusion**

- vWA
- TPOX
- D18S51
- Penta E

**STR Marker Layouts for New U.S. Kits**

- 24plex (5-dye)
- 22 core and recommended loci + 2 additional loci

**STR Kits and Dye Sets Used**

- Example STR Kits |
  - Profiler Plus, SGM Plus, COFlter, Profiler: 5-FAM, JOE, NED, ROX |
  - Identifiler, MiniFiler, NGM, NGM Select: 6-FAM, VIC, NED, PET, LIZ |
  - GlobalFiler: 6-FAM, VIC, NED, TAZ, SID, LIZ (J6) (3500) |
  - PowerPlex 16, 16HS: FL, JOE, TMR, CXR |
  - PowerPlex ESI 16/17, ESX 16/17, 18D, 21, Fusion: FL, JOE, TMR-ET, CXR-ET, CC5 |
  - Qiagen Investigator Kits: B, G, Y, R, D |
  - Research assays: 6-FAM, TET, HEX, ROX

**What is the primary reason for CE instrument sensitivity variation?**

A. Laser strength  
B. CCD camera sensitivity  
C. Optical alignment of laser and detector  
D. All of the above
ISFG 2015: Basic STR Interpretation Workshop
(J.M. Butler & S.N. Gittelson)
31 August 2015

http://www.cstl.nist.gov/strbase/training.htm

Key Points

- On-scale data of STR allele peaks are important to interpretation (both lower and upper limits exist for reliable data)
- Data signals from ABI Genetic Analyzers are processed by proprietary algorithms that include variable binning (adjustment for less sensitive fluorescent dyes), baselining, smoothing, and multi-componenting for separating color channels
- Instrument sensitivities vary due to different lasers, detectors, and optical alignment (remember that signal strength is in "relative fluorescence units", RFUs)

STR Typing Works Best in a Narrow Window of DNA Template Amounts

- Off-scale data with flat-topped peaks
- Allele dropout due to stochastic effects
- “Just right”
- Too much DNA amplified
- Within optimal range
- Typically best results are seen in the 0.5 ng to 1.5 ng range for most STR kits

Critical Points

- On-scale data of STR allele peaks are important to interpretation (both lower and upper limits exist for reliable data)
- Data signals from ABI Genetic Analyzers are processed by proprietary algorithms that include variable binning (adjustment for less sensitive fluorescent dyes), baselining, smoothing, and multi-componenting for separating color channels
- Instrument sensitivities vary due to different lasers, detectors, and optical alignment (remember that signal strength is in "relative fluorescence units", RFUs)

Applied Biosystems (ABI)

A BRIEF HISTORY OF NAME CHANGES FOR APPLIED BIOSYSTEMS

Year Company Name
1986 PE Biosystems
1998 PE Biosystems Group of PE Corporation
2000 Applied Biosystems
2002 Applied Biosystems (ABI)
2008 Life Technologies (after Applied Biosystems merged with Invitrogen)
2014 Life Technologies (after Applied Biosystems merged with Thermo Fisher Scientific)

Typical data “frame”

Data Collection with ABI Prism Genetic Analyzer

- Size Separation
- Sample Separation
- Sample Preparation
- Sample Injection
- Mixture of dye-labeled PCR products from multiplex PCR reaction
- Sample Detection
- Capillary holder with polymer solution
- Processing with GeneMapper® ID software
- Abi Prism spectrograph
- Color Separation
- Fluorescence
- Spatial calibration
- Pixel number
- Spectral calibration
- CCD Camera image
- Even bins
- Larger "red" bin

Depiction of a CCD Camera Image
**Useful Range of an Analytical Method**

- **LOD**: Limit of detection
- **LOQ**: Limit of quantitation
- **LOL**: Limit of linearity
- **LSL**: Limit of sensitivity

**Dilemma of a Threshold in a Continuous World**

(a) Binary Threshold Applied

<table>
<thead>
<tr>
<th>Probability</th>
<th>Threshold</th>
<th>Data Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Too High</td>
<td>Considered reliable</td>
</tr>
<tr>
<td>0</td>
<td>Too Low</td>
<td>Not considered reliable</td>
</tr>
</tbody>
</table>

(b) Continuous Data

Data signal behaves as a continually upward sloping hill.

**Impact of Setting Thresholds Too High or Too Low**

If

- Threshold is set **too high**

  Analysis may miss low-level legitimate peaks (false negative conclusions produced)

- Threshold is set **too low**

  Analysis will take longer as artifacts and baseline noise must be removed from consideration as true peaks during data review (false positive conclusions produced)

**STR Alleles and PCR Amplification Artifacts**

A. Length of the PCR product
B. An allelic ladder
C. Electrophoretic mobility of labeled DNA molecules relative to an internal size standard
D. PCR primer positions

---

http://www.cstl.nist.gov/strbase/training.htm
Key Points

• STR allele designations are made by comparing the relative size of sample peaks to allelic ladder allele sizes
• A common, calibrated STR allele nomenclature is essential in order to compare data among laboratories
• STR allele sizes are based on a measure of the relative electrophoretic mobility of amplified PCR products (defined by primer positions) compared to an internal size standard using a specific sizing algorithm
• STR alleles can vary in their overall length (number of repeat units), with their internal sequence of repeats, and in the flanking region

Identifiler STR Kit Allelic Ladder (internal size standard not shown)

Identifiler data from Boston University (Catherine Grgicak)

Single-Source Sample Profile (1 ng of “C”)

[Diagram showing DNA fragment peaks and allele calls for D18S51]

Transforming Information at a Single STR Locus during Data Processing

[Diagram showing data transformation process]

DNA Size Standard and Sizing Algorithm

(a) GS500-ROX (Applied Biosystems)

(b) [Diagram showing DNA fragment peaks and sizing algorithm]

D18S51 Typing Results on the Same DNA Sample with Three Different STR Kits

(a) Qiagen ESSplex SE Promega PP ESX 17 Applied Biosystems NMG Select

(b) [Diagram showing typing results for D18S51]
Null Alleles

• Allele is present in the DNA sample but fails to be amplified due to a nucleotide change in a primer binding site

• Allele dropout is a problem because a heterozygous sample appears falsely as a homozygote

• Two PCR primer sets can yield different results on samples originating from the same source

• This phenomenon impacts DNA databases

• Large concordance studies are typically performed prior to use of new STR kits

Non-Template Addition

• Taq polymerase will often add an extra nucleotide to the end of a PCR product; most often an “A” (termed “adenylation”)

• Dependent on 5’-end of the reverse primer; a “G” can be put at the end of a primer to promote non-template addition

• This phenomenon impacts DNA databases

• Large concordance studies are typically performed prior to use of new STR kits

Incomplete adenylation

Impact of the 5’ Nucleotide on Non-Template Addition

5’-ACAAG…

Last Base for Primer Opposite Dye Label

(PCR conditions are the same for these two samples)

5’-CCAAG…


Stutter Products

• Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis

• Stutter is less pronounced with larger repeat unit sizes (dinucleotides > tri > tetra > penta)

• Longer repeat regions generate more stutter

• Each successive stutter product is less intense (allele > repeat-1 > repeat-2)

• Stutter peaks make mixture analysis more difficult

Stutter Data from a Set of 345 D18S51 Alleles Measured at NIST Using the PowerPlex 16 Kit

<table>
<thead>
<tr>
<th>Allele</th>
<th>Allele Size (nucleotides)</th>
<th># Measured</th>
<th>Median (%)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>296.9</td>
<td>43</td>
<td>4.8</td>
<td>0.4</td>
</tr>
<tr>
<td>13</td>
<td>300.7</td>
<td>27</td>
<td>5.7</td>
<td>0.5</td>
</tr>
<tr>
<td>14</td>
<td>304.6</td>
<td>35</td>
<td>6.2</td>
<td>0.5</td>
</tr>
<tr>
<td>15</td>
<td>308.5</td>
<td>55</td>
<td>6.9</td>
<td>0.6</td>
</tr>
<tr>
<td>16</td>
<td>312.4</td>
<td>46</td>
<td>7.7</td>
<td>0.5</td>
</tr>
<tr>
<td>17</td>
<td>316.2</td>
<td>47</td>
<td>8.3</td>
<td>0.4</td>
</tr>
<tr>
<td>18</td>
<td>320.2</td>
<td>38</td>
<td>9.0</td>
<td>0.9</td>
</tr>
<tr>
<td>19</td>
<td>324.0</td>
<td>30</td>
<td>9.6</td>
<td>0.9</td>
</tr>
<tr>
<td>20</td>
<td>328.0</td>
<td>24</td>
<td>10.6</td>
<td>0.8</td>
</tr>
<tr>
<td>21</td>
<td>332.0</td>
<td>19</td>
<td>11.3</td>
<td>0.8</td>
</tr>
<tr>
<td>22</td>
<td>336.0</td>
<td>10</td>
<td>12.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Locus Stutter Filter: Average + 3 standard deviations = 7.7 + (3×1.9) = 7.7 + 5.7 = 13.4%
Simplified Illustration of Stutter Trends

<table>
<thead>
<tr>
<th>Repeat Length</th>
<th>% Stutter</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 SD</td>
<td>Tri-nucleotides</td>
</tr>
<tr>
<td>2 SD</td>
<td>Tetra-</td>
</tr>
<tr>
<td>Average</td>
<td>Penta-</td>
</tr>
<tr>
<td></td>
<td>Hexa-</td>
</tr>
</tbody>
</table>

Repeat Length

STR Genotypes
Heterozygote Balance, Stochastic Effects, etc.

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 4

Key Points

- In heterozygous loci, the two alleles should be equal in amount; however, stochastic effects during PCR amplification (especially when the amount of DNA being amplified is limited) create an imbalance in the two detected alleles
- Heterozygote balance (Hb) or peak height ratios (PHRs) measure this level of imbalance
- Under conditions of extreme imbalance, one allele may "drop-out" and not be detected
- Stochastic thresholds are sometimes used to help assess the probability of allele drop-out in a DNA profile

D18S51 Results from Two Samples

<table>
<thead>
<tr>
<th>Individual “C”: 16,18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele 1</td>
</tr>
<tr>
<td>peak height</td>
</tr>
<tr>
<td>peak size</td>
</tr>
<tr>
<td>Allele 2</td>
</tr>
<tr>
<td>peak height</td>
</tr>
<tr>
<td>peak size</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual “D”: 14,20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele 1</td>
</tr>
<tr>
<td>peak height</td>
</tr>
<tr>
<td>peak size</td>
</tr>
<tr>
<td>Allele 2</td>
</tr>
<tr>
<td>peak height</td>
</tr>
<tr>
<td>peak size</td>
</tr>
</tbody>
</table>

Peak Height Ratios (PHRs) or Heterozygote balance (Hb)

728/761 = 0.957 = 95.7%
829/989 = 0.838 = 83.8%
Natural Variation in Peak Height Ratio During Replicate PCR Amplifications

The heights of the peaks will vary from sample-to-sample, even for the same DNA sample amplified in parallel.

Hypothetical Heterozygote Alleles

<table>
<thead>
<tr>
<th>DNA Template Level</th>
<th>Heterozygote Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ng</td>
<td>95%</td>
</tr>
<tr>
<td>0.5 ng</td>
<td>80%</td>
</tr>
<tr>
<td>0.2 ng</td>
<td>60%</td>
</tr>
<tr>
<td>0.1 ng</td>
<td>40%</td>
</tr>
<tr>
<td>0.05 ng</td>
<td>0%</td>
</tr>
</tbody>
</table>

In the extreme, one of the alleles fails to be amplified (this is known as allele drop-out).

