NIST Update

John M. Butler

NIST Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland
NIST Human Identity Project Teams within the Applied Genetics Group

**Forensic DNA Team**

Margarret Kline  
Becky Hill  
Kristen Lewis

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

**DNA Biometrics Team**

Pete Vallone  
Erica Butts  
Kevin Kiesler

*Funding from the FBI S&T Branch through NIST Information Access Division*

**Data Analysis Support**

Dave Duewer

**Office Manager**

Patti Rohmiller

STRBase, Workshops & Textbooks  
Concordance & LT-DNA  
Mixtures, mtDNA & Y

SRM work, variant alleles & Cell Line ID

Rapid PCR, Direct PCR & Biometrics  
ABI 3500 & DNA Extraction  
PLEX-ID & NGS Exploration

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

Forensic STR Information
- STRs101: Brief Introduction to STRs
- Core Loci: FBI CODIS Core STR Loci and European Core Loci
- STR Fact Sheets (observed alleles and PCR product sizes)
- Multiplex STR kits
- Sequence Information (annotated)
- Variant Allele Reports
- Tri-Allelic Patterns
- Mutation Rates for Common Loci
- Published PCR primers
- Y-chromosome STRs
- Low-template DNA Information
- Mixture Interpretation
- Kinship Analysis
- miniSTRs (short amplicons)
- Null Alleles - discordance observed between STR kits
- STR Reference List - now 3400 references

Cataloged as of Dec 2011
605 variant alleles
305 tri-allelic patterns

We invite labs to supply information on variant and tri-alleles observed
Forensic DNA Typing Textbook
3rd Edition is Three Volumes
Now part of my job at NIST (no royalties are received)

For beginning students, general public, & lawyers

Currently being written

Sept 2009
~500 pages

August 2011
~700 pages

Fall 2012
~500 pages
Current NIST Projects
Short Overviews...

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
NIST SRM 2391c

Main Points:
• Traceable physical reference materials to ensure accurate and comparable measurements between laboratories
• Helps meet ISO 17025 needs for traceability to a national metrology institute
• http://www.nist.gov/srm
• SRM 2391c released Aug 2011

Presentations/Publications:
• Profiles in DNA article (Sept 2011)
• ISFG 2011 and ISHI 2011 posters
NIST SRM 2391c

Produced with an entirely new set of genomic DNA samples.

9947A & 9948 are NOT included.

https://www-s.nist.gov/srmors/view_detail.cfm?srn=2391C
**Description of Components in SRM 2391c**

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Quantity (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50 μL of anonymous <strong>female</strong> genomic DNA</td>
<td>1.4 – 1.9 ng DNA/μL</td>
</tr>
<tr>
<td>B</td>
<td>50 μL of anonymous <strong>male</strong> genomic DNA</td>
<td>1.3 – 1.5 ng DNA/μL</td>
</tr>
<tr>
<td>C</td>
<td>50 μL of anonymous <strong>male</strong> genomic DNA</td>
<td>1.3 – 2.0 ng DNA/μL</td>
</tr>
<tr>
<td>D</td>
<td>50 μL of <strong>mixed-source</strong> (Components A and C)</td>
<td>1.4 – 2.0 ng DNA/μL</td>
</tr>
<tr>
<td>E</td>
<td>Two 6 mm punches of CRL-1486 cells spotted on 903 paper</td>
<td>~75,000 cells per punch</td>
</tr>
<tr>
<td>F</td>
<td>Two 6 mm punches of HTB-157 cells spotted on FTA paper</td>
<td>~75,000 cells per punch</td>
</tr>
</tbody>
</table>

\(a\) DNA concentrations and cell counts are nominal values and are **not** intended for use as quantitative standards.
<table>
<thead>
<tr>
<th>Kit Provider</th>
<th>Primer Mixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifiler</td>
<td>Powerplex 16</td>
</tr>
<tr>
<td>Identifiler Plus</td>
<td>Powerplex 16 HS</td>
</tr>
<tr>
<td>NGM</td>
<td>Powerplex ESX 17</td>
</tr>
<tr>
<td>NGM SESelect</td>
<td>Powerplex ESI 17</td>
</tr>
<tr>
<td>COfiler</td>
<td>Powerplex ES</td>
</tr>
<tr>
<td>Profiler</td>
<td>Powerplex S5</td>
</tr>
<tr>
<td>Profiler Plus</td>
<td><strong>Powerplex Y</strong></td>
</tr>
<tr>
<td>Profiler Plus ID</td>
<td>FFFL</td>
</tr>
<tr>
<td>SGM Plus</td>
<td></td>
</tr>
<tr>
<td>SEfiler</td>
<td></td>
</tr>
<tr>
<td>MiniFiler</td>
<td></td>
</tr>
<tr>
<td>Yfiler</td>
<td></td>
</tr>
</tbody>
</table>

All results are concordant across all kits.

