Wednesday, May 13, 2009

12:30 p.m.  Background Information and Introductions
1:00 p.m.  CE Fundamentals
2:30 p.m.  BREAK
2:45 p.m.  CE Troubleshooting
3:15 p.m.  Y-STRs
4:30 p.m.  End first day – provide homework assignment – mixture to solve

HOMEWORK

Thursday, May 14, 2009

8:00 a.m.  qPCR and low level DNA testing
9:15 a.m.  Mixture Interpretation
9:45 a.m.  BREAK
10:00 a.m.  Mixture Interpretation (cont.)
12:00 p.m.  Training concludes – awarding of certificates

Reference Lists of Relevant Articles Supplied for Each Topic
Introductions
J.M. Butler - Utah DNA Training

May 13-14, 2009

Topics and Techniques for Forensic DNA Analysis
Continuing Education Seminar

Introductions

Utah DNA Training
Salt Lake City, UT
May 13-14, 2009

Dr. John M. Butler
National Institute of Standards and Technology
john.butler@nist.gov

Outline for Training Planned

May 13 (afternoon)
• Introductions
• Capillary Electrophoresis Fundamentals
BREAK
• CE Troubleshooting
• Y-STRs
HOMEWORK

May 14 (morning)
• qPCR and low level DNA testing
• Mixture Interpretation
BREAK
• Mixture Interpretation (cont.)

Dr. John M. Butler

Experience
• University of Virginia/FBI Laboratory (1992-1995)
  – Work performed in Bruce McCord’s lab
• NIST NRC Postdoc (1995-1997)
• NIST Human Identity Project Leader (1999-present)
• Forensic DNA Typing textbook (now in its 2nd Edition)
• STRBase website: http://www.cstl.nist.gov/biotech/strbase/
• Family: wife Terilynne and 6 children
• Hobbies: reading, writing, and making PowerPoint slides

NIST Human Identity Project Team
...Bringing traceability and technology to the scales of justice...

John Butler
Group leader
Amy Decker
Becky Hill
Margaret Kline
Jan Redman
Pete Vallone

Dave Dakota (data analysis)
Angie Dolph (summer 2007)
Michelle Burns (summer 2008)

Since 2000:
>100 publications
>250 presentations
>30 training workshops

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

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

http://www.cstl.nist.gov/biotech/strbase/training.htm
Introductions
J.M. Butler - Utah DNA Training

May 13-14, 2009

Current Areas of NIST Effort with Forensic DNA

• Standards
  – Standard Reference Materials
  – Standard Information Resources (STRBase website)
  – Interlaboratory Studies
• Technology
  – Research programs in SNPs, miniSTRs, Y-STRs, mtDNA, qPCR
  – Assay and software development, expert system review
• Training Materials
  – Review articles and workshops on STRs, CE, validation
  – PowerPoint and pdf files available for download
    http://www.cstl.nist.gov/biotech/strbase/NIJprojects.htm

Contributors to These Workshop Slides

Bruce McCord
Florida International University

Angie Dolph
Marshall U/
NIST

Amy Decker
NIST

CE mixtures

NIST and NIJ Disclaimer

Funding: Interagency Agreement 2008-DN-R-121 between the National Institute of Justice and NIST Office of Law Enforcement Standards

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.

SWGDAM Disclaimer…

Background of Participants…

Your name
Your organization

What you hope to learn from this workshop

http://www.cstl.nist.gov/biotech/strbase/training.htm
CE Fundamentals
J.M. Butler - Utah DNA Training

Presentation Outline
- History and background on CE
- Separation
- Injection and sample preparation
- Detection
- Troubleshooting

My Goal:
To help you understand the basic chemistry behind DNA separations and to help make CE instruments less of a "black box"

Pioneers of Capillary Electrophoresis

Stellan Hjertén
Uppsala University (Sweden)

James Jorgenson
University of North Carolina

Barry Karger
Northeastern University

1967
First high voltage CE system (with rotating 3 mm i.d. capillaries)

1981
First "modern" CE experiments (with 75 µm i.d. capillaries)

1988/90
First DNA separations in a capillary (gel-filled/sieving polymer)

Stellan Hjertén
In 2003 at age 75

With first fully automated capillary free zone electrophoresis apparatus in 1967

Received his PhD (1967) under Professor Arne Tiselius who had developed moving boundary zone electrophoresis in 1937 (Nobel Prize in 1948)

A Brief History of Capillary Electrophoresis

- 1937 – Tiselius develops moving boundary electrophoresis
- 1967 – Hjertén uses rotating 3 mm i.d. tubes for CE
- 1981 – Jorgenson and Lukacs demonstrate first high performance CE separations with 75 µm i.d. capillary
- 1988 – Karger’s group shows DNA separations of single stranded oligonucleotides with gel-filled capillaries
- 1990 – Karger’s group shows DNA separations with sieving polymers on DNA restriction fragments
- 1991 – Grossman expands work with sieving polymers
- 1992 – Bruce McCord starts working on PCR product separations with STR allelic ladders

My Experience with CE, STRs, etc.

- May 1993 – began working in Bruce McCord’s lab at Quantico
- Sept 1993 – developed mtDNA amplicon quantitation method (used in FBI casework from 1996 to present)
- Nov 1993 – first demonstration of STR typing by CE (using dual internal standards and TH01 ladder)
- July 1995 – defended Ph.D. dissertation entitled “Sizing and Quantitation of Polymerase Chain Reaction Products by Capillary Electrophoresis for Use in DNA Typing”
- July 1995 – ABI 310 Genetic Analyzer was released

http://www.cstl.nist.gov/biotech/strbase/training.htm
My Experience with CE, STRs, etc.
(cont.)

• 1996-1997 Developed STRBase while a postdoc at NIST
• Nov 1998 – GeneTrace Systems purchased a 310; typed several hundred samples with Profiler Plus and Coffiler kits and compared results to mass spec STR analysis
• 1999-present – Run thousands of samples with all STR kits available (except PP 1.2) and developed a number of new STR multiplex systems
• Jan 2001 – Published “Forensic DNA Typing: Biology and Technology behind STR Markers” (2nd Edition in Feb 2005)
• April 2001-present – Use of ABI 3100 16-capillary array system

In the early 1990s the real question was how to transition from a gel to a capillary

• Cross-linked acrylamide gel filled capillaries were tried first
  – Reusable?
  – Bubble formation
  – Thermal degradation

• Alternative was to not use a gel at all
  – Refillable sieving polymers
  – However, resolution was poor early on

Early Work with CE and STRs

• Barry Karger’s group (1988-1990)
  – Utilized gel-filled capillaries to separate ssDNA
  – Introduced sieving polymers in the form of linear polyacrylamide to separate restriction digests
• Beckman P/ACE 2050 is introduced in 1992 as the first commercially available CE coupled to a laser to enable fluorescence detection
• John Butler and Bruce McCord (1993-1995)
  – First STR typing with single color CE using intercalating dyes and dual bracketing internal size standards
• Rich Mathies’ group (1995)
  – First STR typing with multi-color CE (and multi-capillary) using dye-labeled primers
• ABI 310 is introduced in July 1995 as the first commercially available multi-color CE

First Rapid STR Typing with Capillary Electrophoresis

Single color detection with dual internal size standards

Butler et al. (1994) BioTechniques 17: 1062-1070

Technology Implementation Takes Time – the FBI did not start running casework samples using STRs and CE until January 1999

First Rapid STR Typing with Capillary Electrophoresis

Single color detection with dual internal size standards

Butler et al. (1994) BioTechniques 17: 1062-1070

Technology Implementation Takes Time – the FBI did not start running casework samples using STRs and CE until January 1999

Why Use CE for DNA Analysis?

1. Injection, separation, and detection are automated.
2. Rapid separations are possible
3. Excellent sensitivity and resolution
4. The time at which any band elutes is precisely determined
5. Peak information is automatically stored for easy retrieval

Symbol first used in Oct 1994 at the Promega meeting when I had a poster introducing the use of CE for STR typing
Capillary Electrophoresis (CE)

Data Acquisition and Analysis

DNA Separation occurs in minutes...

Sample tray moves automatically beneath the cathode end of the capillary to deliver each sample in succession.

