Concordance Testing Comparing STR Multiplex Kits with a Standard Data Set

Becky Hill, Margaret Kline, David Duewer, Peter Vallone, and John Butler

National Institute of Standards and Technology

NIST – GMI Seminar
Innsbruck, Austria
September 5, 2011
Outline of Topics to Discuss

• Introduction and importance of concordance testing
  • Overlapping markers with different primer configurations

• NIST role in concordance testing
  – SRM 2391b/2391c concordance with new kits
  – Standard sample set, DNA sequencing

• Commercial STR multiplex kits examined
  – Applied Biosystems, Promega, and Qiagen

• Concordance results with various STR multiplex kits
  – Primer binding site mutations and null alleles

• Summary and conclusions
Why are concordance studies important?
Importance of Concordance Testing

• There are a variety of commercial STR multiplex kits with different configurations of STR markers
  – Different primer sequences are used to amplify the same markers
  – Discordant results can impact DNA databases

• Detection of primer binding site mutations that cause null alleles, or allele drop-out
  – Can only be determined with concordance testing and DNA sequencing

• Concordance with NIST reference materials
  – Important to test with all new STR typing kits

Purpose of Concordance Studies

When different primer sets are utilized, there is a concern that allele dropout may occur due to primer binding site mutations that impact one set of primers but not another.

* represents potential mutations impacting primer annealing

Use of non-overlapping primers permits detection of allele dropout

If no primer binding site mutations

Set 1 Amplicons = Set 2 Amplicons

If a primer binding site mutation exists

Set 1 Amplicons ≠ Set 2 Amplicons
Example Primer Binding Site Mutation that Causes a Null Allele

**D19S433 repeat region**

ggaaggaagtagaggaaggaaggaaggaagtaagtaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag

*This region could potentially represent where the reverse primer is located to include the primer binding site mutation*

*Applied Biosystems does not publish their primer sequences*
To Avoid Overlapping PCR Product Size Ranges with STR Loci in the Same Dye Channel

• **Applied Biosystems (Strategy 1)**
  – **Maintains primer sequences** (except MiniFiler & NGM kits)
  – Utilizes mobility modifiers or additional dyes, no primer redesign is necessary
  – Enables comparison to legacy data with earlier kits but null alleles may go undetected with the potential for incorrect genotypes within data sets

• **Promega Corporation (Strategy 2)**
  – Moves primer sequences to change PCR product size ranges
  – Primer redesign can be difficult, but can be moved from primer-binding-site mutations
  – **Requires concordance studies to check for potential allele dropout**
Why is NIST involved in concordance studies?
Purpose of Concordance Studies

1. To test SRM 2391b/2391c (PCR-based DNA Profiling Standard) components with all new STR multiplex kits and verify results against certified reference values.

2. To gain a better understanding of primer binding site mutations that cause null alleles.
What are the NIST strategies for concordance testing?
Strategies for Concordance Testing

Carolyn R. Hill, Margaret C. Kline, David L. Duewer and John M. Butler
National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, Maryland, USA

Concordance evaluations are important to conduct to determine if there are any allelic dropout or "null alleles" present in a data set. These studies are performed because there are a variety of commercial short tandem repeat (STR) multiplex kits with different configurations of STR markers available to the forensic community. The placement of the markers can vary between kits because the primer sequences were designed to amplify different polymerase chain reaction (PCR) product sizes. When multiple primer sets are used, there is concern that allele dropout may occur due to primer-binding-site mutations that affect one set of primers but not another.

http://www.promega.com/profiles/1301/1301_08.html
The 4 “S’s” of Concordance

• **NIST Standard Samples**
  – Run same samples with multiple kits to compare results

• **Concordance Software**
  – Allows comparison of data sets using NIST developed software
    
    http://www.cstl.nist.gov/biotech/strbase/software.htm

• **DNA Sequencing**
  – To validate and determine the exact cause for the null allele

• **STRBase website**
  – To report verified null alleles and discordant results to the forensic community
    
    http://www.cstl.nist.gov/biotech/strbase/NullAlleles.htm
NIST Concordance Testing Steps