Key Points

- Tri-allelic patterns occasionally occur at STR loci (~1 in every 1000 profiles) and are due to copy number variation (CNVs) in the genome.
- The amelogenin gene is found on both the X and Y chromosomes and portions of it can be targeted to produce assays that enable gender identification as part of STR analysis using commercial kits.
- Due to potential deletions of the amelogenin Y region, additional male confirmation markers are used in newer 24plex STR kits.
- Partial profiles can result from low amounts of DNA template or DNA samples that are damaged or broken into small pieces or contain PCR inhibitors.

STR Profiles
Multiplex PCR, Tri-Alleles, Amelogenin, and Partial Profiles

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 5

Multiplex PCR (Parallel Sample Processing)

- Compatible primers are the key to successful multiplex PCR.
- STR kits are commercially available.
- 15 or more STR loci can be simultaneously amplified.
- Challenges to Multiplexing:
  - Primer design to find compatible primers (no program exists).
  - Reaction optimization is highly empirical often taking months.

Advantages of Multiplex PCR:
- Increases information obtained per unit time (increases power of discrimination).
- Reduces labor to obtain results.
- Reduces template required (smaller sample consumed).

Single-Source DNA Sample Exhibiting a TPOX Tri-Allelic Pattern

Not a mixture as all other loci exhibit single-peak homozygotes or balanced two-peak heterozygotes.
Types of Tri-Allelic Patterns

**More common**

(a) Type 1  
(b) Type 2

(1+2≈3)  
(1≈2≈3)

This classification scheme was developed by Tim Clayton and colleagues at the UK Forensic Science Service (Clayton et al. 2004, J. Forensic Sci. 49:1207-1214)

Tri-Allelic Patterns Occur about 1 in 1000 Profiles but the frequency varies across STR loci

Relative Positions Along the Y-Chromosome of Amelogenin (AMEL Y) and Male Confirmation Markers Used in Newer STR Kits

Deletions of the Y-chromosome can encompass >1 Mb around the AMEL Y region (DYS458 from Y-STR kits is often lost in these situations)

Partial Profiles Can Occur from Poor Quality DNA or Low Amounts of DNA Template

PCR inhibition or degraded, damaged DNA templates often result in only the shorter-size PCR products producing detectable signal

Troubleshooting Data Collection

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 8
Key Points

- The better you understand your instrument(s) and how DNA typing data are generated during the PCR process, the better you will be able to troubleshoot problems that arise.
- Three key analytical requirements for capillary electrophoresis instruments are (1) spectral (color) resolution, (2) size (spatial) resolution, and (3) run-to-run precision.
- Salt levels need to be low in samples in order to effectively inject them into a CE instrument.

Potential Issues and Solutions with Multicolor Capillary Electrophoresis

<table>
<thead>
<tr>
<th>Issue</th>
<th>Cause/Result with Failure</th>
<th>Potential Solutions</th>
</tr>
</thead>
</table>
| Spectral resolution (color separation) | High RFU peaks result in bleed through or pull up that create artificial peaks in adjacent dye channel(s) | Inject less DNA into the CE capillary to avoid overloading the detector |}
| Analytical size resolution | Inner capillary wall coating failures result in an inability to resolve closely spaced STR alleles and in some cases incorrect allele calls can be made | Reinject sample (if a bubble causes poor polymer filling for a single run) or replace the pump (if polymer is not being routinely delivered to fully fill the capillaries) |
| Run-to-run precision | Room temperature changes result in sample alleles running differently compared to allelic ladder alleles and false “off-ladder” alleles are generated | Make adjustments to improve room temperature consistency or reinject samples with an allelic ladder run in an adjacent capillary or a subsequent run |

Single-Source DNA Profile Exhibiting Pull-Up Due to Off-Scale Data at Several Loci

At first glance, the results at this one locus may appear to be a mixture.

Impact of Formamide Quality on Peak Shape and Height

\[
\text{Sample Conductivity Impacts Amount Injected} \\
[\text{DNA}_{\text{inj}}] = \frac{\text{Et}(t^2) (\mu_{\text{ep}} + \mu_{\text{eof}})[\text{DNA}_{\text{sample}}] (\lambda_{\text{buffer}})}{\lambda_{\text{sample}}}
\]

\(\text{[DNA}_{\text{inj}}]\) is the amount of sample injected, \(\text{E}\) is the electric field applied, \(t\) is the injection time, \(r\) is the radius of the capillary, \(\mu_{\text{ep}}\) is the mobility of the sample molecules, \(\mu_{\text{eof}}\) is the electroosmotic mobility, \(\lambda_{\text{buffer}}\) is the buffer conductivity, and \(\lambda_{\text{sample}}\) is the sample conductivity.

Cl- ions and other buffer ions present in PCR reaction contribute to the sample conductivity and thus will compete with DNA for injection onto the capillary.
ISFG 2015: Basic STR Interpretation Workshop  
(J.M. Butler & S.N. Gittelson)  
31 August 2015

Impact of Polymer Filling the Capillary

(a) Incomplete fill → a “meltdown” → poor-quality data and resolution loss

(b) Appropriate fill → high-quality data and sharp peaks → reliable STR typing

Impact of Sample Renaturation

Double-stranded DNA (dsDNA) molecules, which are more rigid than their corresponding single-stranded DNA (ssDNA) counterparts, migrate more quickly through the network of polymer strands inside of a capillary.

When CE conditions permit re-hybridization of the complementary strand, then a shadow peak occurs in front of its corresponding labeled STR allele (or internal size standard DNA fragment).

Extra Peaks Due to Sample Renaturation

(issue mostly like due to the CE instrument temperature control)

Data Interpretation Overview

The Steps of Data Interpretation

Peak (vs. noise)  Allele (vs. artifact)  Genotype (allele pairing)  Profile (genotype combining)

Analytical Threshold  Expected Stutter %  Stochastic Threshold  Peak Height Ratio (PHR)

Moving from individual locus genotypes to profiles of potential contributors to the mixture is dependent on mixture ratios and numbers of contributors

Elements Going into the Calculation of a Rarity Estimate for a DNA Sample

Population allele frequencies  DNA Profile (with specific alleles)  Rarity estimate of DNA profile

There are different ways to express the profile rarity

Acknowledgments

$ NIST Special Programs Office  
Becky (Hill) Steffen

Slides and Discussions on DNA Mixtures

Mike Coble (NIST Applied Genetics Group)  
Robin Cotton & Catharina Gigiuk (Boston U.)  
Bruce Halblieb (Maryland State Police)  
Charlotte Word (consultant)

Contact info:  
john.butler@nist.gov  
+1-301-975-4049

Final version of this presentation will be available at:  
http://www.cstl.nist.gov/strbase/training.htm

http://www.cstl.nist.gov/strbase/training.htm
**Statistical Interpretation 1:**

Introduction to probability and statistics
STR population data collection, calculations, and use
Approaches to calculating match probabilities

Simone N. Gittelson, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

---

**Presentation Outline**

1. Why do we need to do a statistical (probabilistic) interpretation?
2. How do we do a statistical interpretation?
   a. Hardy-Weinberg Equilibrium (HWE)
   b. Recombination and Linkage
   c. Subpopulations
   d. Linkage Equilibrium (LE)
   e. NRC II Report Recommendations
   f. Population Allele Frequencies
   g. Logical Approach for Evidence Interpretation

---

**Acknowledgement and Disclaimers**

I thank John Butler for the discussions and advice on preparing this presentation. I also acknowledge John Buckleton and Bruce Weir for all their helpful explanations on forensic genetics topics.

Points of view in this presentation are mine and do not necessarily represent the official position or policies of 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.

---

**Workshop Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Module (Instructor)</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>0900-0930</td>
<td>Welcome &amp; Introductions</td>
<td>Review expectations and questions from participants</td>
</tr>
<tr>
<td>0930-1100</td>
<td>Data Interpretation 1 (John)</td>
<td>STR kits, loci, alleles, genotypes, profiles Data interpretation thresholds and models Simple-PCR and CE troubleshooting</td>
</tr>
<tr>
<td>1100-1130</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1130-1300</td>
<td>Statistical Interpretation 1 (Simone)</td>
<td>Introduction to probability and statistics STR population data collection, calculations, and use Approaches to calculating match probabilities</td>
</tr>
<tr>
<td>1300-1430</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1430-1600</td>
<td>Data Interpretation 2 (John)</td>
<td>Mixture interpretation: Clayton rules, # contributors Stochastic effects and low-template DNA challenges Worked examples</td>
</tr>
<tr>
<td>1600-1630</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1630-1800</td>
<td>Statistical Interpretation 2 (Simone)</td>
<td>Approaches to calculating mixture statistics Likelihood ratios and formulating propositions Worked examples</td>
</tr>
</tbody>
</table>

---

**Why do we need to do a statistical (probabilistic) interpretation?**

DNA recovered on the crime scene

http://www.cstl.nist.gov/strbase/training.htm
Does the DNA recovered on the crime scene come from the person of interest?

If the DNA recovered on the crime scene comes from the person of interest, we would expect to see peaks for the same genotypes.

If the DNA recovered on the crime scene does not come from the person of interest, we need to know how rare it is to observe the peaks in the EPG of the DNA recovered on the crime scene.

In this case, the observations support the proposition that the DNA recovered on the crime scene came from the person of interest.
DNA recovered on the crime scene

Everyone in the population of potential donors has this observed DNA profile.

DNA of a person of interest

In this case, the observations provide no information on whom the DNA recovered on the crime scene comes from.

A statistical interpretation tells us what our observations mean in a particular case, with regard to a particular question of interest to the court.

**statistical interpretation**

synonym: probabilistic interpretation

definition: A quantitative expression of the value of the evidence.

**How do we do a statistical interpretation?**

**Elements required for a statistical interpretation**

1. DNA profile data (e.g., observed alleles)
2. Appropriate assumptions, models and formulae
3. Population allele frequencies
4. Statistical interpretation of the observations

Hardy-Weinberg Equilibrium (HWE)

Godfrey Harold Hardy
British mathematician

Wilhelm Weinberg
German physician


April 5, 1908: Date of Hardy’s signature in his July 10, 1908 publication in Science 28 (706): 49-50, entitled Mendelian Proportions in a Mixed Population.


Godfrey Harold Hardy
British mathematician

Wilhelm Weinberg
German physician

Hardy Weinberg Equilibrium (HWE)

Assumptions:
1. size of population is infinite
2. no migration
3. random mating
4. no mutations
5. no natural selection

Why is HWE important?
Allele and genotype frequencies in this population remain constant from one generation to the next.

Hardy Weinberg Equilibrium (HWE)

Assumptions:
1. size of population is infinite

Reality:
world = 7.3 billion
Poland = 38.5 million
Krakow = 760,000

Statistics from:

Hardy Weinberg Equilibrium (HWE)

Assumptions:
2. no migration

Reality:
Poland (2013)
220,300
60% Polish, 13% EU, 27% non-EU

Population 1

Population 2

Statistics from:

Hardy Weinberg Equilibrium (HWE)

Assumptions:
3. random mating

Reality:
“...most common model of marriage is between people from the same age group (49.1%) and also similar economical status and especially similar education level (53.4%).”

Quote and statistics from:

Hardy Weinberg Equilibrium (HWE)

Assumptions:
4. no mutations

D21S11: father {28,28}

Mutation rate for locus D21S11: 0.19%

Mutation rate from:

http://www.cstl.nist.gov/strbase/training.htm
**Assumptions:**

5. no natural selection

**Reality:**

Some genes are more likely to lead to diseases than others. However, STR loci used in forensic science come from regions that are not used for coding genes (i.e., they are introns).

**D21S11:**

- allele 28

**Hardy Weinberg Equilibrium (HWE)**

Assumptions:

1. size of population is infinite
2. no migration
3. random mating
4. no mutations
5. no natural selection

If a population is in Hardy-Weinberg Equilibrium, the Hardy-Weinberg Law predicts the genotype frequencies.