Flowchart of CE Sample Analysis

Extract & Quantitate DNA

Prepare Samples

Prepare Capillary And Run Buffer

Load Autosampler

Enter Sample Names into Computer Spreadsheet

Fill Capillary with Sealing Polymer Solution

Apply Voltage to Separate Components

Detect Labeled Components

Analyze Data

Typical Instruments Used for STR Typing

GeneAmp 9700

Thermal Cycler for PCR Amplification

Capillary electrophoresis instruments for separating and sizing PCR products

Single capillary

ABI 310

16-capillary array

ABI 3100

Review Article on STRs and CE


Electrophoresis 2004, JS, 1397-1412

Contents

1. Introduction .......................... 1397
2. General aspects .......................... 1397
3. Early work with CE .......................... 1400
4. Sample preparation and injection .......................... 1401
5. Sample separation .......................... 1402
6. The polymer separation matrix .......................... 1403
7. The buffer .......................... 1404
8. The capillary .......................... 1404
9. Software used .......................... 1404
10. Sample interpretation .......................... 1405
11. Assessing resolution of DNA separations .......................... 1405
12. Applications of forensic DNA testing .......................... 1406
13. Forensic casework .......................... 1407
14. DNA databasing .......................... 1408
15. Increasing sample throughput .......................... 1409
16. Capillary array electrophoresis systems .......................... 1409
17. Microchip CE systems .......................... 1409
18. Future methods for DNA typing with STR markers .......................... 1410
19. References .......................... 1410

Analytical Requirements for STR Typing


- Fluorescent dyes must be spectrally resolved in order to distinguish different dye labels on PCR products
- PCR products must be spatially resolved – desirable to have single base resolution out to >350 bp in order to distinguish variant alleles
- High run-to-run precision – an internal sizing standard is used to calibrate each run in order to compare data over time

http://www.cstl.nist.gov/biotech/strbase/training.htm
Process Involved in 310/3100 Analysis

- **Separation**
  - Capillary – 50μm fused silica, 43 cm length (36 cm to detector)
  - Buffer – TAPS pH 8.0
  - Denaturants – urea, pyrolidinone
- **Injection**
  - electrokinetic injection process (formamide, water)
  - importance of sample stacking
- **Detection**
  - fluorescent dyes with excitation and emission traits
  - CCD with defined virtual filters produced by assigning certain pixels

Ohm’s Law

- \( V = IR \) (where \( V \) is voltage, \( I \) is current, and \( R \) is resistance)

- Current, or the flow of ions, is what matters most in electrophoresis

- CE currents are much lower than gels because of a higher resistance in the narrow capillary

- CE can run a higher voltage because the capillary offers a higher surface area-to-volume ratio and can thus dissipate heat better from the ion flow (current)

DNA and Electrophoresis

“From a practical point of view it is disappointing that electrophoresis cannot be used to fractionate or analyze DNA’s on the basis of size” Olivera, Biopolymers 1964, 2, 245

\[ \mu_p = \frac{q/6\pi\eta r}{\text{small ions with high charge move fastest}} \]

As size increases so does charge!
DNA Separations in Entangled Polymer Sieving Solutions

- Size based separation due to interaction of DNA molecules with entangled polymer strands
- Polymers are not cross-linked (as in slab gels)
- “Gel” is not attached to the capillary wall
- Pumpable -- can be replaced after each run
- Polymer length and concentration determine the separation characteristics

Entangled Polymer Solutions

- Polymers are not cross-linked (above entanglement threshold)
- “Gel” is not attached to the capillary wall
- Pumpable -- can be replaced after each run
- Polymer length and concentration determine the separation characteristics
- Examples:
  - 1% HEC (hydroxyethyl cellulose)
  - 4% polyvinyl pyrrolidone
  - POP-4 and POP-6

Transient Pores Are Formed Above the Entanglement Threshold.

C < C*  C = C*  C > C*

Ogston Sieving  Reptation  Entanglement

\[ \mu \sim \mu_0 e^{-NC} \]
\[ \mu \sim 1/N \]
\[ \mu \sim f(1/CN) \]
What is in POP-4 and Genetic Analyzer Buffer?

POP-4 (4% poly-dimethylacrylamide, 8 M urea, 5% 2-pyrrolidinone)

Running buffer contains 100 mM TAPS and 1 mM EDTA (adjusted to pH 8.0 with NaOH)

TAPS = N-Tris-(hydroxymethyl)methyl-3-aminopropane-sulfonic acid

US Patent 5,552,028 covers POP-4 synthesis

See also Wenz et al. (1998) Genome Research 8: 69-80

Why TAPS instead of Tris-borate (TBE) buffer?

- TBE is temperature/pH sensitive
  - as temperature increases, the pH decreases (0.02 pH units with every 1 °C); this is the principle by which TaqGold activation works
- At lower pH, fluorescence emission of dyes decreases
- Thus when running at 60 °C on the ABI 310, if Tris-borate was used, fluorescent intensity of PCR products would be lower

Capillary Wall Coatings Impact DNA Separations

Electrophoretic flow

Solvated ions drag solution towards cathode in a flat flow profile

Capillary Resolution Differences

(A) Good resolution
(B) Poor resolution


http://www.cstl.nist.gov/biotech/strbase/training.htm
How to Improve Resolution?

1. Lower Field Strength
2. Increase Capillary Length
3. Increase Polymer Concentration
4. Increase Polymer Length

All of these come at a cost of longer separation run times

The electric field strength can influence the shape of the DNA molecule.

The electric field strength can influence the shape of the DNA molecule.

Optimal separations usually 180-200 V/cm

Resolution vs Field Strength

Resolution vs Field Strength

Resolution vs Field Strength

Resolution vs Field Strength

Process of Sizing DNA Fragments Using an Internal Standard

DNA fragment peaks are sized based on the sizing curve produced from the points on the internal size standard

Injection

http://www.cstl.nist.gov/biotech/strbase/training.htm
CE Injection Methods

**CE Injection Methods**

- Hydrodynamic (pressure)
- Electrokinetic (voltage)

**Sample Conductivity Impacts Amount Injected**

\[
\text{DNA}_\text{inj} = \frac{E \pi r^2 (\mu_{ep} + \mu_{eof}) [\text{DNA}]_\text{sample}}{\lambda_{sample}} \lambda_{buffer}
\]

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

**Two Major Effects of Sample Stacking**

1. Sample is preconcentrated. Effect is inversely proportional to ionic strength
2. Sample is focused. Ions stop moving in low electric field
3. Mobility of sample = \(\mu_{ep} = \text{velocity}/\text{electric field}\)

**Steps Performed in Standard Module**

See J.M. Butler (2005) Forensic DNA Typing, 2nd Edition; Chapter 14

- Capillary fill – polymer solution is forced into the capillary by applying a force to the syringe
- Pre-electrophoresis – the separation voltage is raised to 10,000 volts and run for 5 minutes
- Water wash of capillary – capillary is dipped several times in deionized water to remove buffer salts that would interfere with the injection process
- Sample injection – the autosampler moves to position A1 (or the next sample in the sample set) and is moved up onto the capillary to perform the injection; a voltage is applied to the sample and a few nanoliters of sample are pulsed onto the end of the capillary; the default injection is 10 kV (kilovolts) for 5 seconds
- Water wash of capillary – capillary is dipped several times in deionized water to remove any contaminating solution adhering to the outside of the capillary
- Water dip – capillary is dipped in clean water (position 2) several times
- Electrophoresis – autosampler moves to inlet buffer vial (position 1) and separation voltage is applied across the capillary; the injected DNA molecules begin separating through the POP-4 polymer solution
- Detection – data collection begins; raw data is collected with no spectral deconvolution of the different dye colors; the matrix is applied during Genescan analysis

**Typical Sample Preparation for ssDNA**

1. Perform PCR with dye-labeled primers
2. Dilute 1 µL PCR product with 24 µL deionized formamide; add 1 µL ROX-labeled internal sizing standard
3. Denature 2 minutes at 95 ºC with thermocycler
4. Cool to 4 ºC in thermocycler or ice bath
5. Sample will remain denatured for at least 3 days

http://www.cstl.nist.gov/biotech/strbase/training.htm
Comments on Sample Preparation

- **Use high quality formamide** (<100 μS/cm)!
  - ABI sells Hi-Di formamide
  - Regular formamide can be made more pure with ion exchange resin
- **Deionized water vs. formamide**
  - Water works fine but samples are not stable as long as with formamide; water also evaporates over time...
- **Denaturation with heating and snap cooling**
  - Use a thermal cycler for heating and cold aluminum block for snap cooling
  - heat/cool denaturation step is necessary only if water is substituted for formamide...

January 6, 2005 Letter from Applied Biosystems to ABI 310 Customers

- “Testing has shown that Hi-Di Formamide denatures DNA **without the need to heat samples**…”
- In other words, no heat denaturation and snap cooling needed!

Applied Biosystems Okays Use of Deionized Water for DNA Sequencing

**Technical Bulletin #1**

*Applied Biosystems 3730/3730xl DNA Analyzer*

**Subject:** Influence of Sequencing Injection Solution on 3730/3730xl DNA Analyzer Performance

*In this Bulletin:*
- Three Loading Solutions Tested on Page 1
- Loading Solutions Tested on Page 2
- Recommendations on Page 3
- Qualitats for Use on Page 6

*Three Loading Solutions Tested*

**Loading Solution Background**

*Applied Biosystems presently recommends the use of Hi-Di™ Formamide in the single-loading solution for all Applied Biosystems DNA sequencers to ensure sample preservation and integrity in evaporation. However, many users of the 3730/3730xl prefer deionized water or 400 mM EDTA solutions. These choices are driven largely by cost and usability/lactation material considerations.*

Detection

**Detection Issues**

- **Fluorescent dyes**
  - spectral emission overlap
  - relative levels on primers used to label PCR products
  - dye “blobs” (free dye)
- **Virtual filters**
  - hardware (CCD camera)
  - software (color matrix)

**Filters determine which wavelengths of light are collected onto the CCD camera**
Laser Used in ABI 310

- Argon Ion Laser
- 488 nm and 514.5 nm for excitation of dyes
- 10 mW power
- Lifetime ~5,000 hours (1 year of full-time use)
- Cost to replace ~$5,500
- Leads to highest degree of variability between instruments and is most replaced part
- Color separation matrix is specific to laser used on the instrument

Methods for Fluorescently Labeling DNA

- Intercalating Dyes (post-PCR)
- Dye-labeled nucleotide insertion during PCR
- Dye-labeled primer insertion during PCR

Fluorescent Labeling of PCR Products

- Dyes are attached to one primer in a pair used to amplify a STR marker
- Dyes are coupled to oligonucleotides (primers) through NHS-esters and amine linkages on the 5’end of the primer: Dye-(CH₂)₆-primer
- Dye-labeled oligonucleotides are incorporated during multiplex PCR amplification giving a specific color “tag” to each PCR product
- PCR products are distinguished using CCD imaging on the 310

Amine Reactive Dyes used in Labeling DNA

- FAM (Blue)
- JOE (Green)
- TAMRA (Yellow)
- ROX (Red)
- The succinimidyl ester reacts rapidly with amine linkers on DNA bases

Virtual Filters Used in ABI 310

- Visible spectrum range seen in CCD camera

Fluorescent Emission Spectra for ABI Dyes

- NED is a brighter dye than TAMRA

http://www.cstl.nist.gov/biotech/strbase/training.htm
Please Note!