Kit 1

Sample Set

Results

Comparisons with software

Discordant Results

Sequence differences

Vendor changes final kit to correct for primer binding site mutation

Kit 2

Sample Set

Results

Vendor decides not to change the final configuration of the kit

PP-ESX17 is an example

MiniFiler is an example

Sample Set

Report to STRBase
What concordance studies have been completed thus far?
Applied Biosystems AmpFISTR Kits

- Identifiler
- MiniFiler
- Profiler Plus
- SGM Plus
- NGM
- NGM SElect (studies are ongoing)

Promega PowerPlex Systems

- PowerPlex 16
- PowerPlex ESX 17
- PowerPlex ESI 17
- PowerPlex 18D (rapid and direct kit)
Qiagen Investigator HID Kits

- ESSplex
- IDplex
- Hexaplex ESS
- ESSplex SE
What samples are used at NIST to perform concordance testing?
NIST Sample Set (>1450 Samples)

- **NIST U.S. population samples**
  - 254 African American, 261 Caucasian, 139 Hispanic, 3 Asian

- **U.S. father/son paired samples**
  - 178 African American, 198 Caucasian, 190 Hispanic, 198 Asian

- **NIST SRM 2391b**, PCR-based DNA Profiling Standard (highly characterized)
  - 10 genomic DNA samples, 2 cell line samples
  - Includes 9947A and 9948

- **NIST SRM 2391c**, PCR-based DNA Profiling Standard
  - 4 genomic DNA (one mixture)
  - 2 cell lines (903 and FTA paper)
What are the results from the completed concordance studies?
D18S51 Concordance Checking

12 different kits tested

- **ABI**
  - Identifiler
  - NGM
  - NGM SElect
  - Profiler Plus
  - SGM Plus

- **Promega**
  - MiniFiler
  - PP16
  - PP ESX 17
  - PP ESI 17

- **Qiagen**
  - ESSplex
  - ESSplex SE
  - IDplex

- Qiagen

12 different kits tested
<table>
<thead>
<tr>
<th>Marker</th>
<th># of Sets</th>
<th>Marker</th>
<th># of Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amelogenin</td>
<td>13</td>
<td>D2S441</td>
<td>9</td>
</tr>
<tr>
<td>D18S51</td>
<td>12</td>
<td>D19S433</td>
<td>9</td>
</tr>
<tr>
<td>D21S11</td>
<td>12</td>
<td>D1S1656</td>
<td>7</td>
</tr>
<tr>
<td>FGA</td>
<td>12</td>
<td>D12S391</td>
<td>7</td>
</tr>
<tr>
<td>D3S1358</td>
<td>11</td>
<td>SE33</td>
<td>5</td>
</tr>
<tr>
<td>TH01</td>
<td>11</td>
<td>D5S818</td>
<td>4</td>
</tr>
<tr>
<td>D16S539</td>
<td>11</td>
<td>D7S820</td>
<td>4</td>
</tr>
<tr>
<td>vWA</td>
<td>11</td>
<td>D13S317</td>
<td>4</td>
</tr>
<tr>
<td>D8S1179</td>
<td>11</td>
<td>TPOX</td>
<td>3</td>
</tr>
<tr>
<td>D2S1338</td>
<td>10</td>
<td>CSF1PO</td>
<td>4</td>
</tr>
<tr>
<td>D10S1248</td>
<td>9</td>
<td>Penta D</td>
<td>1</td>
</tr>
<tr>
<td>D22S1045</td>
<td>9</td>
<td>Penta E</td>
<td>1</td>
</tr>
</tbody>
</table>
D22S1045 Discordance

ESX 17 (prototype) = 17,17

ESI 17 (prototype) = 15,17

Destabilized some (but uses lower annealing temperature with fewer amplicons in multiplex)

ESX 17 (final) = 15,17

NIST NC01 = 15,17

NIST 23plex = 15,17

G→T 15 bp upstream impacting forward primer binding with ESX17

Promega added additional primer to correct issue
**D22S1045 Null Allele**

Correct type (15,17)

NGM SElect

ABI added an additional primer to correct issue

G→T 15 bp upstream impacting forward primer binding with NGM

4/1449 samples
Amelogenin X Null Allele

3/1449 samples

NGM

NGM SElect

ESX-17

ESI-17
D2S441 Null Allele

9/1449 samples
Correct type (9.1,11)