### Laws of Mendelian Genetics

- **Law of Segregation**
  - The genotype at a locus consists of one maternal allele and one paternal allele. Each child receives a randomly selected allele from each parent.

- **Law of Independent Assortment**
  - The allele transmitted from parent to child at one locus is independent of the allele transmitted from parent to child at a different locus.

### Hardy Weinberg Equilibrium (HWE)

**Punnett Square:**

- **mother**
  - a
  - A
  - B

- **father**
  - a
  - A
  - B

- **homozygote (28, 28)**
  - Pr(28, 28) = probability that a person has genotype (28, 28)

- **heterozygote (13, 16)**
  - Pr(13, 16) = probability that a person has genotype (13, 16)

### Laws of Probability

- **Certain:**
  - 1 = certainty that statement is true
  - EXAMPLE: rolling a 6-sided die
    - Pr(1, 2, 3, 4, 5 or 6) = 1
    - Pr(7) = 0

- **Impossible:**
  - 0 = certainty that statement is false

Laws of Probability

**Law #3 (independent events):** The probability of event A and event B occurring is equal to the probability of event A times the probability of event B.

\[ \Pr(A \text{ and } B) = \Pr(A) \times \Pr(B) \]

**EXAMPLE:** rolling two 6-sided dice
A: rolling a 5 with die 1
B: rolling a 5 with die 2
\[ \Pr(A \text{ and } B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \]

---

Population allele frequencies

<table>
<thead>
<tr>
<th>Allele</th>
<th>Caucasian</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.2</td>
<td>-</td>
<td>-</td>
<td>0.002</td>
<td>-</td>
</tr>
<tr>
<td>25.2</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>26.2</td>
<td>-</td>
<td>-</td>
<td>0.002</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>0.022</td>
<td>0.075</td>
<td>0.028</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>0.159</td>
<td>0.246</td>
<td>0.100</td>
<td>0.057</td>
</tr>
<tr>
<td>28.2</td>
<td>-</td>
<td>-</td>
<td>0.005</td>
<td>-</td>
</tr>
<tr>
<td>29</td>
<td>0.203</td>
<td>0.205</td>
<td>0.208</td>
<td>0.201</td>
</tr>
<tr>
<td>30</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>39</td>
<td>-</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Hardy-Weinberg Law**

Homozygote
\[ \Pr(28,28) = \Pr(\text{paternal allele} = 28) \times \Pr(\text{maternal allele} = 28) \]
\[ \Pr(28,28) = p_{28} \times q_{28} \]

**EXAMPLE:**
father
\[ p_{28} = 0.159 \]
mother
\[ q_{28} = 0.841 \]
\[ \Pr(28,28) = (0.159)^2 = 0.025 \]

---

**Laws of Probability**

**Law #2 (mutually exclusive events):** The probability of event A or event B occurring is equal to the probability of event A plus the probability of event B.

\[ \Pr(A \text{ or } B) = \Pr(A) + \Pr(B) \]

**EXAMPLE:** rolling a 6-sided die
A: rolling an odd number
B: rolling a 2
\[ \Pr(A \text{ or } B) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3} \]

---

**Hardy-Weinberg Law**

Heterozygote
\[ \Pr(13,16) = \Pr(\text{paternal allele} = 13) \times \Pr(\text{maternal allele} = 16) \]
\[ + \Pr(\text{paternal allele} = 16) \times \Pr(\text{maternal allele} = 13) \]
\[ = p_{13} \times p_{16} + q_{13} \times q_{16} \]
\[ = p_{13} \times p_{16} + (1 - p_{13}) \times (1 - p_{16}) \]

**EXAMPLE:**
father
\[ p_{13} = 0.722 \]
mother
\[ q_{16} = 0.722 \]
\[ \Pr(13,16) = 0.722 \times 0.722 + 0.722 \times 0.278 = 0.516 \]

---

According to the Hardy-Weinberg law, what is the probability that a person has genotype \{8,12\}?

A. \(0.144 \times 0.159 = 0.023\)
B. \(2 \times 0.144 \times 0.159 = 0.046\)
C. \(2 \times 0.159 \times 0.159 = 0.051\)
D. \(0.144 + 0.159 = 0.303\)
E. \(2 \times 8 \times 12 = 192\)
F. ???

According to the Hardy-Weinberg law, what is the probability that a person has genotype \{12,12\}?

A. \((0.36)^2 = 0.130\)
B. 0.360
C. 0.36 + 0.36 = 0.720
D. \((12)^2 = 144\)
E. ???

### Recombination and Linkage


**Recombination**

- Paternal DNA
- Maternal DNA
- Recombination
- Gamete DNA 1
- Gamete DNA 2

**Linkage**

- Independence between loci
- Linkage

If the child inherits \(B\), there is a probability of 0.5 that the child inherits \(C\), and a probability of 0.5 that the child inherits \(c\).

If the child inherits \(B\), there is a probability >0.5 that the child inherits \(A\), and a probability of <0.5 that the child inherits \(a\).
Is there linkage?

**Loci on different chromosomes, or on different arms of the same chromosome:**
No, there is no linkage.

**Loci on the same arm of the same chromosome:**
Linkage is possible. This has no impact on unrelated individuals, but should be taken into account for related individuals by incorporating the probability of recombination into the statistical interpretation.

2

**Hardy-Weinberg Equilibrium (HWE)**

Assumptions:
- size of population is infinite
- no migration
- random mating
- no mutations
- no natural selection

2

Subpopulations


http://www.cstl.nist.gov/strbase/training.htm
Subpopulations

General Population

\[ p_{28} = 0.5 \]

50% Subpopulation 1
50% Subpopulation 2

- Subpopulation 1 mates only with members of subpopulation 1
- Subpopulation 2 mates only with members of subpopulation 2

\[ p_{28} = 0.4 \]
\[ p_{28} = 0.6 \]

Not taking into account subpopulations:

General Population

\[ p_{28} = 0.5 \]

\[ Pr(28,28) = 0.25 \times 0.25 = 0.0625 \]

Taking into account subpopulations:

\[ Pr(28,28) = \frac{1}{2} \times 0.16 + \frac{1}{2} \times 0.36 = 0.26 \]

We can use the coancestry coefficient \( F_{ST} \), also called \( \theta \), to take into account the effect of subpopulations when we use the proportion \( p_{28} = 0.5 \) of the general population.
Subpopulations

Allele 28 is identical by state and identical by descent.

The coancestry coefficient $F_{ST}$, also called $\theta$, is the probability that two individuals have an allele identical by descent (IBD).

Subpopulations

Profile probability: probability of observing this profile in a population.

Match probability: probability of observing this profile in a population knowing that this profile has already been observed in one individual in this population.

What is the probability of observing this profile in this population?

If $\theta > 0$: profile probability < match probability

If $\theta = 0$: no relatives, no coancestors

Profile probability = match probability

Subpopulations

We have seen: allele 28

The probability of observing an allele 28 is:

$$\theta + (1 - \theta)p_{28}$$

Where allele 28 is IBD with 28, allele 28 is not IBD with any of the alleles already seen, it is observed by chance.

Subpopulations

Rule of Thumb

If the allele in question has not been seen previously, then it is seen by chance.

If the allele in question has already been seen, then it could be observed again by chance or because it is IBD with an allele that has already been seen.
Subpopulations

What is the probability of seeing allele 28 in this population given that we have already observed allele 28 and allele 28?

Subpopulations

We have seen: allele 28 and allele 28

The probability of observing an allele 28 is:

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

allele 28 is IBD with allele 28
allele 28 is IBD with allele 28
allele 28 is not IBD with any of the alleles already seen, it is observed by chance

Subpopulations

We have seen: allele 28 and allele 28

The probability of observing an allele 28 is:

\[ \frac{2\theta + (1 - \theta)p_{28}}{1 + \theta} \]

Subpopulations

We have seen: allele 28, allele 28 and allele 28

The probability of observing an allele 28 is:

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

allele 28 is IBD with allele 28
allele 28 is IBD with allele 28
allele 28 is IBD with allele 28
allele 28 is not IBD with any of the alleles already seen, it is observed by chance

http://www.cstl.nist.gov/strbase/training.htm
Subpopulations

We have seen: allele 28, allele 28 and allele 28

The probability of observing an allele 28 is:

\[
\frac{3\theta + (1 - \theta)p_{28}}{1 + 2\theta}
\]

Subpopulations

What is the probability of seeing genotype \{28, 28\} in this population given that we have already observed a genotype \{28, 28\}? If \(\theta = 0.02\):

\[
\frac{2\theta + (1 - \theta)p_{28}}{1 + \theta} \times \frac{3\theta + (1 - \theta)p_{28}}{1 + 2\theta}
\]

Subpopulations

What is the genotype probability equal to if \(\theta = 0\)?

A. 0
B. \(\theta\)
C. \(p_{28}^2\)
D. 2\(p_{28}\)
E. 1
F. ???

Subpopulations

What is the probability of seeing allele 13 in this population given that we have already observed allele 13 and allele 16?

Subpopulations

http://www.cstl.nist.gov/strbase/training.htm
We have seen: *allele 13* and *allele 16*

The probability of observing an *allele 13* is:

\[
\begin{align*}
1 \times \theta & + 0 \times \theta + (1 - \theta)p_{13} \\
\text{allele 13 is IBD with 13} & \quad \text{allele 13 is not IBD with any of the alleles already seen, it is seen by chance} \\
1 + \theta &
\end{align*}
\]

We have seen: *allele 13, allele 16* and *allele 13*

The probability of observing an *allele 16* is:

\[
\begin{align*}
0 \times \theta & + 0 \times \theta + 1 \times \theta + (1 - \theta)p_{16} \\
\text{allele 16 is IBD with 16} & \quad \text{allele 16 is IBD with 13} \quad \text{allele 16 is not IBD with any of the alleles already seen} \\
1 + 2\theta &
\end{align*}
\]

We have seen: *allele 13, allele 16* and *allele 15*

The probability of observing an *allele 16* is:

\[
\begin{align*}
2 \times & \theta + (1 - \theta)p_{13} \times \theta + (1 - \theta)p_{16} \\
1 + \theta & \quad 1 + 2\theta
\end{align*}
\]

We have seen: *allele 13* and *allele 16*

The probability of observing an *allele 15* is:

\[
\begin{align*}
\theta & + (1 - \theta)p_{13} \\
1 + \theta &
\end{align*}
\]

We have seen: *allele 13, allele 16* and *allele 15*

The probability of observing an *allele 15* is:

\[
\begin{align*}
\theta & + (1 - \theta)p_{13} \\
1 + \theta &
\end{align*}
\]
Subpopulations

What is the probability of seeing genotype \{13, 16\} in this population given that we have already observed a genotype \{13, 16\}?

\[
2 \times \frac{\theta + (1-\theta)p_{13}}{1 + \theta} \times \frac{\theta + (1-\theta)p_{16}}{1 + 2\theta}
\]

if \(\theta = 0.02\):

\[
2 \times \frac{0.02 + (1-0.02)(0.33)}{1 + 0.02} \times \frac{0.02 + (1-0.02)(0.33)}{1 + 2(0.02)}
\]

\(P_{13} = 0.330\)
\(P_{16} = 0.033\)

What is the genotype probability equal to if \(\theta = 0\)?

A. 0
B. 2\(\theta\)
C. \(p^2_{13}\)
D. \(2p_{13}p_{16}\)
E. 1
F. ???

Response Counter

Linkage Equilibrium (LE)

In a population, the alleles at one locus are independent of the alleles at a different locus.

Linkage Disequilibrium

In a population, the alleles at one locus are not independent of the alleles at a different locus.

Linkage ≠ Linkage Disequilibrium

CAUSES: Linkage Subdivision

EFFECT: Linkage Disequilibrium

Linkage Equilibrium (LE)

Assumptions:
1. size of population is infinite
2. no migration
3. random mating
4. no mutations
5. no selection
6. number of generations is infinite

Why is LE important?
Mendel's Law of Independent Assortment holds.