- There are no filters in a 310
- It's just the choice of pixels in the CCD detector
- All the light from the grating is collected
- You just turn some pixels on and some off

Why Make a Matrix?

The matrix is the solution to a problem:
What's the contribution at any given wavelength (filter set) from each dye?

There are 4 dyes

- Remember algebra from high school?
- To solve a problem with 4 unknowns, you need 4 equations

For Example

\[ I_{540} = bx + gy + zy + rw \]
\[ I_{560} = bx + gy + zy + rw \]
\[ I_{580} = bx + gy + zy + rw \]
\[ I_{610} = bx + gy + zy + rw \]

Where
- \( b \) is the % blue labeled DNA
- \( g \) is the % green labeled DNA, etc.
- \( x, y, z, w \) are the numbers in the matrix (sensitivity to each color)

If you solve \( xyzw \) for each dye individually
Then you can determine dye contribution for any mixture

Matrix Standards (Raw Data)

Matrix Standards (After Color Separation)
The results of the calculation are in a matrix (remember linear algebra?) The values represent the percent spectral overlap from each dye. Values outside this range represent mixtures.

**Matrix File Table from an ABI 310**

<table>
<thead>
<tr>
<th>Reactions</th>
<th>B</th>
<th>G</th>
<th>Y</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B</strong></td>
<td>1.000</td>
<td>0.85</td>
<td>0.134</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td>0.850</td>
<td>1.00</td>
<td>0.752</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>0.134</td>
<td>0.752</td>
<td>1.000</td>
<td>0.110</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>0.000</td>
<td>0.004</td>
<td>0.110</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Deciphering Artifacts from the True Alleles

**Dye Blobs ("Artifacts")**

- Free dye (not coupled to primer) can be injected into the CE capillary and interfere with detection of true STR alleles
- **Dye blobs are wider and usually of less intensity** than true STR alleles (amount depends on the purity of the primers used)
- Dye blobs usually appear at an apparent size that is unique for each dye (e.g., FAM ~120 bp, PET ~100 bp)

**Dye Blob Problems with Some PCR Primers**

Individual Y-STR Locus Amplifications

<table>
<thead>
<tr>
<th>PCR product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
</tr>
<tr>
<td>DYS392</td>
</tr>
</tbody>
</table>

Note higher RFU values due to salt reduction with spin columns.
**Conclusions**

DNA typing by capillary electrophoresis involves:

1. The use of entangled polymer buffers
2. Injection by sample stacking
3. Multichannel laser induced fluorescence
4. Internal and external calibration

---

**Practical Aspects of ABI 310/3100 Use**

- **Dye Artifacts Can Be Removed With Filtration**
  - Residual dye artifacts
  - Elimination of dye blobs with Edge columns

**NIST Y-STR 20plex assay**


**Dye Artifacts Can Be Removed With Filtration**

- Residual dye artifacts
- Elimination of dye blobs with Edge columns

**NIST Y-STR 20plex assay**


**Conclusions**

DNA typing by capillary electrophoresis involves:

1. The use of entangled polymer buffers
2. Injection by sample stacking
3. Multichannel laser induced fluorescence
4. Internal and external calibration

---

**Practical Aspects of ABI 310/3100 Use**

- **ABI Genetic Analyzer Usage at NIST**
  - ABI 310 x 2 (originally with Mac, then NT)
    - 1st was purchased in 1996
    - 2nd was purchased in June 2002
  - ABI 3100 (Data collection v1.0.1)
    - Purchased in June 2002
    - Original data collection software retained
  - ABI 3130xl upgrade (Data collection v3.0)
    - Purchased in April 2001 as ABI 3100
    - Upgraded to ABI 3130xl in September 2005
    - Located in a different room

**Our Use of the ABI 3100**

- Data collection software, version 1.0.1
- **POP-6** with 36 cm capillary array
- STR kits and in-house assays for autosomal STRs, Y-STRs, and miniSTRs
- SNAPSHOT assays for mtDNA SNPS, Y-SNPS, and autosomal SNPS
- DNA sequencing for mtDNA and STR repeat sequencing

**We can routinely get more than 400 runs per capillary array by not changing the polymer between applications**

---

**NIST ABI 3100 Analysis Using POP-6 Polymer**

**High Resolution STR Typing**

**SNAPSHOT SNP Typing**

(Coding Region mtSNP 11plex minisequencing assay)

**mtDNA Sequencing (HVI)**

---

http://www.cstl.nist.gov/biotech/strbase/training.htm
Comparison of ABI 3100 Data Collection Versions

Same DNA sample run with Identifiler STR kit (identical genotypes obtained)

**ABI 3100** (36 cm array, POP-6)
- Data Collection v1.0.1
- 5s@2kV injection

**ABI 3130xl** (50 cm array, POP-7)
- Data Collection v3.0
- 5s@2kV injection

Relative peak height differences are due to “variable binning” with newer ABI data collection versions. Differences in the STR allele relative mobilities (peak positions) are from using POP-6 vs. POP-7.

Consumables for ABI 310/3100

**What we use at NIST**
- A.C.E.™ Sequencing Buffer 10X (Amresco)
  - $155/L = $0.0155/mL 1X buffer (costs 20 times less!)
  - http://www.amresco-inc.com
- 3700 POP-6 Polymer (Applied Biosystems)
  - $530 / 200 mL = $2.65/mL (costs 20 times less!)

**What ABI protocols suggest**
- 10X Genetic Analyzer Buffer with EDTA
  - $78/25 mL = $0.312/mL 1X buffer (ABI)
- 3100 POP-4 Polymer
  - $365 / 7 mL = $52/mL

2004 prices

Reduced Volume PCR Amplifications

**Advantages**
- **Lower cost** since kit contents are stretched
- Improved sensitivity perceived due to use of concentrated PCR products (since 1 uL out of a 5 uL reaction is 20% while 1 uL out of a 50 uL reaction is 2%)

**Disadvantages**
- Less volume of input DNA
  - Tighter control (improved precision) required in DNA quantitation
  - If low amount of DNA, then potential for allelic dropout (S.C.N conditions)
  - If PCR inhibitor is present, then less opportunity for dilution of inhibitor
- Evaporation impacts PCR amplification performance

**Publications**

Identifiler 5 µL PCR Protocol

Identifiler PCR amplification was carried out on a GeneAmp® 9700 using 1 ng of DNA according to kit protocols with the exception of reduced volume reactions (5 µL instead of 25 µL) and reduced cycles (26 instead of 28).

Amplification products were diluted 1:15 in Hi-Di™ formamide and G3500-LIZ internal size standard (0.3 µL) and analyzed on the 16-capillary ABI Prism® 3100 Genetic Analyzer without prior denaturation of samples.

POPs™-6 (3700 POP’s) rather than POP’s™-4 was utilized for higher resolution separations.

Allele calls were made in Genotyper® 3.7 by comparison with kit allelic ladders using the Kazaam macro (20% filter).


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

Identifiler 5 µL PCR

(lower 3100 injection; 5s@2kV instead of 10s@3kV)

**Total cost per sample = $3.87 (Fall 2002)**
Maintenance of ABI 310/3100/3130

- Syringe – leaks cause capillary to not fill properly
- Capillary storage & wash – it dries, it dies!
- Pump block – cleaning helps insure good fill
- Change the running buffer regularly

YOU MUST BE CLEAN AROUND A CE!

Overall Thoughts on the ABI 310/3100/3130

- Settling on a common instrument platform has been good for the forensic DNA community in terms of data consistency (this is also true with the use of common STR kits)
- I am concerned that the community is very dependent primarily on one company…
- I really like using the instrument and can usually get nice data from it
- Like any instrument, it has its quirks…

Thank you for your attention…

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

Questions?

See also http://www.dna.gov/research/nist
http://www.cstl.nist.gov/biotech/strbase
john.butler@nist.gov
Common Laboratory Problems
John M. Butler, PhD
National Institute of Standards and Technology

Bruce R. McCord, PhD
Florida International University

19th International Symposium on Human Identification
Hollywood, CA
October 16, 2008

Bruce McCord’s Profiles in DNA Article
Volume 6 (2), Sept 2003, pp. 10-12

Troubleshooting Capillary Electrophoresis Systems
By Bruce McCord
Associate Professor of Forensic Chemistry, Ohio University, Athens, Ohio

The key to producing good DNA separations is to understand the principles underlying the injection, separation and detection of each allele.