In total there is data for 51 autosomal STRs and 17 Y-STRs.
Insertion/Deletion (InDel) Markers

Main Points:

• InDels (insertion-deletion) or DIPs (deletion-insertion polymorphisms) are short length polymorphisms, consisting of the presence or absence of a short (typically 1-50 bp) sequence.

• Like SNPs, InDels have low mutation rate (value to kinship analysis), small amplicon target sizes (value with degraded DNA), and can be highly multiplexed.

• Can be analyzed on CE instruments like STRs.

• Studied commercial 30plex (Qiagen DIPlex) and a home-brew 38plex in U.S. population samples.

Presentations/Publications:

• FSI Genetics Suppl. Series 2011 article.

• ISFG 2011 poster and ISHI 2011 presentation.
Same DNA Sample Tested with Five STR Kits

Identifiler (Applied Biosystems)

NGM SElect (Applied Biosystems)

PowerPlex 16 (Promega)

PowerPlex ESX 17 (Promega)

ESSplex (Qiagen)
## Kit Concordance Comparisons

<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>SGM-ID</td>
<td>1436</td>
<td>11</td>
<td>15,796</td>
<td>1</td>
<td>99.994</td>
</tr>
<tr>
<td>ID-ProPlus</td>
<td>1427</td>
<td>10</td>
<td>14,270</td>
<td>1</td>
<td>99.993</td>
</tr>
<tr>
<td>ID-IDplex</td>
<td>669</td>
<td>16</td>
<td>10,704</td>
<td>19</td>
<td>99.822</td>
</tr>
<tr>
<td>ID-PP16</td>
<td>662</td>
<td>14</td>
<td>9,268</td>
<td>4</td>
<td>99.957</td>
</tr>
<tr>
<td>ID-MiniFiler</td>
<td>1308</td>
<td>9</td>
<td>11,772</td>
<td>27</td>
<td>99.771</td>
</tr>
<tr>
<td>SGM-NGM</td>
<td>1436</td>
<td>11</td>
<td>15,796</td>
<td>4</td>
<td>99.975</td>
</tr>
<tr>
<td>ID-NGM</td>
<td>1449</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProPlus-NGM</td>
<td>1427</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGM-ESI</td>
<td>1436</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProPlus-ESX</td>
<td>1427</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESI-ESX</td>
<td>1455</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESI-ESSplex</td>
<td>1445</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESX-ESSplex</td>
<td>1445</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESI-NGMSElect</td>
<td>715</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

128 kit-to-kit comparisons
1,104,031 allele comparisons
1224 differences observed
~99.9% concordance

(many corrected now)

Kits (except Identifiler) were kindly provided by **Applied Biosystems, Promega, and Qiagen** for concordance testing performed at NIST
Recent Training Workshops

- **AAFS** (February 22, 2011)
  - Mixture Interpretation (with 6 other speakers)

- **ISFG** (August 30, 2011)
  - CE Fundamentals and Troubleshooting

- **Int. Symp. Human Ident.** (October 3, 2011)
  - Mixture Interpretation (with Boston University)

- **Int. Symp. Human Ident.** (October 6, 2011)
  - Troubleshooting Laboratory Systems

Slide handouts available at http://www.cstl.nist.gov/strbase/training.htm
TrueAllele Mixture Software Evaluation

Main Points:

• Exploring the capabilities and limitations of a probabilistic genotyping approach

• Studying TrueAllele software with a number of different types of mixtures (including low-level and 3-4 person mixtures)

• Work being performed at NIST independently of Cybergenetics

Presentations/Publications:

• ISFG 2011 presentation

• ISHI 2011 mixture workshop

D19S433 result from one replicate of 50,000 simulations

3 person mixture conditioning on the victim

Genotype Probability

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,14</td>
<td>94.8%</td>
</tr>
<tr>
<td>13,14.2</td>
<td>1.0%</td>
</tr>
<tr>
<td>13,16.2</td>
<td>1.7%</td>
</tr>
<tr>
<td>14,14</td>
<td>2.4%</td>
</tr>
<tr>
<td>14,16.2</td>
<td></td>
</tr>
</tbody>
</table>

