Capillary Electrophoresis in DNA Analysis

**Data Interpretation**

NEAFS Workshop
Mystic, CT
September 29-30, 2004
Dr. John M. Butler
Dr. Bruce R. McCord

Outline for Workshop

- Introductions
- STR Analysis
  - Introduction to CE and ABI 310
    - Data Interpretation
  - Additional Topics – Real-time PCR and miniSTRs
    - Higher Throughput Approaches
  - Troubleshooting the ABI 310 (Participant Roundtable)
  - Additional Topics – Y-STRs, validation, accuracy
- Review and Test

Steps in STR Typing

- Sample Injection
- Mixture of dye-labeled PCR products from multiple PCR reaction
- Capillary
- Argon ion LASER (488 nm)
- Fluorescence
- Detection
- CCO Panel (virtual filters)
- Color Separation
- Sample Interpretation
- Processing with GeneScan/Genotyper software

NIST Instruments Used for STR Typing

- GeneAmp 9700
- ABI 310
- ABI 3100
- 480 and 9600
- PowerPlex 16 BIO
- FMBIO III Gel Imager System

We work with all of the different commercial STR kits

- Blue
- Green I
- Profiler™
- Profiler Plus™
- COfiler™
- SGM Plus™

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Sample Interpretation Overview

- Data collection on analytical instrument (e.g., ABI 310 or ABI 3100)
  - Peak identification and sizing
  - Correlation of peak sizes to STR allele repeat number through comparison with previously or concurrently run allelic ladder
  - Review of “called” alleles with editing where needed according to laboratory-specific interpretation guidelines
  - Transfer of final reviewed allele calls from STR profile to table of genotypes

Steps in STR Genotyping Process

- Data Collection
- Peak Identification
- Peak Sizing
- Comparison to Allelic Ladder
- Assignment to Alleles
- Data Review by Analyst/Examiner
- Confirmation of Results by Second Analyst/Examiner

Three Possible Outcomes

- Match – Peaks between the compared STR profiles have the same genotypes and no unexplainable differences exist between the samples. Statistical evaluation of the significance of the match is usually reported with the match report.
- Exclusion – The genotype comparison shows profile differences that can only be explained by the two samples originating from different sources.
- Inconclusive – The data does not support a conclusion as to whether the profiles match. This finding might be reported if two analysts remain in disagreement after review and discussion of the data and it is felt that insufficient information exists to support any conclusion.
Crime Scene STR Profile Compared to Two Suspects

<table>
<thead>
<tr>
<th>STR</th>
<th>Suspect 1</th>
<th>Suspect 2</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBS1179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YVA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comparison of Allelic Ladder to Samples to Convert Size into Allele Repeat Number**

Size deviation of 70 samples and two allelic ladders from one injection of allelic ladder on a single ABI PRISM 310 Genetic Analyzer run.

Precision from Run-to-Run on ABI 310

Data to be Reviewed from Electropherograms

**Final Answer Desired**

What is the DBS1179 genotype?

**Important Information from this Sample**

D8S1179 genotype = 11,14

**SWGDAM STR Interpretation Guidelines**

[SBGDAM](http://www.fbi.gov/hq/lab/sc/sciencetools/strig.htm)

The interpretation of results in casework is a matter of professional judgment and expertise. Not every situation can or should be covered by a preset rule. It is important that the laboratory develops and implements written guidelines for interpretation of analytical results. This document provides a framework for the laboratory to develop short tandem repeat (STR) interpretation guidelines. The laboratory's interpretation guidelines should be based upon validation studies, data from the literature, instrumentation used, and/or casework experience.
SWGDAM STR Interpretation Guidelines

1. Preliminary Evaluation of Data

1.1. The laboratory should develop criteria to determine whether the results are of sufficient internal quality for interpretation purposes using methods appropriate for the detection platform. These criteria should be determined by evaluating data generated by the laboratory.