Effect of Temperature on allele size

Slope is 0.14 bases/degree centigrade
Therefore a small change in temperature has a big effect
(A 1-2 degree shift in temperature of the heat plate can produce an OL allele)

Troubleshooting
Bruce McCord, AAFS 2006 Workshop (Seattle, WA)
February 20, 2006

3. External Factors

- Room temperature
  - Variations in room temperature can cause mobility shifts with band shifts and loss of calibration
  - Temperature is also important due to effects of high humidity on electrical conductance

- Cleanliness
  - Urea left in sample block can crystallize and catalyze further crystal formation causing spikes, clogs and other problems.
  - Best bet is to keep polymer in system and not remove or change block until polymer is used up.

Outline for This Section

1. Chemistry/molecular biology problems – stutter, -A, degradation, inhibition, low copy #
2. Sample and buffer problems – formamide, urea, water, salt concentration, free dye (“dye blobs”)
3. External factors – power supply, room temperature, cleanliness, voltage leaks
4. Instrument problems – optical system, capillary clogging, air bubbles, syringe leaks
5. Troubleshooting benchmarks/QC monitoring

CE Troubleshooting
J.M. Butler - Utah DNA Training
May 13-14, 2009
Temperature Effects: "OL" Alleles

"OL Alleles"

"OL alleles" - look at the 250 peak

"OL allele re-injected"
And the 250 peak...

Monitoring Room Temperature Over Time

And the 250 peak...

Monitoring Room Temperature Over Time

And the 250 peak...

Monitoring Room Temperature Over Time

And the 250 peak...

Monitoring Room Temperature Over Time

And the 250 peak...

Monitoring Room Temperature Over Time

And the 250 peak...
Cleanliness

- Urea sublimates and breaks down to ionic components - these find a path to ground
- Similarly wet buffer under a vial creates paths to ground
- Capillary windows must be clear or matrix effects will occur
- Laser will often assist in this process
- Vial caps will transfer low levels of DNA to capillary

4. Instrumental Factors

- Optical System
  - Sensitivity changes with age, capillary diameter, capillary cleanliness, instrument calibration
- Fluidic System
  - Effects of bubbles, dust, urea crystals, leaks in syringe and capillary ferrule
- Matrix Calculations
  - Changes in buffer, optics, sample dye can alter the software calibrations
- Capillary Problems
  - Chemisorbed materials on capillary surface can produce osmotic flow, DNA band broadening and inconsistent resolution (meltdowns)

The Detection Window

- Make sure that the capillary window is lined up (if it is not, then no peaks will be seen)
- Window may need to be cleaned with ethanol or methanol
- Review start of raw data collection

Beware of Urea Crystals

- Urea crystals have formed due to a small leak where the capillary comes into the pump block
- Urea sublimates and can evaporate to appear elsewhere
- Use a small balloon to better grip the ferrule and keep it tight
- Pump block should be well cleaned to avoid problems with urea crystal formation

These spikes resulted from buffer dilution with poor water. The problem disappeared when the HPLC grade water was purchased to dilute buffer and samples

http://www.cstl.nist.gov/biotech/strbase/training.htm
Storage when ABI 310 is not in use

- Keep inlet of capillary in water...if it dries out then urea crystals from the polymer will clog the opening
- The waste vial (normally in position 3) can be moved into position
- A special device can be purchased from Supelco to rinse the capillary off-line
- Store in distilled water
- Note that the laser is on when the instrument is on

Remember that the water in the open tube will evaporate over time...

Buffer Issues

- The buffer and polymer affect the background fluorescence-affecting the matrix
- Urea crystals and dust may produce spikes
- High salt concentrations may produce reannealing of DNA
- High salt concentrations affect current
- Low polymer concentrations affect peak resolution

Capillary Meltdowns

(A) Good resolution
(B) Poor resolution

Identifier data

Meltdowns can be permanent or transitory as we have seen these may result from sample contamination effects

Does the capillary need to be replaced?

No! The next injection looks fine...

Meltdowns may be the result of

- Bad formamide
- Excess salt in sample/renaturation
- Water in the polymer buffer
- Syringe leak or bottom out
- Poisoned capillary
- Conductive polymer buffer due to urea degradation
- Crack/shift in capillary window
- Detergents and metal ions

5. Troubleshooting benchmarks

- Monitor run current
- Observe syringe position and movement during a batch
- Examine ILS (ROX) peak height with no sample
- Observe “250 bp” peak in GS500 size standard
- Monitor resolution of TH01 9.3/10 in allelic ladder and size standard peak shapes
- Keep an eye on the baseline signal/noise
- Measure formamide conductivity
- Reagent blank – are any dye blobs present?
- See if positive control DNA is producing typical peak heights (along with the correct genotype)
Measurement of Current

- \( V/I = R \) where \( R \) is a function of capillary diameter, buffer, and buffer viscosity.

- In a CE system the voltage is fixed, thus changes in resistance in the capillary will be reflected in the current observed.

- Air bubbles, syringe leaks, alternate paths to ground, changes in temperature, changes in zeta potential, and contamination, will be reflected in the current.

- A typical current for a CE system with POP4 buffer is \( 8-12 \mu A \) (microamps).

Syringe Travel

- The ABI 310 instrument also keeps track of the position of the syringe (in the log file).

- Depending on the resistance to flow, the syringe will travel different lengths.

- Syringe leaks may be reflected in a longer distance traveled prior to each injection.

- These leaks occur around the barrel of the syringe and at the connection to the capillary block.

Use of ABI 310 Log File to Monitor Current and Syringe Travel

Use of ABI 310 log file to track current and syringe position during CE analysis.

Dye Blobs in the Negative Control Sample

Dye blobs in the negative control sample indicating potential contamination or incorrect sample preparation.

Measuring Formamide Conductivity

Measuring formamide conductivity in the negative control sample.

Conclusion:

Troubleshooting is more than following the protocols.

- Monitoring conductivity of sample and formamide.
- Keeping track of current and syringe position in log.
- Watching the laser current.
- Watching and listening for voltage spikes.
- Monitoring room temperature and humidity.
Multiplex_QA Article Published

Research Article

Multiplex_QA: An exploratory quality assessment tool for multiplexed electrophoretic assays

Multiplex_QA is a data analysis tool for monitoring short- and long-term changes in the performance of multiplexed electrophoretic assays. Particularly, the commercial assay format based on the fluorophore lysis method, the Luminex platform is used by the human forensic identity community. A custom of quality metrics are calculated from the signals collected at the internal size standard included in many of the multiplex assays. Three quality metrics are included in the sample: resolution, sensitivity, and reproducibility. Resolution is measured using internal size standard peaks in the electrophoretic patterns. Interchanging graphical display enable the identification of changes in the quality metrics over time. Evaluation of metrics as an exploratory research tool for documenting data quality. The current version of the tool is a part of the STRBase package, with special use for data analysis and to test the tool in various scenarios.

Multiplex_QA Overview

- Research tool that provides quality metrics to review instrument performance over time (e.g., examines resolution and sensitivity using internal size standard peaks)
- Runs with Microsoft Excel macros. Requires STR data to be converted with NCBI’s BatchExtract program into numerical form

Available for download from STRBase:
http://www.cstl.nist.gov/biotech/strbase/software.htm

Acknowledgments

NIST Human Identity Project Team

Leading the Way in Forensic DNA

John Butler (Leader)
Margaret Kline
Pete Vallone
Jan Redman
Amy Decker
Betsy Hill
Dave Duewer

Funding from interagency agreement 2008-DN-R-121 between the National Institute of Justice and the NIST Office of Law Enforcement Standards

Many wonderful collaborators from industry, university, and government laboratories.

Bruce McCord (Florida International University) for many of the slides

http://www.cstl.nist.gov/biotech/strbase/training.htm
Y-Chromosome STRs
J.M. Butler - Utah DNA Training

Presentation Outline

- Why Y is of interest in human identity testing
- Y-STR markers and kits available
- Different population databases and statistics for reporting matches
- Mutation rates, duplications, and deletions and their impact on interpretation
- Value of additional Y-STR loci (beyond the Yfiler 17)

Role of Y-STRs and mtDNA Compared to Autosomal STRs

- Autosomal STRs provide a higher power of discrimination and are the preferred method whenever possible
- Due to capabilities for male-specific amplification, Y-chromosome STRs (Y-STRs) can be useful in extreme female-male mixtures (e.g., when differential extraction is not possible such as fingernail scrapings)
- Due to high copy number, mitochondrial DNA (mtDNA) may be the only source of surviving DNA in highly degraded specimens or low quantity samples such as hair shafts

Lineage Markers:

Advantages
- Extend possible reference samples beyond a single generation (benefits missing persons cases and genetic genealogy)
- Family members have indistinguishable haplotypes unless mutations have occurred

Disadvantages
- Lower power of discrimination due to no genetic shuffling with recombination
- Family members have indistinguishable haplotypes unless mutations have occurred

http://www.cstl.nist.gov/biotech/strbase/training.htm
Genetic Genealogy Companies

The rapidly growing field of genetic genealogy is expanding the use of mtDNA and Y-STRs.

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

Questions #45a & #45b

• Is your lab using or validating Y-STRs?
  – 51 Yes (30%)
    28 Yfiler, 15 PowerPlex Y, some both kits
  – 114 No
  – 6 no response

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

• Y-STR data can be entered in CODIS similar to entering the current STR loci in CODIS. Do you think CODIS should include Y-STR loci in Popstats calculations?
  – Yes – 116 (68%)
  – No – 18
  – No response – 37

Past Law & Order episodes have discussed the “CODIS Y-STR database” and its capabilities for familial searching...