Rapid PCR and Rapid DNA Testing

Main Points:

• **Performing research on reducing the total time required for STR typing**
  – Focusing on the multiplex amplification of commercial STR kits with faster polymerases and thermal cyclers
  – Single-source reference samples (sensitivity > 200 pg)

• **Designing testing plans for rapid DNA typing devices**
  – NIST will be examining rapid DNA instruments with FBI collaboration

• **Exploring direct PCR protocols** with FTA and 903 papers

Presentations/Publications:

• Vallone et al. (2008) FSI Genetics - on rapid PCR
• ISFG 2011 and ISHI 2011 presentations by Tom Callaghan (FBI)
• ISFG 2011 presentation and poster on direct PCR
ABI 3500 Validation Studies

Main Points:
• The 3500 has proven to be reliable, reproducible and robust in our hands – we have provided feedback to ABI to improve use.
• Produces excellent DNA sequencing results.
• Signal strength is different compared to ABI 3130xl and requires studies to set analytical and stochastic thresholds.
• **Dye-specific analytical thresholds** resulted in less allelic and full locus dropout than applying one analytical threshold to all dyes.
• RFID tracking decreases flexibility in our research experience.

Presentations/Publications:
• MAAFS talk (May 2011)
• ABI road show talks (July & Aug 2011)
• ISFG presentation (Sept 2011)
• ISHI poster (Oct 2011)
**ABI 3500 Open Letter Update**

**Concerns Expressed in 3/31/11 Open Letter**

1. RFID tags
2. New .hid file structure requires new software
3. Short shelf life of reagents – would like to see data for expiration times

At the Promega ISHI meeting (Oct 2011), ABI described data for studies around reagent expiration through a poster at their booth. Sailus, Wheaton, Fisher, Calandro. “Understanding the Consumables on the 3500 Genetic Analyzers in the context of a Human Identification (HID) Laboratory”

They have promised that **polymer and buffer expiration dates will no longer be a hard stop** but only a warning with the future Windows 7 software upgrade (3500 Data Collection v1.3).
Performance Assessment of PlexID

Abbott Ibis Biosciences
PLEX-ID System

- In collaboration with FBI
- Evaluating ESI-TOF mass spectrometer for mtDNA
- Base composition of the control region determined from 8 triplex PCRs
- Started running the PlexID platform mid-October 2011
- Scheduled to complete experiments in February 2012
Contamination Check

Checks run weekly on the PlexID to monitor baseline noise and potential contamination

No signal detected in ‘red’ wells
PLEX-ID Evaluations Performed Thus Far…

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of Plates</th>
<th>Number of Unique Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixtures</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Concordance</td>
<td>33</td>
<td>247</td>
</tr>
<tr>
<td>Sensitivity / Limit of Detection</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Contamination</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td><strong>71</strong></td>
<td><strong>6816 wells examined</strong></td>
</tr>
</tbody>
</table>

- Mixtures can be detected with minor component present at 5-10%
- Concordance with Sanger sequencing (98.8%) (n=247)
- Limit of detection ≈ 2.5 pg/well
- 1-2 plates run daily on the platform since mid-October
Future Projects Planned

- New book in progress on interpretation issues
- Additional mixture software evaluation
- Rapidly mutating Y-STR loci (European collaboration)
- More concordance testing with new STR kits
- Complete PLEX-ID mass spec validation with mtDNA base composition (FBI collaboration)
- Rapid DNA test device evaluation (FBI collaboration)
- Exploration of Next-Generation Sequencing
- Digital PCR for human DNA quantitation
Characterizing New STR Loci

Main Points:

• In April 2011, the FBI announced plans to expand the core loci for the U.S. beyond the current 13 CODIS STRs

• Our group is collecting U.S. population data on new loci and characterizing them to aid understanding of various marker combinations

• We are collecting all available information from the literature on the 24 commonly used autosomal STR loci

Presentations/Publications:

• AAFS 2011 presentation
• Hares (2012) Expanding the U.S. core loci… *FSI Genetics* 6(1): e52-e54
Article in the January 2012 issue of *Forensic Science Review*

Available at [http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm](http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm)

**Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis**

**REFERENCE:** Butler JM, Hill CR: Biology and genetics of new autosomal STR loci useful for forensic DNA analysis; *Forensic Sci Rev* 24:15; 2012.