If a population is in Linkage Equilibrium, the product rule predicts the genotype frequencies.
Linkage Equilibrium

In a population, the alleles at one locus are independent of the alleles at a different locus.

Product rule:
probability of genotype at multiple loci
= product of genotype probabilities at each locus

Assumption: Linkage Equilibrium

Product rule:

\[ Pr(\{13,16\}, \{28,28\}, \{8,12\} \text{ and } \{12,12\}) = Pr(13,16) \times Pr(28,28) \times Pr(8,12) \times Pr(12,12) \]

What is the probability that a person has genotype \{13,16\}, \{28,28\}, \{8,12\} and \{12,12\}?

A. \[0.025 \times 0.022 \times 0.046 \times 0.130 = 3.3 \times 10^{-6}\]
B. \[0.130 - 0.025 - 0.022 - 0.046 = 0.037\]
C. \[0.025 + 0.022 + 0.046 + 0.130 = 0.223\]

NRC II Report Recommendations


Fixation indices (\(F\)-statistics)

<table>
<thead>
<tr>
<th>(F)-statistics</th>
<th>Alternative notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_{IS})</td>
<td>Individual to Subpopulation</td>
<td>the correlation of alleles within an individual within a subpopulation</td>
</tr>
<tr>
<td>(F_{IT})</td>
<td>Individual to Total population</td>
<td>the correlation of alleles within an individual (“inbreeding”)</td>
</tr>
<tr>
<td>(F_{ST})</td>
<td>Subpopulation to Total population</td>
<td>the correlation of alleles of different individuals in the same subpopulation (“coancestry”)</td>
</tr>
</tbody>
</table>


NRC II Report Recommendations

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardy-Weinberg Law</td>
<td>Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the population</td>
</tr>
<tr>
<td>Includes possibility that the individual's two alleles are IBD (“inbreeding”).</td>
<td>Corrects for Hardy-Weinberg Disequilibrium in the population caused by population subdivision. Assumes Linkage Equilibrium in the population.</td>
</tr>
<tr>
<td>Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the sub-populations.</td>
<td>Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the sub-populations.</td>
</tr>
</tbody>
</table>


http://www.cstl.nist.gov/strbase/training.htm


### NRC II Report Recommendations

<table>
<thead>
<tr>
<th>Hardy-Weinberg Law:</th>
<th>Homozygotes</th>
<th>Heterozygotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Includes possibility that the individual’s two alleles are IBD (&quot;inbreeding&quot;):</td>
<td>$p_1^2$</td>
<td>$2p_1p_2$</td>
</tr>
<tr>
<td>$F = 0.02$:</td>
<td>$0.025$</td>
<td>$0.022$</td>
</tr>
</tbody>
</table>

### NRC II Report Recommendations

<table>
<thead>
<tr>
<th>Hardy-Weinberg Law:</th>
<th>Homozygotes</th>
<th>Heterozygotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Includes possibility that the individual’s two alleles are IBD (&quot;inbreeding&quot;):</td>
<td>$\theta = 0.02$:</td>
<td>$0.040$</td>
</tr>
</tbody>
</table>

---

### Population Allele Frequencies

#### D8S1179:

<table>
<thead>
<tr>
<th>Allele</th>
<th>Total 2N=2072</th>
<th>2N = 722</th>
<th>2N = 684</th>
<th>2N = 472</th>
<th>2N = 194</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total #</td>
<td>Total %</td>
<td>Caucasian</td>
<td>Black</td>
<td>Hispanic</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>1.06</td>
<td>0.014</td>
<td>0.007</td>
<td>0.0148</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0.48</td>
<td>0.006</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>10</td>
<td>163</td>
<td>7.87</td>
<td>0.102</td>
<td>0.031</td>
<td>0.093</td>
</tr>
<tr>
<td>11</td>
<td>139</td>
<td>6.71</td>
<td>0.076</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>12</td>
<td>294</td>
<td>14.20</td>
<td>0.168</td>
<td>0.130</td>
<td>0.129</td>
</tr>
<tr>
<td>13</td>
<td>556</td>
<td>26.80</td>
<td>0.330</td>
<td>0.219</td>
<td>0.273</td>
</tr>
<tr>
<td>14</td>
<td>484</td>
<td>23.40</td>
<td>0.166</td>
<td>0.298</td>
<td>0.263</td>
</tr>
<tr>
<td>15</td>
<td>291</td>
<td>14.00</td>
<td>0.104</td>
<td>0.190</td>
<td>0.129</td>
</tr>
<tr>
<td>16</td>
<td>101</td>
<td>4.87</td>
<td>0.033</td>
<td>0.064</td>
<td>0.032</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>0.39</td>
<td>0.001</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>0.19</td>
<td>-</td>
<td>0.003</td>
<td>0.002</td>
</tr>
</tbody>
</table>

---

### ISFG 2015: Basic STR Interpretation Workshop

(J.M. Butler & S.N. Gittelson)

http://www.cstl.nist.gov/strbase/training.htm
Steps in generating and validating a population database

1. Decide on number of samples and ethical/racial grouping
2. Gather samples
3. Analyze samples at desired genetic loci
4. Determine allele frequencies for each locus


Validating a Population Database

Key question: What are you going to do with the database?

Validating a database requires validating the population genetic model that will be used.

Statistical tests examine whether the data in the database performs as expected according to the population genetic model.


Sampling Variation

If we profiled everyone, the true frequencies are:

- $f_A = 0.015$
- $f_B = 0.550$
- $f_C = 0.075$
- $f_D = 0.025$

Population database frequencies:

- $f_A = 0$
- $f_B = 0.500$
- $f_C = 0.044$
- $f_D = 0.022$

close, but not the same

Slide courtesy to Dr. John Buckleton
**Options**

Factor of 10 and minimum allele frequency of 5/2N


- Multiply match probability by 10
- Use 5/2N as the minimum allele frequency

**Population allele frequencies**

D8S1179:

<table>
<thead>
<tr>
<th>Allele</th>
<th>Total</th>
<th>2N = 2072</th>
<th>2N = 722</th>
<th>2N = 472</th>
<th>2N = 194</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total #</td>
<td></td>
<td>0.014</td>
<td>0.007</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0.48</td>
<td>0.004</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>163</td>
<td>7.87</td>
<td>0.311</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>139</td>
<td>6.71</td>
<td>0.053</td>
<td>0.053</td>
<td>0.119</td>
</tr>
<tr>
<td>12</td>
<td>294</td>
<td>14.20</td>
<td>0.130</td>
<td>0.129</td>
<td>0.119</td>
</tr>
<tr>
<td>13</td>
<td>556</td>
<td>26.80</td>
<td>0.219</td>
<td>0.273</td>
<td>0.201</td>
</tr>
<tr>
<td>14</td>
<td>484</td>
<td>23.40</td>
<td>0.298</td>
<td>0.263</td>
<td>0.201</td>
</tr>
<tr>
<td>15</td>
<td>291</td>
<td>14.00</td>
<td>0.190</td>
<td>0.129</td>
<td>0.129</td>
</tr>
<tr>
<td>16</td>
<td>101</td>
<td>4.87</td>
<td>0.036</td>
<td>0.032</td>
<td>0.093</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>0.39</td>
<td>0.004</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>0.19</td>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Logical Approach for Evidence Interpretation**

**Framing the question**

Different questions have different answers.

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Profile probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the probability of observing this profile in the population?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 2</th>
<th>Match probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the probability of observing this profile in the population if we have already observed one person with this profile in this population?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 3</th>
<th>Combined probability of inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the probability that a person selected randomly in the population would be included (or not excluded) as a possible donor of the DNA?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 4</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>By how much do the DNA typing results support the person of interest being the donor?</td>
<td></td>
</tr>
</tbody>
</table>

**Likelihood Ratio (LR)**

The probability of observing the DNA typing results given that the prosecution's proposition is true divided by the probability of observing the DNA typing results given that the defense's proposition is true.

\[
\frac{\Pr(E|H_p)}{\Pr(E|H_d)}
\]
Logical Framework for Updating Uncertainty

Odds form of Bayes' theorem:

\[
\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}
\]

posterior odds
prior odds

Likelihood Ratio

4

Likelihood Ratio (LR)

\[
\frac{\Pr(E|H_p)}{\Pr(E|H_d)}
\]

if \( \Pr(E|H_p) > \Pr(E|H_d) \)

- 1

0

if \( \Pr(E|H_p) < \Pr(E|H_d) \)

0

Acknowledgements

John Butler

Slides and discussions on forensic genetics
John Buckleton and Bruce Weir

Contact Info.: simone.gittelson@nist.gov
+1-301-975-4892

Final version of this presentation will be available at:

http://www.cstl.nist.gov/strbase/NISTpub.htm
Data Interpretation 2:
Mixture interpretation: Clayton rules, # contributors
Stochastic effects and low-template DNA challenges
Worked examples

John M. Butler, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

<table>
<thead>
<tr>
<th>Time</th>
<th>Module (Instructor)</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>0900-0930</td>
<td>Welcome &amp; Introductions</td>
<td>Review expectations and questions from participants</td>
</tr>
<tr>
<td>0930 – 1100</td>
<td>Data Interpretation 1 (John)</td>
<td>STR kits, loci, alleles, genotypes, profiles, Data interpretation thresholds and models, Simple PCR and CE troubleshooting</td>
</tr>
<tr>
<td>1100 – 1130</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1130 – 1300</td>
<td>Statistical Interpretation 1 (Simone)</td>
<td>Introduction to probability and statistics, STR population data collection, calculations, and use Approaches to calculating match probabilities</td>
</tr>
<tr>
<td>1300 – 1330</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>1430 – 1600</td>
<td>Data Interpretation 2 (John)</td>
<td>Mixture interpretation: Clayton rules, # contributors, Stochastic effects and low-template DNA challenges, Worked examples</td>
</tr>
<tr>
<td>1600 – 1630</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>1630 – 1800</td>
<td>Statistical Interpretation 2 (Simone)</td>
<td>Approaches to calculating mixture statistics, Likelihood ratios and formulating propositions, Worked examples</td>
</tr>
</tbody>
</table>

Information that goes into a DNA rarity estimate (i.e., where errors can occur)

1. Estimates are derived from testing a small subsection of a population
   - Population allele frequencies (with specific alleles/genotypes)
   - Rarity estimate of DNA profile (e.g., RMP or LR)

2. “All models are wrong – but some are useful” (George Box, 1979)
   - Genetic formulas and assumptions made

3. The risk of error goes up with complexity of the DNA profile (e.g., >2 person mixture or low-quality, low-template DNA sample)
   - Evidentiary DNA Profile (with specific alleles/genotypes)

Recent FBI Erratum on Allele Frequencies Errors Made in 1999

- Genotyping errors were made in 27 samples, affecting the reported frequencies of 51 alleles
- For alleles requiring a frequency correction, the magnitude of the change in frequencies ranged from 0.000012 to 0.018 (average 0.0020 ± 0.0025)
- “The authors are of the view that these discrepancies require acknowledgment but are unlikely to materially affect any assessment of evidential value”