Value of Y-Chromosome Markers

J.M. Butler (2005) Forensic DNA Typing, 2nd Edition; Table 9.1

<table>
<thead>
<tr>
<th>Application</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forensic casework on sexual assault evidence</td>
<td>Male-specific amplification (can avoid differential extraction to separate sperm and epithelial cells)</td>
</tr>
<tr>
<td>Paternity testing</td>
<td>Male children can be tied to fathers in motherless paternity cases</td>
</tr>
<tr>
<td>Missing persons investigations</td>
<td>Patrilineal male relatives may be used for reference samples</td>
</tr>
<tr>
<td>Human migration and evolutionary studies</td>
<td>Lack of recombination enables comparison of male individuals separated by large periods of time</td>
</tr>
<tr>
<td>Historical and genealogical research</td>
<td>Surnames usually retained by males; can make links where paper trail is limited</td>
</tr>
</tbody>
</table>

Y-STRs can permit simplification of male DNA identification in sexual assault cases

http://www.cstl.nist.gov/biotech/strbase/training.htm
Forensic Advantages of Y-STRs

- **Male-specific amplification** extends range of cases accessible to obtaining probative DNA results (e.g., fingernail scrapings, sexual assault without sperm)
- **Technical simplicity due to single allele profile**; can potentially recover results with lower levels of male perpetrator DNA because there is not a concern about heterozygote allele loss via stochastic PCR amplification; number of male contributors can be determined
- Courts have already widely accepted STR typing, instrumentation, and software for analysis (Y-STR markers just have different PCR primers)
- Acceptance of statistical reports using the counting method due to previous experience with mtDNA

Scenarios Where Y-STRs Can Aid Forensic Casework

- Sexual assaults by vasectomized or azoospermic males (no sperm left behind for differential extraction)
- Extending length of time after assault for recovery of perpetrator’s DNA profile (greater than 48 hours)
- Fingernail scrapings from sexual assault victims
- Male-male mixtures
- Other bodily fluid mixtures (blood-blood, skin-saliva)
- Gang rape situation to include or exclude potential contributors
- Confirmation of amelogenin Y negative males

Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)
- Paternal lineages possess the same Y-STR haplotype (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another
- Not as informative as autosomal STR results
  - More like addition (10 + 10 + 10 = 30) than multiplication (10 x 10 x 10 = 1,000)

Y-STRs Identify the Male Component even with Excess Female DNA

- Male DNA only
- Mixture of Male and Female DNA
- 800X female DNA

Confirmation of Amelogenin Negative Males

- Often due to deletion of that entire region of the Y-chromosome rather than a primer binding site mutation
- Most commonly seen in males of Indian subcontinent origin
- Y-STRs help demonstrate that the AMEL X sample is really male
    - 12/649 Malaysian males showed no AMEL Y
    - 5/77 Nepal males showed no AMEL Y
- AMEL X only
- Normal AMEL X,Y male
- A new section on the NIST STRBase website will be created on this topic soon

What has happened in the past few years...

- “Full” Y-chromosome sequence became available in June 2003; over 350 Y-STR loci identified (only ~20 in 2000)
- Selection of core Y-STR loci (SWGDAM Jan 2003)
- Commercial Y-STR kits released
  - PowerPlex Y (9/03), Yfiler (12/04)
- Many population studies performed and databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

http://www.cstl.nist.gov/biotech/strbase/training.htm
History of Y-STR Marker Discovery

1992 - DYS19 (Roewer et al.)
1994 - YCAI a/b, YCAII a/b, DYS356 (Mathias et al.)
1996 - DYS358III, DYS359, DYS367, DYS352, DYS351 (Roewer et al.)
1997 - DYS386, DYS388 (Kaye et al.)
1998 - DYS385 a/b (Schneider et al.)
1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)
2000 - DYS354, DYS365, DYS367, DYS352, DYS351 (Kaye et al.)
2000 - G09411 (DYS462), G10123 (de Knijff unpublished)
2001 - DYS441, DYS442 (ida et al.)
2002 - DYS445, DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d (Redd et al.)
2002 - DYS468-DYS476 (129 new Y STRs; Manfred Kayser GDB entries)
2003 - DYS397-DYS411 (50 new Y STRs; Manfred Kayser GDB entries)
2004-2006 - DYS441, DYS442 (Iida et al.)
2002 - DYS443, DYS444, DYS445 (Iida et al.)
2000 - DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d (Redd et al.)
2002 - DYS468-DYS476 (129 new Y STRs; Manfred Kayser GDB entries)
2003 - DYS597-DYS604 (50 new Y STRs; Manfred Kayser GDB entries)
2004-2006 - DYS448-726 (GDB entries)

Y-STR Typing of Duplicated Regions

"multi-copy loci"

Multiple primer binding sites occur giving rise to more than one PCR product for a given set of primers

Y-STR loci are often counted by the number of amplicons rather than the number of PCR primer pairs

Physical Map of the Human Y-Chromosome


Y-STR Typing of Duplicated Regions

"multi-copy loci"

Multiple primer binding sites occur giving rise to more than one PCR product for a given set of primers

Y-STR loci are often counted by the number of amplicons rather than the number of PCR primer pairs

Core Y-STR Characteristics

<table>
<thead>
<tr>
<th>STR Marker</th>
<th>Position (Mb)</th>
<th>Repeat Motif</th>
<th>Allele Range</th>
<th>Mutation Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS393</td>
<td>3.17</td>
<td>AGAT</td>
<td>8-17</td>
<td>0.05%</td>
</tr>
<tr>
<td>DYS19</td>
<td>10.12</td>
<td>TTAG</td>
<td>10-19</td>
<td>0.20%</td>
</tr>
<tr>
<td>DYS391</td>
<td>12.54</td>
<td>TCTA</td>
<td>6-14</td>
<td>0.40%</td>
</tr>
<tr>
<td>DYS439</td>
<td>12.95</td>
<td>AGAT</td>
<td>8-15</td>
<td>0.38%</td>
</tr>
<tr>
<td>DYS389 III</td>
<td>13.05</td>
<td>[TCGG] [TCTA]</td>
<td>9-17</td>
<td>0.20%, 24-34</td>
</tr>
<tr>
<td>DYS438</td>
<td>13.38</td>
<td>TTCT</td>
<td>6-14</td>
<td>0.09%</td>
</tr>
<tr>
<td>DYS390</td>
<td>15.71</td>
<td>[TCTA] [TCGG]</td>
<td>17-28</td>
<td>0.32%</td>
</tr>
<tr>
<td>DYS385 a/b</td>
<td>19.19, 19.23</td>
<td>GAAA</td>
<td>7-28</td>
<td>0.23%</td>
</tr>
<tr>
<td>DYS392</td>
<td>20.97</td>
<td>AT</td>
<td>6-20</td>
<td>0.05%</td>
</tr>
</tbody>
</table>

Positions (in megabases) along the Y-chromosome were determined with NCBI build 35 (Mar 2004) using BLAT. Allele ranges represent the full range of alleles reported in the literature. Mutation rates summarized from YHRD (http://www.yhrd.org, accessed 6 Apr 2005).

http://www.cstl.nist.gov/biotech/strbase/training.htm
Y-Chromosome STRs

J.M. Butler - Utah DNA Training

Y-Chromosome Standard NIST SRM 2395

Available Y-STR Loci, Kits and Databases

Haplotype Databases for Y-STR Kits

Y-STR Database Content (as of May 2009)

http://www.cstl.nist.gov/biotech/strbase/training.htm
SWGDAM Guidelines on Y-STR Interpretation

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.1) Population Databases

- Loci on NRY should be considered linked as a single locus
- Source of population database should be documented
- Relevant population(s) for which the frequency will be estimated should be identified
- Consolidated US Y-STR database should be used for population frequency estimation

http://www.usystrdatabase.org

(5.2) Haplotype Searches

- Should be conducted using all loci for which results were obtained from the evidentiary sample
- In cases where less information is obtained from the known sample, only those loci for which results were obtained from both the known and evidentiary sample should be used in the population database search

(5.3) Haplotype Frequency Estimation

- Counting method endorsed with application of a confidence interval to correct for database size and sampling variation
- Reporting a haplotype count with a confidence interval is acceptable as a factual statement regarding observations in the database

\[
p = 1,96, \sqrt{\frac{(p(1-p)}{n}} \quad 1 - (0.05)^{1/n}
\]

where \( p = \frac{x}{n}, n = \) database size, \( x = \) number of haplotypes in database


(5.4) Y-STR Mixtures

- Calculations can be performed for probability of exclusion and likelihood ratios

http://www.cstl.nist.gov/biotech/strbase/training.htm
Y-Chromosome STRs  
J.M. Butler - Utah DNA Training  
May 13-14, 2009

**SWGDAM Y-STR Interpretation Guidelines**  
Section 5. Statistical Interpretation

(5.5) Joint Match Probability

- **The product rule may be utilized** to combine the autosomal STR genotype match probability and Y-STR haplotype frequency information

- Citation to Walsh et al. (2008) Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Sci. Int.* 174: 234-238

---

**Example Y-STR Haplotype**

<table>
<thead>
<tr>
<th>Core US Haplotype</th>
<th>Matches by Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YHRD (9 loci)</td>
</tr>
<tr>
<td></td>
<td>- 7 matches in 27,773</td>
</tr>
<tr>
<td></td>
<td>YHRD (11 loci)</td>
</tr>
<tr>
<td></td>
<td>- 0 matches in 6,281</td>
</tr>
<tr>
<td></td>
<td>ReliaGene (11 loci)</td>
</tr>
<tr>
<td></td>
<td>- 0 matches in 3,403</td>
</tr>
<tr>
<td></td>
<td>PowerPlex Y (12 loci)</td>
</tr>
<tr>
<td></td>
<td>- 0 matches in 4,004</td>
</tr>
<tr>
<td></td>
<td>Yfiler (17 loci)</td>
</tr>
<tr>
<td></td>
<td>- 0 matches in 3,561</td>
</tr>
</tbody>
</table>