1.1.1. When quantitative results (e.g., peak height ratios) are used to evaluate STR profiles, the results should be examined to determine if they meet the laboratory's defined statistical and interpretational thresholds.

1.1.1.1. The analytical threshold(s) is defined as the minimum and maximum density thresholds that are determined to assign alleles.

1.1.1.2. The interpretation thresholds should be defined empirically.

1.1.2. When qualitative results are not used, the laboratory should establish criteria to interpret alleles based on visual inspection of gel images.

1.2. The laboratory should develop criteria to evaluate internal lane size standards and/or allelic ladders.

1.3. Controls are required to assess analytical procedures.

1.3.1. The laboratory should establish criteria for evaluation of the following controls, including but not limited to: negative lanes, amplification lanes, and positive control.

1.3.2. The laboratory should develop criteria for the interpretation and documentation of results in the event that the controls do not perform as expected.

1.4. A laboratory using STR multiplexes that contain redundant loci should establish criteria regarding the concurrence of such data.


---

3. Interpretation of Results

3.1. The laboratory should evaluate the results of DNA testing by determining whether the DNA is from a single person or more than one person. This may be accomplished by an examination of the number of alleles at each locus, peak height ratio, and/or band intensities.

3.1.1. Single Contributor: A sample may be considered to be from a single contributor when the observed number of alleles at each locus and the signal intensity ratio of alleles at a locus are consistent with a profile from a single contributor. All loci should be evaluated in making this determination.

3.1.2. Mixture With Single/Cosingle Contribution: A sample may be considered to consist of a mixture of major and minor contributors if the minor contributor(s) do not generate a profile that is consistent with a single contributor profile. All loci should be evaluated in making this determination.

3.1.3. Mixture With Known Donors: In some cases, when one of the contributors (e.g., the victim) is known, the genetic profile of the unknown contributor may be inferred. Depending on the profiles in the mixture, an unknown contributor can be distinguished because of similarity in signal intensities or the presence of shared or masked alleles. The genetic profile of the unknown contributor may be inferred as either greater than (+) or less than (−) the respective ladder allele.

3.1.4. Mixture With Unknown or Other Contributors: When major or minor contributors cannot be distinguished because of similarity in signal intensity or the presence of shared or masked alleles, mixtures may still be evaluated or excluded as possible contributors.

3.2. The laboratory should establish guidelines to interpret profiles that exhibit potential stochastic effects (e.g., PCR inhibitors). The laboratory should prepare guidelines for formulating conclusions resulting from comparisons of single source samples and mixtures with known reference samples.


---

5. Statistical Interpretation

5.1. The source of the population database(s) used should be documented. Relevant population(s) for which the frequency will be calculated should be identified.

5.2. The formulas used in calculating the frequency of a DNA profile should be defined for the following:

5.2.1. Heterozygote profiles

5.2.2. Homozygote profiles

5.2.3. Composite profiles (i.e., multiple locus profiles)

5.2.4. Minimum allele frequencies

5.2.5. Mixture calculations

5.2.6. Biological relationships, where appropriate

5.3. When used, criteria for the declaration of source attribution should be documented.


---

6. References/Suggested Readings


DNA Advisory Board: Quality assurance standards for convicted offender DNA databasing laboratories (approved April 1999), Forensic Science Communications (July 2000) 2. Available at www.fbi.gov/programs/lab/fsc/backissu/july2000/codisp.htm


---

NEAFS CE-DNA Workshop (Butler and McCord) Sept 29-30, 2004

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
GeneScan and Genotyper

Review of STR Data

GeneScan

- Apply matrix, size standard, and analysis parameters
- Remove the peak designation from the “250 bp” peak in the GS500 ROX/LIZ internal size standard
- Confirm that all alleles in the allelic ladder have been designated as peaks

Genotyper

- Scan sizes of all peaks in the internal size standard to confirm that they are correct (especially the 340 bp peak)
- Run genotyping macro (Kazaam or Power)
- Review electropherograms and edit data primarily through removing labels on peaks determined not to be alleles
- Create a table of final allele calls for export