How Book Chapters Map to Data Interpretation Process

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Input Information</th>
<th>Decision to be made</th>
<th>How decision is made</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Data file</td>
<td>Peak or Noise</td>
<td>Analytical threshold</td>
</tr>
<tr>
<td>3</td>
<td>Peak</td>
<td>Allele or Artifact</td>
<td>Stutter threshold; precision sizing bin</td>
</tr>
<tr>
<td>4</td>
<td>Allele</td>
<td>Heterozygote or Homozygote or Allele(s) missing</td>
<td>Peak heights and peak height ratios; stochastic threshold</td>
</tr>
<tr>
<td>5</td>
<td>Genotype/full profile</td>
<td>Single-source or Mixture</td>
<td>Numbers of peaks per locus</td>
</tr>
<tr>
<td>6</td>
<td>Mixture</td>
<td>Deconvolution or not</td>
<td>Major/minor mixture ratio</td>
</tr>
<tr>
<td>7</td>
<td>Low level DNA</td>
<td>Interpret or not</td>
<td>Complexity threshold</td>
</tr>
<tr>
<td>8</td>
<td>Poor quality data</td>
<td>Replace CE components (buffer, polymer, array) or call service engineer</td>
<td>Review size standard data quality with understanding of CE principles</td>
</tr>
</tbody>
</table>

http://www.cstl.nist.gov/strbase/training.htm
A Brief History of DNA Mixtures (1)

- 1991 – Ian Evett article (with single-locus RFLP probes)
- 1995 – Mixtures presented in OJ Simpson trial
- 1996 – 9plex STR kits (Profiler Plus, PowerPlex 1.1)
- 1997 – Weir et al using Likelihood Ratios (LRs) for mixture statistics
- 1998 – Clayton et al (FSS) DNA mixture deconvolution
- 2000 – initial SWGDAM Interpretation Guidelines published
- 2000 – Combined Probability of Inclusion (CPI) statistic is allowed by DNA Advisory Board and pushed by the FBI
- 2000 – 16plex STR kits (PP16 and Identifier)
- 2005 – NIST Interlaboratory Mixture Study (MIX05) finds extensive variation in laboratory approaches

A Brief History of DNA Mixtures (2)

- 2006 – ISFG Mixture Recommendations published emphasizing that LRs are a better method over CPI
- 2007 – Informal SWGDAM study finds most labs doing 2-person mixture data under two hypotheses; in its simplest form LRs become CPI
- 2008 – NUJ study shows value of DNA in burglary cases and more touch DNA samples with complex mixtures begin being processed
- 2010 – SWGDAM Interpretation Guidelines emphasize need for statistics and stochastic thresholds with CPI; probabilistic genotyping approach is mentioned
- 2012 – ISFG publishes LR with drop out potential of allele dropout
- 2013 – Another NIST Interlaboratory Study (MIX13) finds extensive variation in laboratory approaches
- Present – a number of software programs exist to help with calculations but no universal approach exists

Statistical Approaches with Mixtures

1. Random Match Probability (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 – CPI/RMP – Calculation of the probability that a random (unrelated) person would be excluded/included as a contributor to the observed DNA mixture

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

\[ LR = \frac{Pr(E | H_I)}{Pr(E | H_J)} \]
The FBI DNA Advisory Board (DAB) Recommendations on Statistics

February 23, 2000

Forensic Sci. Comm. 2(3); available on-line at

“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)

NIST Interlaboratory Mixture Studies

http://www.cstl.nist.gov/strbase/interlab.htm

- Provide a big-picture view of the community
  - not graded proficiency tests
  - offers laboratories an opportunity to directly compare themselves to others in an anonymous fashion
- Some lessons learned:
  - instrument sensitivities can vary significantly
  - amount of input DNA plays important role in ability to detect minor component(s)
  - protocols and approaches are often different between forensic labs
- Studies Conducted

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th># Labs</th>
<th># Samples</th>
<th>Mixture Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSS 1</td>
<td>1997</td>
<td>22</td>
<td>11 stains</td>
<td>ss, 2p, 3p</td>
</tr>
<tr>
<td>MSS 2</td>
<td>1999</td>
<td>45</td>
<td>11 stains</td>
<td>ss, 2p, 3p</td>
</tr>
<tr>
<td>MSS 3</td>
<td>2000-01</td>
<td>74</td>
<td>7 extracts</td>
<td>ss, 2p, 3p</td>
</tr>
<tr>
<td>MIX05</td>
<td>2005</td>
<td>69</td>
<td>4 cases (fsa)</td>
<td>only 2p</td>
</tr>
<tr>
<td>MIX13</td>
<td>2013</td>
<td>108</td>
<td>5 cases (fsa)</td>
<td>2p, 3p, 4p</td>
</tr>
</tbody>
</table>

MSS: mixed stain study

Elements of DNA Mixture Interpretation

Principles (theory)

Protocols (validation)

Practice (training & experience)

ISFG Recommendations

SWGDAM Guidelines

Your Laboratory

SOPs

Your Laboratory

SOPs

Training within Your Laboratory

Consistency across analysts

Periodic training will aid accuracy and efficiency within your laboratory

Have you read the 2006 ISFG DNA Commission Recommendations on Mixture Interpretation?

A. Yes
B. No
C. Never heard of them!

0% 0% 0%

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

http://www.cstl.nist.gov/strbase/training.htm
Responses to ISFG DNA Commission Mixture Recommendations

- **UK Response**
  - Gill et al. (2008) FSI Genetics 2(1): 76–82
  - Schneider et al. (2006) Rechtsmedizin 16:401-404 (German version)
  - Schneider et al. (2009) Int. J. Legal Med. 123: 1-5 (English version)

- **German Stain Commission**

- **ENFSI Policy Statement**

- **New Zealand/Australia Support Statement**
  - Stringer et al. (2009) FSI Genetics 3(2):144-145

- **SWGDM – Interpretation Guidelines**
  - Approved Jan 2010 and released April 2010 on FBI website

**ISFG Recommendations on Mixture Interpretation**

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

**“Clayton Rules”: Steps in the Interpretation of DNA Mixtures**

Step 1: Identify the Presence of a Mixture
- Examine the number of peaks present in a locus
  - More than 2 peaks at a locus (except for tri-allelic patterns at perhaps one of the loci examined)

Step 2: Designate Allele Peaks

Step 3: Identify the Number of Potential Contributors

Step 4: Estimate the Relative Ratio of the Individuals Contributing to the Mixture

Step 5: Consider All Possible Genotype Combinations

Step 6: Compare Reference Samples

**Is a DNA Profile Consistent with Being a Mixture?**


If the answer to any one of the following three questions is yes, then the DNA profile may very well have resulted from a mixed sample:

- Do any of the loci show more than two peaks in the expected allele size range?
- Is there a severe peak height imbalance between heterozygous alleles at a locus?
- Does the stutter product appear abnormally high (e.g., >15-20%)?
**Peak Height Ratio (PHR) Patterns**

(a) Single-Source vs. (b) Mixture Profiles

**Heterozygote Balance across 10 STRs Under Different Conditions**

N = 422 heterozygous loci

N = 1688 heterozygous loci

**Stutter Ratios** across 10 STRs Under Different Conditions

Impact of Template DNA Amount on Variation in Peak Height Ratio

**Step #2: Designate Allele Peaks**

- Use regular data interpretation rules to decipher between true alleles and artifacts
- Use stutter filters to eliminate stutter products from consideration (although stutter may hide some of minor component alleles at some loci)
- Consider heterozygote peak heights that are highly imbalanced (<60%) as possibly coming from two different contributors

**Data Interpretation Steps**

The Steps of Data Interpretation

Analytical Threshold

Expected Stutter %

Stochastic Threshold

Peak Height Ratio (PHR)

Next step: Examine feasible genotypes to deduce possible contributor profiles

Moving from individual locus genotypes to profiles of potential contributors to the mixture is dependent on mixture ratios and numbers of contributors
Step #3: Identifying the Potential Number of Contributors

- **Important for statistical calculations**
- Typically if 2, 3, or 4 alleles then 2 contributors
- If 5 or 6 alleles per locus then 3 contributors
- If >6 alleles in a single locus, then >4 contributors
- Also pay attention to relative peak heights and potential genotype combinations

### DNA Mixture Example for this Workshop

#### Potential Problems with Amelogenin

- Works best with 2-person male/female mixtures (such as sexual assault cases)
  - Male/male mixture or multiple males with single female component limit usefulness
- Molecular reasons for alteration of expected ratio
  - Deletion of AMEL Y (or primer site mutation)
  - Deletion of AMEL X (or primer site mutation)

![Image of DNA mixture example](image)

#### Comparison of Expected and Simulated Mixture Results

**Expected Results when estimating # of contributors:**
- If 2, 3, or 4 alleles are observed at every locus across a profile then 2 contributors are likely present
- If a maximum of 5 or 6 alleles at any locus, then 3 contributors are possible
- If >6 alleles in a single locus, then >3 contributors

**Results from Simulation Studies:**
- Buckleton et al. (2007) found with a simulation of four person mixtures that 0.02% would show four or fewer alleles and that 76.35% would show six or fewer alleles for the CODIS 13 STR loci.
- Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

![Image of expected and simulated mixture results](image)

### Levels of Locus Heterozygosity Impact the Number of Alleles Observed in Mixtures

<table>
<thead>
<tr>
<th>Loci</th>
<th>No. of alleles</th>
<th>Simulated 2-Person Mixture</th>
<th>Results from a 2-Person Mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3</td>
<td>0.011</td>
<td>0.249</td>
<td>0.559</td>
</tr>
<tr>
<td>vWA</td>
<td>0.018</td>
<td>0.144</td>
<td>0.548</td>
</tr>
<tr>
<td>D16</td>
<td>0.016</td>
<td>0.287</td>
<td>0.512</td>
</tr>
<tr>
<td>D2</td>
<td>0.016</td>
<td>0.664</td>
<td>0.462</td>
</tr>
</tbody>
</table>

![Image of levels of locus heterozygosity impact](image)

---

http://www.cstl.nist.gov/strbase/training.htm
Impact of Allele Sharing on Ability to Clearly Determine the Number of Contributors

Real data shows variation due to stochastic (random) effects

<table>
<thead>
<tr>
<th>Allele (repeat #)</th>
<th>Peak height (RFU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

Impact of Additional STR Loci on Mixture Assumptions

Probability of incorrectly assigning the specific number of contributors based on observed alleles (not considering peak height imbalances)

<table>
<thead>
<tr>
<th>True # of contributors</th>
<th>Using NIST Caucasians (Hill et al. 2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>CODIS13 1.75E-40 6.34E-09 0.161242 0.945657 0.999873</td>
</tr>
<tr>
<td></td>
<td>CODIS22 0 (&lt; E-99) 9.59E-21 5.32E-05 0.188138 0.859901</td>
</tr>
<tr>
<td>5</td>
<td>CODIS13 9.78E-33 2.10E-06 0.41432 0.989651</td>
</tr>
<tr>
<td></td>
<td>CODIS22 6.36E-61 7.01E-15 0.004837 0.610149</td>
</tr>
<tr>
<td>4</td>
<td>CODIS13 7.02E-25 0.000515 0.785495</td>
</tr>
<tr>
<td></td>
<td>CODIS22 3.50E-46 3.49E-09 0.16523</td>
</tr>
<tr>
<td>3</td>
<td>CODIS13 8.42E-17 0.059486</td>
</tr>
<tr>
<td></td>
<td>CODIS22 5.77E-31 0.000433</td>
</tr>
<tr>
<td>2</td>
<td>CODIS13 8.70E-08 0.999873</td>
</tr>
<tr>
<td></td>
<td>CODIS22 2.05E-15 0.999873</td>
</tr>
</tbody>
</table>

Step #4: Estimation of Relative Ratios for Major and Minor Components to a Mixture

- Mixture studies with known samples have shown that the mixture ratio between loci is fairly well preserved during PCR amplification
- Thus it is generally thought that the peak heights (areas) of alleles present in an electropherogram can be related back to the initial component concentrations
- For 2-person mixtures, start with loci possessing 4 alleles...