---

**Frequency Estimate Calculations**

In cases where a Y-STR profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

\[ p = \frac{X}{N} \]

An upper bound confidence interval can be placed on the profile's frequency using:

\[ p \pm 1.96 \sqrt{\frac{p(1-p)}{N}} \]

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

---

**SWGDAM Y-STR Interpretation Guidelines**  
Section 5. Statistical Interpretation

(5.6) Population Substructure

- **Studies have shown that** \( F_{st} \) values are very small for most populations

- **Use of the counting method** that incorporates the upper bound estimate of the count proportion offers an appropriate and conservative statistical approach to evaluating the probative value of a match

No need to use theta correction, but no discussion of partial profiles

---

**When there is no match...**

In cases where the profile has not been observed in a database, the upper bound on the confidence interval is

\[ 1 - \alpha \frac{1}{N} \]

where \( \alpha \) is the confidence coefficient (0.05 for a 95% confidence interval) and N is the number of individuals in the database.

\[ 1 - \alpha \frac{1}{N} = 1 - (0.05)^{\frac{1}{4,004}} = 0.000748 \]

\[ = 0.075\% \ (\sim 1 \ \text{in 1,340}) \]

If using database of 2,443, then the best you can do is 1 in 816

---
Y-Chromosome STRs
J.M. Butler - Utah DNA Training

May 13-14, 2009

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

The Meaning of a Y-Chromosome Match

Conservative statement for a match report:

The Y-STR profile of the crime sample matches the Y-STR profile of the suspect (at xxx number of loci examined). Therefore, **we cannot exclude the suspect** as being the donor of the crime sample. In addition, we cannot exclude all patrilineal related male relatives and an unknown number of unrelated males as being the donor of the crime sample.

Y-STR Mutations

Mutations will impact kinship testing involving Y-STRs
(e.g., use of a paternal relative as a reference for a missing persons case)

NIST Work with Father-Son Samples

- Samples obtained from paternity testing laboratory as buccal swabs, extracted with DNA-IQ, quantified, diluted to 0.5 ng/µL
- To-date: 100 father-son pairs of U.S. Caucasian, African American, Hispanic, and Asian (800 samples)
- Verified autosomal STR allele sharing with Identifiler (QC for gender and potential sample switches)
- Typed with Yfiler (17 Y-STRs) – examined mutations

Separating Brothers with 47 Y-STRs

- Two suspected brothers (ZT79338 and ZT79339) are part of our ~660 U.S. sample dataset at NIST.
- Thus far, we have evaluated 47 Y-STR allele calls on these samples.
- A mutation at DYS391 separates these individuals (one contains allele 11 and the other allele 10).
- These samples share autosomal STR alleles and contain identical mtDNA sequences.

Y-STR Mutation Rates for the 17 Yfiler Loci

<table>
<thead>
<tr>
<th>Yfiler REF locus</th>
<th>Literature Summary*</th>
<th>NIST Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TOTAL</td>
</tr>
<tr>
<td>DYf1919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf389I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf389II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf392</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf394</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYf395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA-H4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Literature summary from www.YHRD.org and papers in press


http://www.cstl.nist.gov/biotech/strbase/training.htm
Mutations Seen in 100 African American Father-Son Pairs

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Sample</th>
<th>Locus</th>
<th>Allele (Mother)</th>
<th>Allele (Father)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>1057</td>
<td>Y-DATA</td>
<td>11</td>
<td>9</td>
<td>loss of 2 repeats</td>
</tr>
<tr>
<td>African American</td>
<td>465</td>
<td>DYS389I</td>
<td>14,20</td>
<td>13,20</td>
<td>loss of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>585</td>
<td>DYS389I</td>
<td>14,32</td>
<td>15,33</td>
<td>gain of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>185</td>
<td>DYS390</td>
<td>24</td>
<td>23</td>
<td>loss of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>925</td>
<td>DYS390</td>
<td>15</td>
<td>16</td>
<td>gain of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>825</td>
<td>DYS390</td>
<td>18</td>
<td>18</td>
<td>gain of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>520</td>
<td>DYS390</td>
<td>18</td>
<td>18</td>
<td>gain of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>705</td>
<td>DYS390</td>
<td>22</td>
<td>22</td>
<td>gain of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>725</td>
<td>DYS390</td>
<td>22</td>
<td>22</td>
<td>gain of 1 repeat</td>
</tr>
<tr>
<td>African American</td>
<td>975</td>
<td>DYS390</td>
<td>17,2,19,20</td>
<td>17,2,19,20</td>
<td>Duplication</td>
</tr>
<tr>
<td>African American</td>
<td>335</td>
<td>DYS389I</td>
<td>19,20</td>
<td>19,20</td>
<td>Duplication</td>
</tr>
<tr>
<td>African American</td>
<td>335</td>
<td>DYS389I</td>
<td>19,20</td>
<td>19,20</td>
<td>Duplication</td>
</tr>
<tr>
<td>African American</td>
<td>335</td>
<td>DYS390</td>
<td>19,20</td>
<td>19,20</td>
<td>Duplication</td>
</tr>
</tbody>
</table>

Mutations in both DY3458 and DY3635 were observed in father and son 16B

PowerPlex Y Population Study

Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America

- DYS19
- DYS389I
- DYS439
- DYS389II
- DYS438
- DYS437
- DYS392
- DYS393
- DYS390
- DYS385 a/b
- Y GATA H4
- DYS437
- DYS438
- DYS390
- DYS385 a/b

Locus Duplication and Deletion

Events that impact Y-STR interpretation

Duplication Observed at DYS390

Yfiler data

Concordant with our previous PowerPlex Y and 20plex data

PowerPlex Y results on sample from Ann Marie Gross (MN BCA)

http://www.cstl.nist.gov/biotech/strbase/training.htm
DYS448 Triplication
Seen in Both Father and Son

Demonstrates full inheritance of this Y-STR locus triplication

DYS389I, DYS389II, DYS439 Deletions
Seen in Both Father and Son

Full inheritance of these Y-STR locus deletions

Duplication at Multiple Loci with Single-Source Sample

Y-chromosome mapping

PowerPlex Y data

DYS389I DYS439 DYS389II

Entire region of Y-chromosome has likely been duplicated and then diverged

Most duplications have a single repeat spread in allele patterns

Duplication and Divergence Model

Locus # dup >1 repeat
DYS19 23 2
DYS389I 5 0
DYS389II 9 2
DYS390 1 0
DYS391 3 1
DYS392 0 0
DYS393 3 0
DYS385a/b 17 0

*from www.yhrd.org, literature, and our work

Since single-step mutations are most common, then single repeat spacing in duplicated alleles is expected

Deciphering between a Mixture of Multiple Males and Locus Duplication

• Note the number of loci containing >1 allele (other than multi-copy DYS385)
• Consider relative position on the Y-chromosome if multiple loci have two alleles
• See if repeat spread is >1 repeat unit
• Examine DYS385 for presence of >2 alleles

Locus duplication along the Y-chromosome is in many ways analogous to heteroplasmy in mitochondrial DNA, which depending on the circumstances can provide greater strength to a match between two DNA samples.

http://www.cstl.nist.gov/biotech/strbase/training.htm
Deletions of some Y-STRs can be an inadvertent diagnosis of male infertility


- AZFa deletion (<1 in 100,000 men): expected to lack DYS389I/II, DYS437, DYS438, DYS439
- AZFb deletion (very rare): expected to lack DYS385 and DYS392
- AZFc deletion (1 in 4,000 men): expected to lack DYS464
- Possible that “incomplete” haplotypes are not being submitted to the Y-STR haplotype databases
- Thus, Y-STRs are not neutral with respect to fertility information

Practical Information on Y Deletions

- If DYS458 is deleted in Yfiler, then your sample is likely to lack an Amelogenin Y amplicon as DYS458 and AMEL Y are 1.13 Mb apart on the short arm of the human Y-chromosome
- Many Y-chromosomes are more complicated than originally thought!

Value of Additional Loci

Promega sells a Y-deletion test for infertility testing

http://www.cstl.nist.gov/biotech/strbase/training.htm
Y-Chromosome STRs
J.M. Butler - Utah DNA Training

May 13-14, 2009

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

Going Beyond Commercial Y-STR Kits

- Most forensic DNA laboratories (certainly in the U.S.) will only use commercially available kits due to quality control issues.
- Using these kits as a starting point, are there additional loci that would be beneficial in separating samples with common types, which could be advocated to companies for possible future adoption in Y-STR kits?
- Is it possible to regularly resolve individuals from the same paternal lineage (e.g., fathers and sons) if enough Y-STRs are examined?