GeneScan® Software

- Calls peaks (based on threshold values)
- Separates colors with matrix file
- Sizes peaks with internal size standard

ABI manual is P/N 4303189

GeneScan Software Screen Views

- Raw Data
- Sample Information
- Analyzed Data
- Instrument Performance Indicators (Current, etc.)
- Size Curve

Baseline Subtraction

Peaks are filtered to produce normalized data

Analysis Parameters

Flowchart of Steps Used by GeneScan Software to Produce Analyzed STR Data
GeneScan Analysis Parameters

Macintosh version 3.1

NT version 3.7

GeneScan Analysis Parameter Differences

NT version 3.7 Default

NT version 3.7.1 Default

GeneScan® Analysis version 3.7.1 enables analysis of sample files that, under certain circumstances, might have failed under v3.7. Several minor improvements were implemented to increase the robustness of the sizecaller and improve the baselining function to eliminate negative peak area.

### Differences between Mac and NT versions of 310 Data Collection

J. Forensic Sci. 2004, 49(1):92-95

The NT version of GeneScan captures peak height information prior to smoothing.

### Be Careful when Limiting the Analysis Range

Do not put too close to peaks that are needed.

There will be some variation from run to run in terms of scan numbers.

Start point (e.g., 4000 data points)

Sample runs a little slower

Sample runs a little faster

Peaks Seem to Disappear from Electropherograms if the Analysis Start Position is too Late

Must go back to Genescan software and reduce Analysis start position to capture the 75 bp peak…
Genotyper 3.7 view

Peak appears out of +/-0.5 bp sizing window due to loss of 75 bp peak from the GS500 ROX sizing standard

Appearance of “poor precision” may be due to failure to include a peak in the internal size standard for that sample

Be careful about the stop point for data collection and analysis

There will be some variation from run to run in terms of scan numbers

If a stop point of 25 minutes was used, the 400 bp peak would not be collected or analyzed...

Approx. 1 minute difference in these runs

Data from Debbie Hobson (2001) FBI Laboratory

Internal Size Standards and Sizing Algorithms

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

Internal Sizing Standards

GS500 ROX (Applied Biosystems)

IL600 CXR (Promega)

GeneFlo 625 ROX (CHIMERx-no longer available?)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Local Southern Sizing Algorithm

- Local Southern is commonly used but may not be the best in all situations.
- Local Southern involves using 2 peaks above and 2 peaks below an unknown peak from the internal size standard to make a calculated DNA size.

\[
\text{Size of STR Allele} = \frac{X_1 + X_2}{2}
\]


Global Southern vs Local Southern Methods

- **Global Southern Method:** Generates best-fit curve from all matched fragments in the size standard.
- **Local Southern Method:** Generates best-fit curve from only nearby internal lane standard data points (usually two peaks on either side of the unknown).

Local Southern Sizing

- Normally the 250 bp peak is not included because it does not always migrate uniformly compared to the other peaks in the size standard.

Global Southern Sizing

- **Global Southern Method:** Generates best-fit curve from all matched fragments in the size standard.
- **Local Southern Method:** Generates best-fit curve from only nearby internal lane standard data points (usually two peaks on either side of the unknown).

There is value in Global Southern Sizing for STRs when temperature variations within a run exist:


Thoughts on Size Standards

- Be consistent in use if you want to be able to compare data over time.
- All size standards I have tested work.
- Allele sizes are different with different internal sizing standards.
- GS500 has a large “hole” in its sizing ability when using the local Southern algorithm for medium-sized STR alleles because of the 250 bp peak that cannot be used; also must be run out to 450 bp to be able to type large FGA alleles with ABI kits.