Mixture Ratio and Mixture Proportion Determined Using D18S51 Peak Heights

If we assume:
- Major: 16,18
- Minor: 14,20

Mixture Ratio
\[ M_R = \frac{\phi_{16} + \phi_{18}}{\phi_{14} + \phi_{20}} = \frac{288 + 274}{53 + 65} = 4.76 \]

Mixture Proportion
\[ M_X = \frac{\phi_{16} + \phi_{18}}{\phi_{14} + \phi_{18} + \phi_{18} + \phi_{20}} = \frac{288 + 274}{53 + 266 + 274 + 65} = 0.826 = 83% \]

Step #5: Consider All Possible Genotype Combinations

<table>
<thead>
<tr>
<th>DNA Template Amount</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625 ng</td>
<td>tt, tt, tt, tt, tt, tt</td>
</tr>
<tr>
<td>0.125 ng</td>
<td>tt, tt, tt, tt, tt, tt, tt, tt</td>
</tr>
<tr>
<td>0.25 ng</td>
<td>tt, tt, tt, tt, tt, tt, tt, tt, tt</td>
</tr>
<tr>
<td>0.5 ng</td>
<td>tt, tt, tt, tt, tt, tt, tt, tt, tt, tt</td>
</tr>
</tbody>
</table>

Detectability of Alleles and Peak Height Ratios Can Vary with DNA Template Amount

http://www.cstl.nist.gov/strbase/training.htm
**Steps in DNA Interpretation**

- **Question sample**
- **Mixture**
- **Peak (vs. noise)**
- **Allele (vs. artifact)**
- **Genotype (allele pairing)**
- **Profile (genotype combining)**
- **Weight of Evidence**
- **Match probability**

**Known sample**

- **Reference Sample(s)**
- **Report Written & Reviewed**

**It's the potential Genotypes NOT the Alleles that matter in mixtures!**

---

**Considering Genotype Combinations**

- **PR**
- **QS**
- **PQ**
- **RS**
- **QR**
- **PS**

Peak Height Ratios (PHR)
- Minimum Peak Height (mPH)
- Proportion (p) or mixture proportion (M)

---

**Restricted vs Unrestricted Genotype Combinations**

**Example data**

- **P**
- **Q**
- **R**
- **S**

**Suspect genotype**

- **13**
- **14**
- **15**
- **16**

**Unrestricted genotype combinations**

- **PQ & RS**
- **RS & PQ**
- **PR & SQ**
- **SQ & PR**
- **PS & RQ**
- **RQ & PS**

**Restricted genotype combinations**

- **PQ & RS**
- **RS & PQ**

---

**Possible genotype combinations in 2-person mixtures**

- **A**
- **B**

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

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

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

- **1 allele**
  - Homozygote + homozygote, overlapping allele

---

**Possible Genotype Combinations with Two-Person Mixtures**

- **1 allele (P)**
- **2 alleles (P, Q)**
- **3 alleles (P, Q, R)**
- **4 alleles (P, Q, R, S)**

- **Homozygote + homozygote, overlapping allele**
- **Homozygote + homozygote, no overlapping alleles**
- **Heterozygote + heterozygote, overlapping allele**
- **Heterozygote + heterozygote, no overlapping alleles**
- **Heterozygote + homozygote, overlapping allele**
- **Heterozygote + homozygote, no overlapping alleles**
- **Homzygote + homzygote, overlapping allele**
- **Homzygote + homzygote, no overlapping alleles**

---

**Potential Genotype Combinations with Three Contributors**

- **23 “families” of possibilities**
- **3 allele pattern has 8 “families”**

**150 total combinations**

This “family” has 30 possibilities

---

[http://www.cstl.nist.gov/strbase/training.htm](http://www.cstl.nist.gov/strbase/training.htm)
Step #6: Compare Reference Samples

- If there is a suspect, a laboratory must ultimately decide to include or exclude him…

- If no suspect is available for comparison, does your laboratory still work the case? (Isn’t this a primary purpose of the national DNA database?)

- Victim samples can be helpful to eliminate their allele contributions to intimate evidentiary samples and thus help deduce the perpetrator

### DNA Mixture Example for this Workshop

#### D18S51 Results

- **Major:** 16,18
- **Minor:** 14,20

- Individual “C”
- Individual “D”

#### Single-Source Sample Profile (1 ng of “C”)

#### Single-Source Sample Profile (1 ng of “D”)

### 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
**ISFG 2015: Basic STR Interpretation Workshop**  
(J.M. Butler & S.N. Gittelson)  
31 August 2015

*Step #1:* Identify the Presence of a Mixture  
*Step #2:* Designate Allele Peaks  
*Step #3:* Identify the Number of Potential Contributors  
*Step #4:* Estimate the Relative Ratio of the Individuals Contributing to the Mixture  
*Step #5:* Consider All Possible Genotype Combinations  
*Step #6:* Compare Reference Samples  
*Step #7:* Determine statistical weight-of-evidence

**German Mixture Classification Scheme**  
(German 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

---

**Mixed DNA Sample**  
Determine STR profile and compute RMP  
Probability of Exclusion (PDE)  
Likelihood Ratio (LR)

**Type A**  
Type C  
A biostatistical analysis must be performed  
*Assume number of contributor(s)*

**Type B**  
Probability of Exclusion (PDE)  
Likelihood Ratio (LR)

**Type C**  
A biostatistical analysis must be performed  
*Assume number of contributor(s)*

---

**Allele stacking with overlapping genotypes**  
(a) observed data  
(b) contributor genotypes

---

**Potential Impact of Allele Sharing**  
**Allele Q may not represent the "major" contributor**

---

http://www.cstl.nist.gov/strbase/training.htm
Peaks in the Stutter Position of Major Alleles May Need to Be Considered as Pairing with Alleles from Minor Contributors

If this result is from 2 contributors, then the likely genotype combinations are 14,16 (PQ) and 18,19 (RS)

If >2 contributors, then genotypes 13,18 or 15,19 (and many other combinations) may be possible contributors...

More Sensitive Assays and Instruments

- **Superb sensitivity is available** with DNA amplification using the polymerase chain reaction and laser-induced fluorescence detection with capillary electrophoresis
- Since 2007 (beginning with the release of the MiniFiler STR kit), **improved buffers and enzymes** have been used to boost DNA sensitivities in all STR kits
  - In 2010 the ABI 3500 Genetic Analyzer was released with 4X signal over the previous ABI 3100 and ABI 310 instruments
  - Energy-transfer dyes are used with some of the STR kits
  - Some labs increase the sensitivity dial with additional PCR cycles
- So what is wrong with having improved sensitivity?

Improved Sensitivity is a Two-Edged Sword


"As sensitivity of DNA typing improves, laboratories' abilities to examine smaller samples increases. This improved sensitivity is a two-edged sword. With greater capabilities comes greater responsibilities to report meaningful results. Given the possibility of DNA contamination and secondary or even tertiary transfer in some instances, does the presence of a single cell (or even a few cells) in an evidentiary sample truly have meaning?..."

Ian Evett and Colleagues’ Case Assessment and Interpretation: Hierarchies of Propositions

**TABLE 16.2** Hierarchical Levels of Propositions Originally Developed by the UK Forensics 1998a, 1998b, Evett et al. 2000a, 2000b, Gill 2001)

<table>
<thead>
<tr>
<th>Hierarchy Levels</th>
<th>Propositions</th>
<th>Decision Maker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level III</td>
<td>Offense</td>
<td>Responsibility of the jury or judge</td>
</tr>
<tr>
<td></td>
<td>Supplies the probability that a suspect has committed a criminal offense</td>
<td></td>
</tr>
<tr>
<td>Level II</td>
<td>Activity</td>
<td>Juror or possibly scientist if given adequate case circumstances</td>
</tr>
<tr>
<td></td>
<td>Informing regards of the kinds of activities which may have produced the forensic evidence</td>
<td></td>
</tr>
<tr>
<td>Level I Source</td>
<td>Source</td>
<td>Scientist</td>
</tr>
<tr>
<td>Sub-level I Sub-source</td>
<td>With low amounts of DNA, the scientist may not be able to infer how the DNA arrived at the site where the DNA sample was collected</td>
<td></td>
</tr>
</tbody>
</table>

More Touch Evidence Samples

- **More poor-quality samples are being submitted**
  - Samples with <100 pg of DNA submitted in Belgium: 19% (2004) → 45% (2008)
    - Michel 2009 FSIGSS 2:542-543
  - AAFS 2014 presentations showed poor success rates
    - NYC (A110): only 10% of >9,500 touch evidence swabs from 2007 to 2011 produced usable DNA results
    - Allegheny County (A114): examined touch DNA items processed from 2008 to 2013 across different evidence types (e.g., 8 of 56 car door handles yielded “resolvable profiles”)

http://www.cstl.nist.gov/strbase/training.htm
New Options Exist for Statistical Analysis

- Increase in approaches to try and cope with potential allele dropout → number of probabilistic genotyping methods have grown since Balding & Buckleton 2009 article
- Many possible choices for probabilistic genotyping software with commercial interests at stake


Math Analogy to DNA Evidence

<table>
<thead>
<tr>
<th>Basic Arithmetic</th>
<th>Algebra</th>
<th>Calculus</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 + 2 = 4</td>
<td>$2x^2 + x = 10$</td>
<td>$\int_{x=0}^{10} f'(x)dx$</td>
</tr>
</tbody>
</table>

Single-Source DNA Profile (DNA databasing)

Sexual Assault Evidence (2-person mixture with high-levels of DNA)

Touch Evidence (>2-person, low-level, complex mixtures perhaps involving relatives)

Many laboratories are not prepared to cope with complex mixtures

- Have appropriate validation studies been performed to inform proper interpretation protocols? (curriculum & classroom instruction)
- Are appropriately challenging proficiency tests being given? (graded homework assignments)
- Would we want to go into a calculus exam only having studied algebra and having completed homework assignments involving basic arithmetic?

Are We Facing a “Perfect Storm” for DNA Testing and Interpretation?

- Increase in assay and instrument sensitivity
- Increase in challenging casework samples (touch evidence)
- Increase in possible statistical techniques used with complex mixtures
- Increase in number of loci examined with new STR kits

Perhaps We Should Slow Down with Some of the DNA Mixtures That We (Scientists and Lawyers) Are Taking On...

<table>
<thead>
<tr>
<th>Poor Quality Conditions</th>
<th>Large Numbers of Contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet surface leads to hydroplaning</td>
<td>Slick, mountain road</td>
</tr>
<tr>
<td>Foggy, wet conditions</td>
<td>Curve, poor visibility</td>
</tr>
</tbody>
</table>
### Decisions during Data Interpretation

<table>
<thead>
<tr>
<th>Input Information</th>
<th>Decision to be made</th>
<th>How decision is made</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data file</td>
<td>Peak or Noise</td>
<td>Analytical threshold</td>
</tr>
<tr>
<td>Peak</td>
<td>Allele or Artifact</td>
<td>Stutter threshold; precision sizing bin</td>
</tr>
<tr>
<td>Allele</td>
<td>Heterozygote or Homozygote or Allele(s) missing</td>
<td>Peak heights and peak height ratios; stochastic threshold</td>
</tr>
<tr>
<td>Genotype/Profile</td>
<td>Single-source or Mixture</td>
<td>Numbers of peaks per locus</td>
</tr>
<tr>
<td>Mixture</td>
<td>Deconvolution or not</td>
<td>Major/minor mixture ratio</td>
</tr>
<tr>
<td>低 level DNA</td>
<td>Interpret or not</td>
<td>Complexity/uncertainty threshold</td>
</tr>
<tr>
<td>Poor quality data</td>
<td>Replace CE components (buffer, polymer, array) or call service engineer</td>
<td>Review size standard data quality with understanding of CE principles</td>
</tr>
</tbody>
</table>

(M. Butler (2015), Advanced Topics in Forensic DNA Typing: Interpretation, Table 1.1, p. 6)

### Results Depend on Assumptions

- “Although courts expect one simple answer, statisticians know that the result depends on how questions are framed and on assumptions tucked into the analysis.”

- We inform our assumptions with data from validation studies...