Data Set Used to Examine Common Types

- Yfiler kit (17 Y-STR loci) run on all NIST male U.S. population samples – makes up ~20% of Applied Biosystems database – submitted to the YHRD
- Additional 20 Y-STR loci run on full set of NIST population samples (and several less polymorphic ones only on subset of samples)

Subdividing Common Types with Additional Loci

656 males from 3 U.S. populations

U.S. Haplotype

(19, 389I, 390, 391, 392, 393, 385 a/b) (26)
(7) (15) (4)

PowerPlex Y

(437) (42)
(12) (1) (1)

Yfiler

(448, 456, 458, 635, H4)

Most common type

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 13
DYS393 – 13
DYS385 a/b – 11, 14

A) Identical: DYS…444, 446, 485, 495, 505, 508, 534, 540, 556
B) Subdivide into two groups (2)(1): DYS…449, 463, 520, 532, 533, 557, 570, 594, 643
C) Subdivide into three groups (1)(1)(1): DYS522, DYS576

Adding either DYS522 or DYS576 to Yfiler loci resolves all 26 samples with the most common type.
Y-Chromosome STRs

J.M. Butler - Utah DNA Training

May 13-14, 2009

Subdividing Unresolved Yfiler Haplotypes (1)

Subdividing Unresolved Yfiler Haplotypes (2)

Identifiler Results from Two Unresolved Hispanic Males

Following Typing with 37 Y-STR Loci

http://www.cstl.nist.gov/biotech/strbase/training.htm
Summary on Subdividing Common Types

- 640 haplotypes were observed in the 656 U.S. population samples with the Yfiler loci: 626 were unique, 2 were observed 3 times, and 12 haplotypes were observed twice.
- With the addition of 20 new Y-STR loci, all but two sample pairs are resolved.
- In this sample set, the 7 Y-STRs (DYS532, DYS522, DYS576, DYS570, DYS565, DYS449, DYS534) have the same ability to resolve the sample haplotypes as all 20 new loci.
- These 7 loci will be the focus of future studies and multiplex assays.

NIST Activities with Y-STRs

- Sequenced variant alleles – http://www.cstl.nist.gov/biotech/strbase/STRseq.htm

26 publications since 2001 on NIST Y-chromosome work

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Capillary Electrophoresis Literature Listing

Prepared by John Butler (NIST) – May 2009

General Information


Early Work


ABI Prism 310 Genetic Analyzer


ABI 3100 and Other Capillary Array Systems


Energy Transfer Dyes


DNA Separation Mechanisms

Useful References on Y-Chromosome Markers and Assays
Prepared by John Butler (May 2009)

Y-STR Haplotype Databases (see http://www.cstl.nist.gov/biotech/strbase/y_strs.htm)
Genealogy Y-STR Database: http://www.ysearch.org/

ISFG Recommendations (and NIST recommendations on nomenclature)

STR Assays and Kits

Additional Loci
Useful References on Y-Chromosome Markers and Assays
Prepared by John Butler (May 2009)

Population Variation and Data Interpretation


Duplication/Deletion


Mixture Statistics


Joint Match Probability


Mutation Rates


Useful References on Y-Chromosome Markers and Assays  
Prepared by John Butler (May 2009)

Other NIST Publications


Low-Copy Number (LCN) and Touch DNA Evidence


Literature Listing prepared by John Butler (May 2009)


MIXTURES


RECOMMENDATIONS


STRATEGY


Gill, P. et al. (1998) Interpretation of simple mixtures when artifacts such as stutters are present—with special reference to multiplex STRs used by the Forensic Science Service. Forensic Sci. Int. 95: 213-224.


SOFTWARE


FSS-i3 software (from FSS and Promega):
http://www.promega.com/fss/3
http://www.forensic.gov.uk/forensic_t/i3/index.htm

GeneMapperID-X (from Applied Biosystems):
http://idx.appliedbiosystems.com


TrueAllele (from Cybogenetics):
http://www.cybogen.com/

STATISTICS


INTERLABORATORY STUDIES


NIST Mixture Interpretation Study 2005 (MIX05): http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm
The German Stain Commission: recommendations for the interpretation of mixed stains

P. M. Schneider · R. Fimmers · W. Keil · G. Molsberger · D. Patzelt · W. Pflug · T. Rothämel · H. Schmitter · H. Schneider · B. Brinkmann

Received: 8 April 2008 / Accepted: 16 April 2008 / Published online: 5 July 2008
© Springer-Verlag 2008

Abstract In the course of forensic DNA analysis, the interpretation of DNA profiles of mixed stains, i.e. cell material from more than a single donor, has become increasingly more important. The German Stain Commission, a joint commission of Institutes of Forensic Science and Legal Medicine, has therefore developed guidelines aiming to harmonize the evaluation of mixed stains in German criminal cases.

Keywords Short tandem repeat typing · Biostatistical analysis · Likelihood ratio · Probability of exclusion · Mixtures

Preface

Since the beginning of forensic stain analysis, mixed stains have been observed [1, 2]. Over the past few years, they have...
gained importance as evidence due to improved analytical methods and the enormous increase in the numbers of investigated stains [3, 4]. While the interpretation of single source stains usually does not cause problems [5], the evaluation and interpretation of mixed DNA stains requires particular attention [6–8]. Our recommendations – first published in German [9] – are intended to build a framework for an adequate means of treating typical cases. However, it is beyond the scope of these basic recommendations to address all possible constellations.

Definitions

A stain exhibiting more than two alleles in a single DNA system\(^1\) shall be considered a mixed stain except in the case of genetic irregularities (e.g., trisomy, somatic mosaicism, or duplication). If more than two alleles are observed in at least two DNA systems, the presence of a mixed stain shall be assumed.

The number of possible contributors to a mixed stain shall be derived, if possible:

- In general, the presence of not more than four alleles in a given system allows the assumption of at least two independent stain donors.
- In general, the presence of not more than six alleles in a given system allows the assumption of at least three independent stain donors.
- In general, if more than six alleles are observed in a given system, the exact number of stain donors cannot be reliably determined.

Classification of mixed stains

Type A has no obvious major contributor with no evidence of stochastic effects.\(^2\) Type B has clearly distinguishable major and minor DNA components; consistent peak height ratios of approximately 4:1 (major to minor component) across all heterozygous systems, and no evidence of stochastic effects. Type C has mixtures with no major component(s) and evidence of stochastic effects.

Evaluation criteria

Peak analysis

The morphology of a peak shall be typical and fully consistent with an allele of a given short tandem repeat system. Generally, reproducible peaks with heights >50 relative fluorescence units (RFU) can be considered regular peaks if the noise of the baseline is low and the number of PCR cycles recommended by the manufacturer was used.

The presence of peaks exhibiting a low signal strength (i.e., typically below 100 RFU) and/or peaks exhibiting clearly variable intensities shall be annotated in the table of observed alleles. Tables in the final report shall be accompanied by a legend explaining the designations of peak characteristics.

Stutter peaks

Both \(n-1\) and \(n+1\) stutter peaks may occur. Their heights depend on the DNA systems and the amplification conditions. A stutter peak may, in certain cases, exhibit up to 15% of the height of the corresponding main peak. Furthermore, the following shall be considered for the evaluation of a stutter peak:

- The relative stutter intensities of the alleles of a locus, as well as those between loci of a multiplex amplification.
- The possibility that a stain allele is in the position of a stutter peak.

In case of reasonable doubt, a peak in the position of a stutter peak shall be considered a true allele and part of the DNA profile and shall be included in the biostatistical calculation.

Inclusion/exclusion criteria

Inclusion

If all alleles of a person in question are uniformly present in a mixed stain, the person shall be considered a possible contributor to the stain.

Exclusion

If alleles of a person in question are not present in a mixed stain, the person shall not be considered as a possible contributor to the stain.

Grey area between inclusion and exclusion

The following effects may occur in type C mixtures due to imbalances between the mixture components and may cause

---

\(^1\) A DNA system is a genetic locus exhibiting a short tandem repeat polymorphism amplified with a pair of defined primers using the polymerase chain reaction (PCR).

\(^2\) DNA profiles obtained from the amplification of samples with low DNA content and/or poor DNA quality, where the occurrence of allelic drop out and/or locus drop out has to be assumed.
difficulties in reaching an unambiguous decision about inclusion or exclusion across all analyzed DNA systems:

- Locus drop out and allelic drop out (e.g., caused by the sensitivity of the amplification system, as well as by stochastic effects).
- Allelic drop out is more likely to occur for longer than for shorter alleles, and in particular for DNA systems with long amplicon sizes.

Additional criteria

In every case, the decision about inclusion or exclusion shall be made after careful consideration of the issues described under the “Grey area between inclusion and exclusion” section. The reasons shall be explained in detail. If appropriate, it shall be stated why a clear decision about inclusion or exclusion was not possible.

Biostatistical calculations for mixed stains

Basis

The basis for all calculations is the knowledge of the allele frequencies in the relevant population.

Probability of exclusion \( (P_E) \)/probability of inclusion \( (P_I) \)

\( P_I \) represents the combined probability of all combinations of genotypes that cannot be excluded to have contributed to the DNA profile of a stain based on the criteria given in the “Inclusion” section. \( P_I \) is equivalent to the match probability in the case of a stain originating from a single person.

The calculation of \( P_I \) is independent of assumptions about the number of possible contributors to a stain, the genotypes, and the ethnic origin of persons involved in a given case. It is equivalent to the probability that a randomly selected person is a contributor to the stain \( \text{[random man not excluded (RMNE)]} \). The probability of exclusion \( P_E=1-P_I \) indicates the probability of excluding a randomly selected person as a contributor to a given stain.