**ABSTRACT:** Short tandem repeats (STRs) are regions of tandemly repeated DNA segments found throughout the human genome that vary in length (through insertion, deletion, or mutation) with a core repeated DNA sequence. Forensic laboratories commonly use tetranucleotide repeats, containing a four base pair (4-bp) repeat structure such as GATA. In 1997, the Federal Bureau of Investigation (FBI) Laboratory selected 13 STR loci that form the backbone of the U.S. national DNA database. Building on the European expansion in 2009, the FBI announced plans in April 2011 to expand the U.S. core loci to as many as 20 STRs to enable more global DNA data sharing. Commercial STR kits enable consistency in marker use and allele nomenclature between laboratories and help improve quality control. The STRBase website, maintained by the U.S. National Institute of Standards and Technology (NIST), contains helpful information on STR markers used in human identity testing.

**Key Words:** Autosomal genetic markers, CODIS STRs, core loci, DNA typing, European Standard Set, expanded U.S. core loci, short tandem repeat (STR), STR kits.

Discusses the 24 autosomal STR loci available in commercial kits.
What concerns have been raised?

Recent Public Criticism of Efforts to Expand the CODIS Core Loci in the U.S.

**October 4, 2011** Presentation at Promega’s International Symposium on Human Identification

**October 16, 2011** Follow-up article by BBC News on ISHI presentation

**FBI’s DNA Database Upgrade Plans Come Under Fire**

By: Paul Rincon, Science editor Published October 16, 2011 by the BBC News Website  A major upgrade of the Federal Bureau of Investigation’s (FBI) DNA database system has come under fire from members of the forensic science community. The Codis system is used to generate the genetic profiles stored in the US national DNA [...]
Recent Publication by Budowle et al. Summarizing Criticisms Raised in ISHI Talk

Jianye Ge, Arthur Eisenberg, Bruce Budowle
*Investigative Genetics* 2012, 3:1 (6 January 2012)
Developing criteria and data to determine best options for expanding the core CODIS loci

Available at [http://www.investigativegenetics.com](http://www.investigativegenetics.com)
What concerns have been raised?

**Concerns Raised** in Public Criticisms of Expanded CODIS Core Loci Selection

- **Not enough data behind decisions** – need more community involvement rather than a small committee making decisions
- **Casework needs** should drive decisions
- **Large loci fail in casework samples and should be avoided** – miniSTR capabilities are preferred
- **Large multiplexes** may adversely impact performance
- **DYS391 is a poor choice** and AMEL Y nulls are not a significant concern
- **Y-STRs should be included** as core loci to benefit familial searches of the future
- **No definition of performance goals** are provided
What data exist behind decisions made so far and what additional data are there for consideration to help address concerns raised?
### The 11 STR Loci Beyond the CODIS 13

<table>
<thead>
<tr>
<th>STR Locus</th>
<th>Location</th>
<th>Repeat Motif</th>
<th>Allele Range*</th>
<th># Alleles*</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2S1338</td>
<td>2q35</td>
<td>TGCC/TTCC</td>
<td>10 to 31</td>
<td>40</td>
</tr>
<tr>
<td>D19S433</td>
<td>19q12</td>
<td>AAGG/TAGG</td>
<td>5.2 to 20</td>
<td>36</td>
</tr>
<tr>
<td>Penta D</td>
<td>21q22.3</td>
<td>AAAGA</td>
<td>1.1 to 19</td>
<td>50</td>
</tr>
<tr>
<td>Penta E</td>
<td>15q26.2</td>
<td>AAAGA</td>
<td>5 to 32</td>
<td>53</td>
</tr>
<tr>
<td>D1S1656</td>
<td>1q42</td>
<td>TAGA</td>
<td>8 to 20.3</td>
<td>25</td>
</tr>
<tr>
<td>D12S391</td>
<td>12p13.2</td>
<td>AGAT/AGAC</td>
<td>13 to 27.2</td>
<td>52</td>
</tr>
<tr>
<td>D2S441</td>
<td>2p14</td>
<td>TCTA/TCAA</td>
<td>8 to 17</td>
<td>22</td>
</tr>
<tr>
<td>D10S1248</td>
<td>10q26.3</td>
<td>GGAA</td>
<td>7 to 19</td>
<td>13</td>
</tr>
<tr>
<td>D22S1045</td>
<td>22q12.3</td>
<td>ATT</td>
<td>7 to 20</td>
<td>14</td>
</tr>
<tr>
<td>SE33</td>
<td>6q14</td>
<td>AAAG‡</td>
<td>3 to 49</td>
<td>178</td>
</tr>
<tr>
<td>D6S1043</td>
<td>6q15</td>
<td>AGAT/AGAC</td>
<td>8 to 25</td>
<td>25</td>
</tr>
</tbody>
</table>

*Allele range and number of observed alleles from Appendix 1, J.M. Butler (2012) Advanced Topics in Forensic DNA Typing: Methodology; ‡SE33 alleles have complex repeat structure
Concern: Large loci fail in casework samples and should be avoided – miniSTR capabilities are preferred

- We agree that miniSTRs (smaller amplicons) work best with degraded DNA that is often present in casework samples

- How often are high molecular weight loci failing?