### Ian Evett on Interpretation

“The crucial element that the scientist brings to any case is the interpretation of those observations. This is the heart of forensic science: it is where the scientist adds value to the process.”


### Acknowledgments

$ NIST Special Programs Office
Simone Gittelson

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

Contact info:
john.butler@nist.gov
+1-301-975-4049

Final version of this presentation will be available at:
http://www.cstl.nist.gov/strbase/NISTpub.htm
Statistical Interpretation 2:
Approaches to calculating mixture statistics
Likelihood ratios and formulating propositions
Worked examples
Simone N. Gittelson, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Presentation Outline

1. Combined Probability of Inclusion (CPI)
2. modified Random Match Probability (mRMP)
3. Likelihood Ratio (LR)
4. Formulating Propositions for Likelihood Ratios

Acknowledgement and Disclaimers

I thank John Butler for the discussions and advice on preparing this presentation. I also acknowledge John Buckleton for all his helpful explanations on DNA mixture interpretation.

Points of view in this presentation are mine and do not necessarily represent the official position or policies of 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.

Combined Probability of Inclusion (CPI)

the probability that a random person would be included as a contributor to the mixture

the probability that a random person would not be excluded as a contributor to the mixture

“random man not excluded”

http://www.cstl.nist.gov/strbase/training.htm
Information it takes into account

- presence of alleles

Assumptions

The population is in Hardy-Weinberg and Linkage equilibrium.

All genotypes considered are equally likely.

We do not take into account the number of contributors in the calculation.

Combined Probability of Inclusion (CPI)

The probability that a random person would be included as a contributor to this mixture:

\[
P_I = p_{10}^2 + 2p_{10}p_{11} + 2p_{10}p_{12} + p_{11}^2 + 2p_{11}p_{12} + p_{12}^2
\]

Where:

- \( p_{10} = 0.220 \)
- \( p_{11} = 0.309 \)
- \( p_{12} = 0.360 \)

\[
P_I = (0.220 + 0.309 + 0.360)^2
\]

\[
P_I = 0.79
\]

Combined Probability of Inclusion (CPI)

The probability that a random person would be included as a contributor to this mixture:

\[
P_I = p_{14}^2 + 2p_{14}p_{16} + 2p_{14}p_{18} + 2p_{14}p_{20} + p_{16}^2 + 2p_{16}p_{18} + 2p_{16}p_{20} + p_{18}^2 + 2p_{18}p_{20} + p_{20}^2
\]

Where:

- \( p_{14} = 0.134 \)
- \( p_{16} = 0.147 \)
- \( p_{18} = 0.078 \)
- \( p_{20} = 0.018 \)

\[
P_I = (0.134 + 0.147 + 0.078 + 0.018)^2
\]

\[
P_I = 0.14
\]
**Combined Probability of Inclusion (CPI)**

\[ CPI = PI_{CSF1PO} \times PI_{D18S51} \]

\[ = 0.79 \times 0.14 \]

\[ = 0.11 \]

**D18S51**

\[ P_{14} = 0.134 \]
\[ P_{16} = 0.147 \]
\[ P_{18} = 0.078 \]
\[ P_{20} = 0.018 \]

\[ PI = p_{14}^2 + 2p_{14}p_{16} + 2p_{14}p_{18} + 2p_{14}p_{20} + p_{16}^2 + 2p_{16}p_{18} + 2p_{16}p_{20} + p_{18}^2 + 2p_{18}p_{20} + p_{20}^2 \]

\[ = (0.134 + 0.147 + 0.078 + 0.018)^2 \]

\[ = 0.14 \]

The probability that a random person would be included as a contributor to this mixture

**HOWEVER, CPI can only be applied if...**

...there is no possibility of allele drop-out.

If there is a possibility of allele drop-out, then everyone would be included as a possible contributor. In this case, the probability that a random person would be included is equal to 1.
**Combined Probability of Inclusion (CPI)**

\[
CPI = PI_{DBS1179} \times PI_{D21S11} \times PI_{D7SB20} \times PI_{CSF1PO} \\
\times PI_{D35135B} \times PI_{TH01} \times PI_{D13S317} \times PI_{D16S539} \\
\times PI_{D25I33B} \times PI_{D19S433} \times PI_{vWA} \times PI_{TPOX} \\
\times PI_{D18S51} \times PI_{D58B18} \times PI_{FGA} \\
= 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \\
= 1
\]

Everyone is included. No one is excluded.

**modified Random Match Probability (mRMP)**


**Information it takes into account**

- presence of genotypes
- list of genotype combinations that are possible
- presence of alleles and peak heights
- peaks below stochastic threshold (where allele drop-out is possible)

**Assumptions**

- The population is in Hardy-Weinberg and Linkage equilibrium (because we are doing the calculation with \( \theta = 0 \)).
- The number of contributors is ____.
- Only the genotypes satisfying the mixture deconvolution rules are possible.
EPG of the crime stain:

Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

stochastic threshold = 150 rfu

CSF1PO

1. major contributor: 10,11
2. mRMP = \(2p_{10}p_{11}\)
   \(= 2(0.220)(0.309)\)
   \(= 0.136\)

D18S51

1. major contributor: 16,18
2. mRMP = \(2p_{16}p_{18}\)
   \(= 2(0.147)(0.078)\)
   \(= 0.023\)

TPOX

peak at 11 is above the stochastic threshold

1. major contributor: 28, F
2. mRMP = \(2p_{28}(1 - p_{28}) + p_{28}^2\)
   \(= 2(0.159)(0.841) + 0.159^2\)
   \(= 0.293\)
### Genotype of major contributor?

A. 8,11  
B. 8,8  
C. 11,11  
D. 8,F  
E. 15,15  
F. ???

### Genotype of minor contributor?

A. 11,11  
B. 11,11 or 8,11  
C. 11,F  
D. 8,F  
E. 15,15  
F. ???

### mRMP for major contributor?

A. \((0.252)^2 = 0.064\)  
B. \(2(0.525)(0.252) = 0.265\)  
C. \((0.525)^2 = 0.276\)  
D. 0.525  
E. 1  
F. ???  
G. I forgot to bring my math skills to this workshop

### mRMP for minor contributor?

A. \((0.252)^2 = 0.064\)  
B. \((0.252)^2 + 2(0.525)(0.252) = 0.328\)  
C. \(2(0.252) - (0.252)^2 = 0.440\)  
D. \(2(0.525) - (0.525)^2 = 0.774\)  
E. 1  
F. ???  
G. All answers look the same to me (I drank too much vodka last night)

### mRMP for D8S1179

<table>
<thead>
<tr>
<th>X Chromosome</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{13})</td>
<td>0.330</td>
<td>(p_{14})</td>
<td>0.166</td>
</tr>
</tbody>
</table>

### Genotype of major contributor?

A. 13,13  
B. 13,14  
C. 13,15  
D. 14,15  
E. 16,17  
F. ???

---

http://www.cstl.nist.gov/strbase/training.htm
**mRMP for major contributor?**

A. $(0.166)^2 = 0.028$
B. $2(0.166)(0.104) = 0.035$
C. $2(0.330)(0.104) = 0.069$
D. $2 - 0.166 - 0.104 = 1.73$
E. $2 + 0.166 + 0.104 = 2.27$
F. ???
G. All answers look the same to me
   (I drank vodka for breakfast)

**mRMP for minor contributor?**

A. $2(0.330)(0.166 + 0.104) = 0.178$
B. $2(0.166) - (0.166)^2 = 0.304$
C. $0.330$
D. $2(0.330) - (0.330)^2 = 0.551$
E. $1$
F. ???
G. I used up all my math skills on the previous questions

**Genotype of minor contributor**

A. 13,14 or 13,15
B. 14,F
C. 13,F
D. 15,F
E. 16,17
F. ???

**mRMP for the Major Contributor**

**mRMP for the Minor Contributor**

---

http://www.cstl.nist.gov/strbase/training.htm
mRMP for the Minor Contributor

<table>
<thead>
<tr>
<th>Locus</th>
<th>mRMP (minor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>0.551</td>
</tr>
<tr>
<td>D21511</td>
<td>0.293</td>
</tr>
<tr>
<td>D7S820</td>
<td>0.267</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>0.510</td>
</tr>
<tr>
<td>D8S1358</td>
<td>0.419</td>
</tr>
<tr>
<td>TH01</td>
<td>0.571</td>
</tr>
<tr>
<td>D13S317</td>
<td>0.219</td>
</tr>
<tr>
<td>D165539</td>
<td>0.529</td>
</tr>
<tr>
<td>D21338</td>
<td>0.199</td>
</tr>
<tr>
<td>D195433</td>
<td>0.445</td>
</tr>
<tr>
<td>VWA</td>
<td>1</td>
</tr>
<tr>
<td>TP10</td>
<td>0.328</td>
</tr>
<tr>
<td>D18551</td>
<td>0.005</td>
</tr>
<tr>
<td>D5S818</td>
<td>0.266</td>
</tr>
<tr>
<td>YCA</td>
<td>1</td>
</tr>
</tbody>
</table>

All Loci: $2.3 \times 10^{-8}$

Likelihood Ratio (LR)


There are two sides to every story...

$H_p$: The crime stain came from the person of interest (POI).

$H_d$: The crime stain did not come from the POI. It came from some other person.

prosecution’s proposition
defense’s proposition

The likelihood ratio gives the value of the findings with regard to the prosecution’s and defense’s standpoints in the case.

Information it takes into account

- presence of genotypes
- peaks below stochastic threshold (where allele drop-out is possible)
and
- the standpoints of the prosecution and the defense (i.e., the two competing propositions)

Assumptions

DEPEND ON THE MODEL USED

Here, we will use the classical binary model and make the same assumptions as we did for mRMP:
- The population is in Hardy-Weinberg and Linkage equilibrium.
- The number of contributors is ___.
- Only the genotypes satisfying the mixture deconvolution rules are possible.
Standpoints of the prosecution and the defense

$H_p$: The DNA came from the POI and an unknown contributor.

$H_d$: The DNA came from two unknown contributors.

Likelihood Ratio (LR)

The probability of observing the DNA typing results of the crime stain given the POI's genotype and that the DNA came from the POI and one unknown contributor:

$$LR = \frac{Pr(G_{CS} | G_{POI}, H_p)}{Pr(G_{CS} | G_{POI}, H_d)}$$

Numerator:

$$Pr(G_{CS} | G_{POI}, H_p) = Pr(16,18) \times Pr(14,20) = 2p_{16}p_{18} \times 1 = 2p_{16}p_{18}$$

Denominator:

$$Pr(G_{CS} | G_{POI}, H_d) = Pr(16,18) \times Pr(14,20) = 2p_{16}p_{18} \times 2p_{14}p_{20} = 4p_{16}p_{18}p_{14}p_{20}$$

What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype [14,20]?