Abstract

Early studies have established the Local Southern algorithm as a precise tool for sizing DNA fragments. As a result, the Local Southern algorithm of the PE Applied Biosystems' software GeneScan® Analysis (PE Applied Biosystems, Foster City, California), is the manufacturer’s recommended method for sizing short tandem repeats (STRs). However, this recommendation is made with the warning that size estimates may be imprecise if any of the standard fragments run anomalously. Specifically, the GeneScan®-500 (GS-500) internal standard fragments of 250 and 340 bases in length ran anomalously under non-optimal conditions on the ABI Prism® 310 Genetic Analyzer (PE Applied Biosystems, Foster City, California).

The California Department of Justice DNA Laboratory currently uses the GS-500 size standard without the 250-base standard assigned and the Local Southern method to size AmpFLSTR Profiler Plus™ alleles. However, even with the manufacturer's recommended instrument running conditions, studies in this laboratory demonstrate that ambient temperature variation over the course of a 310 run can result in anomalous migration of GS-500 standard fragments. When ambient temperature varies, a simple analysis method change can improve precision.

This study suggests that the Global Southern method may provide improved precision over the Local Southern method when using the GS-500 internal standard with the ABI Prism® 310 Genetic Analyzer. In addition, this study shows that precision for fragments greater than 300 bases is further improved by excluding the 340-base GS-500 fragment in conjunction with using the Global Southern method. When ambient temperature shifts occur, this sizing method change should reduce the number of sample reruns necessary.


http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
**Summary of Steps in GeneScan Data Analysis**

Create New Project (Load sample files)
Establish Matrix
Define Size Standard Peaks
Set Analysis Parameters
- Minimum peak height (e.g., 50 RFU)
- Scan range evaluated (best to avoid primer region)
- Sizing algorithm options (Local Southern is default)

Data Review Windows
- Analysis Control Window
- Results Control Window

**Steps for Processing Samples:**
- Install matrix onto each (highlighted) sample
- Select appropriate size standard (or create a new one)
- Check analysis parameter values (or select pre-defined analysis parameter file)
- Click “Analyze” button
- Examine data in Analysis Control by double clicking on sample file name (this allows you to pull up files in all colors and select or de-select color of interest)

**Effect of Peak Detection Threshold**

Failure to have these alleles designated will result in Genotyper macro failure and inability to type samples

Peak threshold too high so that some alleles are not designated as peaks by the software

**Genotyper Software**

<table>
<thead>
<tr>
<th>Macintosh</th>
<th>Windows</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>NT</td>
</tr>
<tr>
<td>2.5</td>
<td>3.7 (5-dye)</td>
</tr>
<tr>
<td>2.5.2 (5-dye)</td>
<td></td>
</tr>
</tbody>
</table>

- Converts GeneScan sized peaks into genotype calls using macros
- Genotyping performed by comparison of allele sizes in allelic ladder to sample alleles

ABI manual is P/N 904648

**Summary of Genotyper Data Analysis**

Open template file containing typing macro specific for STR kit used
Import PROCESSED/REVIEWED GeneScan files
Check internal size standard peaks
Run Kazaam (or Power) macro
Review allelic ladder to confirm that all alleles are called correctly
Review sample data by dye color
Remove calls to stutter peaks or instrument artifacts
Create allele table
Export allele table

**Rapid Genotyper Data Review**

An overlay of electropherograms permits a rapid assessment of sizing precision and the number of alleles seen in a particular sample set…

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
**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)

**Poor primer purity**

**Removal of Dye Artifacts Following PCR Amplification**

**Common Errors in Genotyping**
- Genotyper table does not import the third allele in a tri-allelic pattern
- Bleedthrough (pull-up) between dye colors results in a peak that falls into a possible allele bin in an adjacent color
- Clicking off a peak label for a true allele by accident and failing to restore the label
- Accidentally clicking on a peak and inserting a label near the beginning of a locus size range that is not a true allele, which causes the second true allele to not show up in the final table of results since only two alleles are imported for each locus