8/9 null alleles were from Asian samples
D2S441 Sequencing

“A” base insertion = 9.1

G → A SNP 25/26 bp upstream of the repeat

True Genotype = 9.1,11
NGM Genotype = 11,11
Primer Changes with ABI Kits

D2S441 and D22S1045 have an additional primer in NGM and NGM SElect

Table 4 from “Development of the AmpFlSTR NGM SElect Kit: New Sequence Discoveries and Implications for Genotype Concordance”, Forensic News (January 2011)
D19S433 Discordance

Identifiler & NGM = 14,14

Allele 13 was missing in two different Asian samples with ABI primers = 2/2886 = 0.07% discordance

ESX 17 = 13,14

ESI 17 = 13,14

AF45A (Asian)

Frequencies [for] the silent allele were determined to be 0.0114 in 176 people from Shizuoka (Honshu) and 0.0128 in 156 people from Okinawa

T→A 8 bp downstream impacting reverse primer binding with Identifiler (and thus SGM Plus)

Natsuko Mizuno,1 D.V.M.; Tetsushi Kitayama,1 M.Sc.; Koji Fujii,1 Ph.D.; Hiroaki Nakahara,1 D.V.M.; Kanako Yoshida,1 Ph.D.; Kazumasa Sekiguchi,1 Ph.D.; Naoto Yonezawa,2 Ph.D.; Minoru Nakano,2 Ph.D.; and Kentaro Kasai,1 Ph.D.

A D19S433 Primer Binding Site Mutation and the Frequency in Japanese of the Silent Allele It Causes

DOI: 10.1111/j.1556-4029.2006.00806.x
Available online at: www.blackwell-synergy.com
D18S51 Null Allele

Correct type (13, 15)

C→T SNP 172 bp downstream from repeat
D3S1358 Null Allele

Correct type (14,17)

G → C SNP 11 bp downstream from repeat

1/1449 samples
Completed Concordance Studies

<table>
<thead>
<tr>
<th>Kits compared</th>
<th>Samples</th>
<th>Loci Compared</th>
<th>Comparisons</th>
<th># Differences</th>
<th>Concordance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>102,345</td>
<td>1,021</td>
<td>948,301</td>
<td>1,109</td>
<td>99.883</td>
</tr>
</tbody>
</table>

948,301 allele comparisons
1,109 total differences
99.88% concordance

Kits (except Identifiler) were kindly provided by Promega, Qiagen and Applied Biosystems for concordance testing performed at NIST.
Was there complete concordance with SRM 2391b and SRM 2391c?
SRM 2391b/2391c
PCR-Based Profiling Standard

- The first set of samples run with new STR multiplex kits is SRM 2391b/SRM 2391c

- All new kits tested have been completely concordant with the certified values of all markers for each component for SRM 2391b and 2391c

- One exception for SRM 2391b: MiniFiler
  - Genomic 8 with D16S539
SRM 2391b Genomic 8 with D16S539

All allele calls with MiniFiler for CSF1PO, D7S820, D13S317, D18S51, D21S11, FGA, and D16S539 (with the exception noted below) match previously certified values.

*Due to primer binding site mutation

Null Allele

Slight imbalance with allele 11
D16S539 SRM 2391b Genomic 8

T→C mutation 34 bp downstream of the repeat

Position of the T→C probably affects the reverse primer of Minifiler and is the 3rd base found the 5’end of the Reverse PP16 primer. This could explain the imbalance of the allele seen when using PP16.
Summary & Final Thoughts
Conclusions

• Concordance testing is valuable when different sets of primers are used to amplify the same markers

• Null alleles and discordant results are reported on STRBase:
  http://www.cstl.nist.gov/biotech/strbase/NullAlleles.htm

• NIST plays an important role in concordance testing to aid the community
  – SRM 2391b/2391c concordance
  – Several null alleles have been fixed before the final release of new STR multiplex kits
Acknowledgments

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

**NIST Disclaimer:** 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.

**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.

**NIST Team for This Work**

John Butler  
Dave Duewer  
Margaret Kline  
Pete Vallone  

A special thanks to Applied Biosystems, Promega, and Qiagen for providing the kits used in this study

**Contact Info:** becky.hill@nist.gov, 301-975-4275