Likelihood ratio

The calculation of the likelihood ratio (LR) is based on the assumption of two mutually excluding hypotheses. This imperatively requires the description of a distinct scenario for a given stain case. Both hypotheses explicitly describe alternative scenarios for the origin of a stain. Each of these hypotheses shall clearly state who contributed to the stain and how many unknown contributors are assumed. Then, a calculation of the likelihood for the occurrence of the DNA profile of the stain is performed based on the assumption of the respective hypotheses: \( L(\text{stain}|H) \). The LR

\[
LR = \frac{L(\text{stain}|H_1)}{L(\text{stain}|H_2)}
\]

allows the evidential value of a stain to be calculated with reference to a specific person involved in a case, e.g., an accused stain donor.

Given a two-person mixed stain \( M \) and that all observed alleles can be explained by the genotype of the victim, \( G_v \), and the genotype of the suspect, \( G_s \), the hypotheses can be formulated as follows:

- **Hypothesis \( H_p \) (view of the prosecution):** The stain \( M \) originates from the victim \( V \) and the suspect \( S \).
- **Hypothesis \( H_d \) (view of the defense):** The stain originates from the victim \( V \) and from an unknown person \( U \) unrelated to the suspect.

\[
LR = \frac{L(M|H_p)}{L(M|H_d)} = \frac{L(M|G_v,G_s)}{L(M|G_v,G_u)}
\]

The resulting LR provides a numerical value, which indicates how many times more likely the observed DNA profile is under the assumption of the scenario described in \( H_p \) compared to the scenario described in \( H_d \).

Procedures

**Calculation for a mixed stain with an unambiguous major component from one person**

The conclusion of a major DNA profile from a single contributor in a mixed stain shall only be drawn if a peak height ratio of at least 4:1 (major vs minor component) is observed across all heterozygous DNA systems (see “Definitions” section). In this case, the major DNA profile can be considered equivalent to that of a stain originating from a single person, and all calculations can be performed accordingly.
Calculation based on probability of exclusion/inclusion

If a major DNA profile cannot be identified based on unambiguous DNA profiles, or if the number of contributors cannot be determined, calculations of the probability of exclusion \( P_E \) or the probability of inclusion \( P_I \), respectively, for randomly selected persons is appropriate. Also, the calculation of \( P_E \) and \( P_I \) is always possible for type A and type B mixtures.

Supplementary recommendations

Further calculations that may result in erroneous interpretations of the evidence shall not be performed (e.g. reporting the genotype frequency of a non-excluded suspect, if the mixed stain does not allow a meaningful biostatistical interpretation).

Validated computer programmes for the calculation of complex mixed stains are available.

Acknowledgements We thank K. Anslinger (Munich), P. Berschick (Düsseldorf), M. Eckert (Wiesbaden), C. Hohoff (Münster), S. Jung (Würzburg), and J. Schnee-Griese (Stuttgart) for helpful collaboration. Furthermore, we thank N. Morling (Copenhagen) and S. Rand (Antwerp) for a critical revision of the English text.

Appendix

Examples of the calculations of \( P_I \) and \( P_E \)

The probability of inclusion \( P_I \) is calculated from the sum of all genotypes of possible stain contributors. In a stain case, where \( a, b, \) and \( c \) denote the alleles of a DNA system detected in the mixture, the sum of all relevant genotypes can be calculated as follows (assuming that allele frequency data conform to Hardy–Weinberg equilibrium):

\[
P_I = a^2 + b^2 + c^2 + 2ab + 2bc + 2ac
\]

This term can be simplified using the formula for the binominal distribution:

\[
a^2 + b^2 + c^2 + 2ab + 2bc + 2ac = (a + b + c)^2
\]

Assuming a frequency of 0.1 for alleles \( a, b, \) and \( c \), the following result is obtained:

\[
P_I = 0.3^2 = 0.09
\]

Thus, it is expected that 9% of a group of randomly selected persons (=RMNE). The probability of exclusion is calculated from the difference

\[
P_E = 1 - P_I = 1 - 0.09 = 0.91
\]

Thus, it is expected that 91% of a group of randomly selected persons will be excluded as stain contributors. For several DNA systems, \( S_1, S_2, \ldots, S_n \), which are genetically unlinked (i.e., in linkage equilibrium), the general expression of \( P_E(S_1, S_2, \ldots, S_n) \) can be derived from the product of the individual inclusion probabilities \( P(S_j) \) as follows:

\[
P_E(S_1, S_2, \ldots, S_n) = 1 - [P(S_1) \cdot P(S_2) \cdot \ldots \cdot P(S_n)]
\]

Examples for the calculation of the LR

Simple scenario

Consider a case with a mixed stain \( M \) with three alleles, \( a, b, \) and \( c \), composed from a victim and a perpetrator. The victim \( V \) has the genotype \( AB \), and the suspect \( S \) has the genotype \( BC \). The hypotheses can be given as follows:

\( H_p \): The stain \( M \) originates from the victim \( V \) and the suspect \( S \).

\( H_d \): The stain \( M \) originates from the victim \( V \) and from an unknown person unrelated to the suspect.

Let us first derive the numerator of the LR. The prosecution claims that the stain \( M \) can be explained by a combination of the genotypes of the victim and the suspect, as there are no unaccounted alleles. Hence, the numerator results as

\[
L(M|H_p) = L(M|G_V, G_s) = 1
\]

The defense, however, claims that the suspect has not contributed to the stain. The genotype of the suspect is not relevant since the presence of allele \( c \) in the mixture must be explained by the contribution of an unknown person. As allele \( c \) may have been contributed either by a person homozygous for allele \( c \) or from a person heterozygous for \( c \) in combination with allele \( a \) or \( b \), the denominator is as follows:

\[
L(M|H_d) = L(M|G_u, G_s) = 2ac + 2bc + c^2
\]

And, thus, the entire expression is given as

\[
LR = \frac{1}{2ac + 2ab + c^2}
\]

Assuming a frequency of 0.1 for alleles \( a, b, \) and \( c \), the following result is obtained:

\[
LR = \frac{1}{0.02 + 0.02 + 0.01} = \frac{1}{0.05} = 20
\]
The result can be described by the following statement: It is 20 times more likely to observe the DNA profile if the mixed stain originated from the victim and the suspect than if it originated from the victim and an unknown person (who is unrelated to the suspect).

**Complex scenario**

Let us consider a case with a mixed stain \( M \) with four alleles \( a, b, c, \) and \( d \) found on the victim’s clothes. The victim’s genotype is EF and, hence, the corresponding alleles \( e \) and \( f \) are not observed in the stain. Suspect \( S \) has genotype AB, but there is no known second person who may have contributed the alleles \( e \) and \( d \). The hypotheses can be given as follows:

- \( H_p: \) Stain \( M \) originates from suspect \( S \) and an unknown person \( U \).
- \( H_d: \) Stain \( M \) originates from two unknown persons \( U1 \) and \( U2 \).

The prosecution claims that the stain can be explained by a combination of the suspect’s genotype and a second person with the genotype CD. Hence, the numerator results as

\[
L(M|H_p) = L(M|G_S, G_U) = 2cd
\]

The defense claims that no genotypes of the contributors are known. Thus, the sum of all possible genotype combinations from two persons \( U1 \) and \( U2 \) must be considered for the denominator:

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Combined frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U1 )</td>
<td>( U2 )</td>
</tr>
<tr>
<td>AB</td>
<td>CD</td>
</tr>
<tr>
<td>AC</td>
<td>BD</td>
</tr>
<tr>
<td>AD</td>
<td>BC</td>
</tr>
<tr>
<td>BC</td>
<td>AD</td>
</tr>
<tr>
<td>BD</td>
<td>AC</td>
</tr>
<tr>
<td>CD</td>
<td>AB</td>
</tr>
</tbody>
</table>

\[
L(M|H_d) = L(M|G_{U1}, G_{U2}) = 24abcd
\]

After reducing the term and by assuming a frequency of 0.1 for alleles \( a, b, c, \) and \( d \), the following result is obtained:

\[
LR = \frac{2cd}{24abcd} = \frac{1}{12ab} = 0.12 = 8.3
\]

Thus, it is eight times more likely to observe the DNA profile if the mixed stain originated from the suspect and an unknown person than if it originated from two unknown persons. If two suspects \( S1 \) and \( S2 \) with the genotypes AB and CD are considered for the same mixed stain scenario, the hypotheses and, hence, the LR change, as no unknown person remains for \( H_d: \)

- \( H_p: \) Stain \( M \) originates from the suspects \( S1 \) and \( S2 \).
- \( H_d: \) Stain \( M \) originates from two unknown persons \( U1 \) and \( U2 \).

Thus, the numerator of the LR is, again, 1. The term cannot be reduced further and the resulting LR is as follows:

\[
LR = \frac{1}{24abcd} = \frac{1}{0.0024} = 416.7
\]

Thus, it is 416 times more likely to observe the DNA profile if the mixed stain originated from suspects \( S1 \) and \( S2 \) than if it originated from two unknown persons.

We give the following caveat: Additional hypotheses, which are not discussed here, can be formulated. Depending on the precise scenario, such additional hypotheses may be highly relevant in a given case, such as (a) \( H_p: \) the stain originates from \( S1 \) and \( S2 \); (b) \( H_d: \) the stain originates from \( U1 \) and \( U \), or (b) \( H_p: \) the stain originates from \( S1 \) and \( S2 \); (b) \( H_d: \) the stain originates from \( S2 \) and \( U \). Depending on the genotype frequencies of \( S1 \) and \( S2 \), the resulting LR may differ significantly.

### References

Mixture Homework

Data courtesy of Amy Christen, Marshall University NEST Project Team