---

**GeneMapperID v 3.1 Software**

Released in Nov 2003

Since June 2004, Applied Biosystems no longer sells GeneScan and Genotyper software (they will support these programs until June 2009)...ABI is encouraging the adoption of GeneMapperID which replaces the functions of GeneScan and Genotyper.
310 Data Collection Software and GeneMapperID

- Controls 310 run conditions
- Translates light on CCD camera into electropherogram (raw data)
- Sample sheets and injection lists are created

ABI manual is P/N 904958B

Use of GeneMapperID v3.1

- Projects
- Kit, panel, marker, bin
- Provides a flexible format and different methods for viewing data
- Can analyze results from multiple STR kits within the same GeneMapper project file
- Not designed to work with pentanucleotide repeat loci (version 3.2 will be able)

The Project Window

Analysis Parameters Window

Algorithms:
- Basic is the default for many applications
- Advanced is similar to the Windows NT version of GeneScan
- Classic is similar to Macintosh version of GeneScan

Sample Plot Window

Genetyper-like features
Review Multiple Loci from a Single Sample

Review Multiple Samples at a Single Locus

Panel Manager Window

Process Component Based Quality Values (PQVs)

A kit is a collection of panels.
A panel is a collection of markers.
Bin sets are collections of the expected allele locations for markers contained with a kit.

Expert Systems Under Development

• Difference between an “expert system” and a “quality assurance” tool

Examples
• TrueAllele (Cybergenetics, Pittsburg, PA)
• STRess2 (Forensic Science Service, UK)
• OSIRIS (NCBI/NIJ, Steve Sherry’s team)
• CompareCalls (Myriad Genetics, Salt Lake City, UT)

Model for Future Use of Expert Systems

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
TrueAllele® is designed to operate independently of other allele calling systems.

A single reviewer with TrueAllele may be used in place of the current two-person review process.

Examined a dataset of 2,048 convicted offender STR profiles and found only four samples with differences between TrueAllele and Genotyper at a single locus each (spikes, calling alleles outside of ladder range, and Genotyper missing the third allele in a tri-allelic pattern).

**Steps for Putting STR Data into PowerPoint Presentations**

- Open GeneScan or Genotyper file and put data into desired display format
  - It is better to keep image a medium size that can be stretched within PowerPoint to make the lines thicker
- Take a picture of the desired image on screen
  - Windows: press "print screen", paste image into Paint program, then crop portion of image desired
  - Mac: shift-Apple-4 keys, draw box around desired image, open picture image under Macintosh hard drive
- Open PowerPoint and paste copied image from Paint
- Label image within PowerPoint
- Print as 2-per-page handout to have a nice size annotated figure

**Example with Profiler Plus Allelic Ladder**

- Label Loci...
- Image is too narrow for a nice balanced plot
- Stretch the image
- Results in thicker lines and more easily read numbers

**Steps in Creating PowerPoint Files from GeneScan and Genotyper Data**

**Collecting Information**

Presenting Information Collected on the ABI 310

**Journal of Forensic Sciences** (July 2004), volume 49(2), pp. 660-667
The same thing is done in Genotyper—display image in software, print screen, crop within Paint and paste into PowerPoint, stretch to fill slide, add labels as needed...

### Databank vs Casework Data Challenges

- **Databank** (single source samples)
  - Too much DNA may be added to the PCR reaction resulting in pull-up between dye colors
  - Lots of data to review—often produced by contractors

- **Casework** (mixtures or low level samples)
  - Often limited DNA material to work with
  - Low copy number samples can result in allele dropout
  - Can produce complicated STR profiles to interpret

### Common Casework Challenges

- **Degraded DNA**
  - Loss of signal at larger size loci

- **Mixture**
  - More than two alleles at multiple loci

### Forensic Casework DNA Profiles

**Crime scene evidence**

(MIXTURE of victim and perpetrator)

All of the suspect’s alleles are included in the evidentiary DNA profile


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