What is the probability of obtaining these DNA typing results for the crime stain if the POI is not a contributor?
**Likelihood Ratio (LR)**

\[
\text{D18S51} \\
\begin{array}{l}
P_{14} = 0.134 \\
P_{16} = 0.147 \\
P_{18} = 0.078 \\
P_{20} = 0.018 \\
\end{array}
\]

\[G_{POI} = \{14, 20\}\]

\[
LR = \frac{2P_{14}P_{18}}{4P_{14}P_{16}P_{18}P_{20}} \\
= \frac{1}{2P_{14}P_{20}} \\
= 207.30
\]

The DNA typing results are 207 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

---

**Likelihood Ratio (LR)**

\[
\text{CSF1PO} \\
\begin{array}{l}
P_{10} = 0.220 \\
P_{11} = 0.309 \\
P_{12} = 0.360 \\
\end{array}
\]

\[G_{POI} = \{12, 12\}\]

**Numerator:**

\[
\text{peak at 12 is above the stochastic threshold}
\]

What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype \(\{12, 12\}\)?

\[
Pr(10, 11) \times Pr(12, 12) \\
= 2P_{10}P_{11} \times 1 \\
= 2P_{10}P_{11}
\]

**Denominator:**

\[
\text{peak at 12 is above the stochastic threshold}
\]

What is the probability of obtaining these DNA typing results for the crime stain if the POI is not a contributor?

\[
Pr(10, 11) \times Pr(10, 12) + Pr(10, 11) \times Pr(10, 12) + Pr(10, 11) \times Pr(11, 12) \\
= 2P_{10}P_{11} \times P_{12}^2 + 2P_{10}P_{11} \times 2P_{10}P_{12} + 2P_{10}P_{11} \times 2P_{11}P_{12} \\
= 2P_{10}P_{11}P_{12}(P_{12} + 2P_{10} + 2P_{11})
\]

---

http://www.cstl.nist.gov/strbase/training.htm
Likelihood Ratio (LR) for CSF1PO:

\[ LR = \frac{2p_{10}p_{11}}{2p_{10}p_{11}(p_{12} + 2p_{10} + 2p_{11})} \]

\[ = \frac{1}{p_{12} + 2p_{10} + 2p_{11}} \]

\[ = 1.96 \]

Likelihood Ratio (LR) for D21S11:

\[ LR = \frac{2p_{30}p_{32}}{2p_{30}p_{32}(2p_{28} - p_{28}^2)} \]

\[ = 1 \]

The DNA typing results are about 2 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.
What does a $LR \approx 3$ mean?

A. The person of interest committed the crime.
B. A total of 3 peaks were observed at this locus.
C. It is about 3 times more probable that the DNA came from the person of interest and an unknown contributor than that the DNA came from two unknown contributors.
D. There are 3 contributors to this DNA mixture.
E. The DNA typing results are about 3 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.
F. ???

### Likelihood Ratio (LR)

**D21S11**

$p_{28} = 0.159$

$p_{32} = 0.090$

$G_{POI} = (28,28)$

Peak at 28 is below the stochastic threshold

### Transposed Conditional

$H$: the animal is an elephant

$E$: the animal has four legs

Pr($E|H$) = 1

Pr($H|E$) = $\frac{1}{5,000}$

Pr($E|H$) is not equal to Pr($H|E$)

### Boston University Mixture

Transposed Conditional

$\text{Pr}(H|E) = \frac{\text{prior odds}}{100} \times \frac{\text{likelihood ratio}}{100}$

These DNA typing results indicate that the probability of the prosecution’s proposition being true is 100 times greater than the probability of the defense’s proposition being true.

These DNA typing results indicate that the probability of the defense’s proposition being true is 1,000 times greater than the probability of the prosecution’s proposition being true.

EPG of the crime stain:

Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

Posterior odds: 100,000

Prior odds: 1

Likelihood ratio: 100

Pr($H|E$) = $\frac{1}{5,000}$

Pr($E|H$) is not equal to Pr($H|E$)
Likelihood Ratio (LR) for all loci

\[ H_p: \text{The DNA came from the POI and an unknown contributor.} \]
\[ H_d: \text{The DNA came from two unknown contributors.} \]

If \( H_p \) is true, is the POI the major contributor or the minor contributor?

If \( H_p \) is true, the POI could be either the major contributor or the minor contributor. Let us consider these possibilities to be equally probable. So if \( H_p \) is true, there is a probability of \( \frac{1}{2} \) that the POI is the major contributor and a probability of \( \frac{1}{2} \) that the POI is the minor contributor.
We can only observe these DNA typing results if the POI is the minor contributor.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Major</th>
<th>Minor</th>
<th>(G_{\text{POI}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>D18S51</td>
<td>16,18</td>
<td>14,20</td>
<td>(G_{\text{POI}} = {14,20})</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>10,11</td>
<td>12,12</td>
<td>(G_{\text{POI}} = {12,12})</td>
</tr>
<tr>
<td>D2IS51</td>
<td>30,32</td>
<td>28,28</td>
<td>(G_{\text{POI}} = {28,28})</td>
</tr>
<tr>
<td>TPOX</td>
<td>8,8</td>
<td>11,11</td>
<td>(G_{\text{POI}} = {11,11})</td>
</tr>
</tbody>
</table>

Likelihood Ratio (LR) for all loci

- **H\(_p\)**: The DNA came from the POI and an unknown contributor.
- **H\(_d\)**: The DNA came from two unknown contributors.

Numerator:

Because these DNA typing results are only possible when the POI is the minor contributor, and the POI is the minor contributor with a probability of \(\frac{1}{2}\), we multiply the numerator of the likelihood ratio for the entire profile by \(\frac{1}{2}\).

Likelihood Ratio (LR)

\[ LR = 2.5 \times 10^7 = 25 \text{ million} \]

The DNA typing results are about 25 million times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

Formulating Propositions for Likelihood Ratios


http://www.cstl.nist.gov/strbase/training.htm
Formulating Propositions for Likelihood Ratios

If the propositions change, the likelihood ratio changes.

The propositions depend on the case circumstances and the standpoints of the prosecution and the defense.

Consider the following 4 cases for a 2-person mixture.

2-person mixture

Case 1: Alleged Rape Case
The crime sample is a vaginal swab taken from the complainant V. Standpoints of the prosecution and the defense:
prosecution: “POI raped V.”
defense: “POI did not rape V. Someone else raped V.”
case circumstances: V had no consensual partner at the time of this event.

What is $H_p$? What is $H_d$?

2-person mixture

Case 2: Stabbing Case
A person V is found stabbed to death. The crime sample is taken from the POI’s shirt sleeve shortly after the discovery of V. Standpoints of the prosecution and the defense:
prosecution: “POI stabbed V.”
defense: “POI did not stab V. POI has never seen V before. Someone else stabbed V.”

case circumstances: V had no consensual partner at the time of this event.

What is $H_p$? What is $H_d$?

2-person mixture

Case 3: Assault Case
A person V is found unconscious in an alleyway. There are indications that V was hit on the head with a hard object. The crime sample is taken from a metal bar found on the ground nearby.

Standpoints of the prosecution and the defense:
prosecution: “POI hit V with the metal bar.”
defense: “POI did not hit V. Someone else hit V.”
case circumstances: The metal bar is associated with neither V nor POI.

What is $H_p$? What is $H_d$?
2-person mixture

Case 3: Assault Case

\( H_p \): The DNA came from the victim V and the POI.

\( H_d \): The DNA came from two unknown contributors.

Case 4: Shooting

The crime sample is taken from the trigger of a handgun found on the crime scene.

Standpoints of the prosecution and the defense:
- prosecution: “POI shot this gun.”
- defense: “Someone else shot this gun. POI never touched this gun.”

What is \( H_p \)? What is \( H_d \)?

2-person mixture

Case 4: Assault Case

\( H_p \): The DNA came from POI and an unknown contributor.

\( H_d \): The DNA came from two unknown contributors.

Summary

<table>
<thead>
<tr>
<th>Takes into account</th>
<th>Models:</th>
</tr>
</thead>
<tbody>
<tr>
<td>presence/absence of alleles</td>
<td>allele drop-out and allele drop-in</td>
</tr>
<tr>
<td>CPI</td>
<td>X</td>
</tr>
<tr>
<td>mRMP</td>
<td>X</td>
</tr>
<tr>
<td>LR (binary)</td>
<td>X</td>
</tr>
<tr>
<td>LR (semi-continuous)</td>
<td>X</td>
</tr>
<tr>
<td>LR (fully continuous)</td>
<td>X</td>
</tr>
</tbody>
</table>

Models:
- CPI
- mRMP
- LR (binary)
- LR (semi-continuous)
- LR (fully continuous)

Acknowledgements

John Butler

Discussions on DNA mixture interpretation

John Buckleton

Contact Info.:
simone.gittelson@nist.gov
+1-301-975-4892

Final version of this presentation will be available at: http://www.cstl.nist.gov/strbase/training.htm
Additional Training Resources

Boston University DNA Mixture Training: http://www.bu.edu/dnamixtures/


NIST STRBase Mixture Information: http://www.cstl.nist.gov/strbase/mixture.htm

Simone Gittelson workshop on Mixture Interpretation & Statistics at the Bode East 14th Annual DNA Technical Conference (Orlando, FL), May 29, 2015

Guidance for DNA Interpretation


DNA Commission of the ISFG: http://www.isfg.org/Publications/DNA+Commission


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


**Software for DNA Analysis and Interpretation**

**Armed Xpert** (NicheVision): http://www.armedxpert.com/


**DNAMIX** (Bruce Weir): http://www.biostat.washington.edu/~bsweir/DNAMIX3/webpage/

**DNA Mixture Separator** (Torben Tvedebrink): http://people.math.aau.dk/~tvede/mixsep/


**Forensic DNA Statistics** (Peter Gill): https://sites.google.com/site/forensicdnastatistics/

**Forensim** (Hinda Haned): https://forensim.r-forge.r-project.org/


**GenoProof Mixture** (Qualitytype): http://www.qualitytype.de/en/qualitytype/genoproof-mixture

**ISFG Software Resources Page:** http://www.isfg.org/software

**Lab Retriever** (Scientific Collaboration, Innovation & Education Group): http://www.scieg.org/lab_retriever.html

**likeLTD** (David Balding): https://sites.google.com/site/baldingstatisticalgenetics/software/likeld-r-forensic-dna-r-code

**LRmix** (Hinda Haned): https://sites.google.com/site/forensicdnastatistics/PCR-simulation/lrmix
LRmix Studio (Hinda Haned): http://lrmixstudio.org/


STRmix (Ducan Taylor, Jo-Anne Bright, John Buckleton): http://strmix.com/

TrueAllele Casework (Cybergenetics): http://www.cybgen.com/systems/casework.shtml

**STR Kits, Loci, and Population Data**


STRBase: http://www.cstl.nist.gov/strbase

**Setting Thresholds**


Statistical Approaches


Statistical Interpretation for Linked Loci


**Subpopulations and F-Statistics**


**Stutter Products & Peak Height Ratios**


### Estimating the Number of Contributors


**Mixture Ratios & Deconvolution**


**Stochastic Effects & Allele Dropout**


Low Template DNA Mixtures


Probabilistic Genotyping


### U.S. Caucasian Population Data

Number of individuals: N = 361
Number of alleles: 2N = 722

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>CSF1PO</th>
<th>TPOX</th>
<th>D7S820</th>
<th>D8S1179</th>
<th>D18S51</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>0.028</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.006</td>
<td>0.525</td>
<td>0.144</td>
<td>0.014</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8.1</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.014</td>
<td>0.127</td>
<td>0.168</td>
<td>0.006</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.22</td>
<td>0.05</td>
<td>0.256</td>
<td>0.102</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>10.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.309</td>
<td>0.252</td>
<td>0.205</td>
<td>0.076</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.36</td>
<td>0.042</td>
<td>0.159</td>
<td>0.168</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.082</td>
<td>0.001</td>
<td>0.035</td>
<td>0.33</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>13.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.01</td>
<td>0.004</td>
<td>0.166</td>
<td>0.134</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>14.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>0.104</td>
<td>0.17</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.033</td>
<td>0.147</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>16.2</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.001</td>
<td>0.139</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>-</td>
<td>0.078</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.04</td>
<td>-</td>
<td>0.04</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.018</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.007</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>D21S11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>25.2</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>26.2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.159</td>
</tr>
<tr>
<td></td>
<td>28.2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>29.2</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>29.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.283</td>
</tr>
<tr>
<td></td>
<td>30.2</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>30.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>31.2</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>32.2</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>33.1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>33.2</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>34.2</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>-</td>
</tr>
</tbody>
</table>

Reference:
Basic STR Interpretation Workshop
John M. Butler & Simone N. Gittelson
Krakow, Poland