- What data exist on success rates of loci for profiles stored in Forensic Index of CODIS based on PCR product size?
Palm Beach Sheriff’s Office Crime Lab

LDIS Forensic Unknowns – PowerPlex 16 data

**Single-source**

2,452 profiles total
- Loss of Penta D: 633
- Loss of Penta E: 323
- Loss of FGA: 202
- Loss of all 3 loci: 130

\[
\frac{130}{2452} = 5.3\%
\]

FGA loss = 8.2%

**Mixtures**

841 profiles total
- Loss of Penta D: 297
- Loss of Penta E: 296
- Loss of FGA: 179
- Loss of all 3 loci: 55

\[
\frac{55}{841} = 6.5\%
\]

FGA loss = 21.3%

Larger loci are lost in a fraction of casework samples...

Data courtesy of Cecelia Crouse & Tara Sessa (PBSO)
## Additional Data from VA and CA

<table>
<thead>
<tr>
<th></th>
<th>Virginia</th>
<th>California</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># Forensic Unknowns</strong></td>
<td>13,488</td>
<td>37,024</td>
</tr>
<tr>
<td>(single-source profiles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No FGA</strong> (largest of current</td>
<td>68</td>
<td>1,936</td>
</tr>
<tr>
<td>CODIS 13 core loci)</td>
<td>(0.5%)</td>
<td>(4.5%)</td>
</tr>
<tr>
<td><strong>Profiles missing at least</strong></td>
<td>1,609</td>
<td>4,440</td>
</tr>
<tr>
<td>one locus**</td>
<td>(12%)</td>
<td>(12%)</td>
</tr>
</tbody>
</table>

Data courtesy of George Li, Brad Jenkins, and Ken Konzak
Will Performance with Large Multiplexes Be Adversely Impacted with Additional Loci Added?

• There has been significant improvement in kit development in recent years
  – In addition, 6-dye capability of ABI 3500 instruments may play a role in future kits…

• What assay or kit data exist with 20plex (or greater) STR multiplexes?
  – NIST 26plex
  – PowerPlex 21 data collected at NIST
PowerPlex 21

• Promega STR kit to be released in early 2012
  – NIST has been working with this kit since spring 2011
    primarily for concordance testing and has permission
    from Promega to discuss results

• **Contains 20 autosomal STRs + amelogenin**

• **Enables examination of performance characteristics** similar to a future U.S.
  megaplex containing at least 20 loci
DNA Dilution Series with PowerPlex 21

As expected with any STR kit/assay, allele dropout occurs below 100 pg...

Data courtesy of Becky Hill (NIST)
Measurement of Allele Dropout and Extreme Peak Height Imbalance for 2 STR Kits

Three fully heterozygous (except PT83 at Penta D) pristine DNA samples were examined in a dilution series with PowerPlex 21 and Identifiler Plus. Results are ordered by amplicon size and dye color.

**PowerPlex 21 - 30 cycles (5s@2kV)**

**Identifiler Plus - 28 cycles (10s@3kV)**

Data courtesy of Becky Hill (NIST)

Having 5 additional loci did not adversely impact success rates

92% detected

Total alleles possible = 875
Total alleles present = 805

Total alleles possible = 672
Total alleles present = 619

92% detected
Concern: DYS391 is a poor choice and AMEL Y nulls are not a significant concern

• AMEL Y nulls happen…
  – Common practice in some labs is a follow-up test with Y-STRs to confirm that a sample is male
  – Some labs have implemented an additional ChrY test (SRY) to confirm AMEL Y nulls

• A further purpose of having a single Y-STR is to aid QC checks if further Y-STR testing is performed for familial searching or casework purposes
  – DYS391 result will enable a QC check to Yfiler or PowerPlex Y results like D3 and D7 did for Profiler Plu/COfiler (albeit a rather weak one because it is not very polymorphic)
  – By itself, DYS391 is not polymorphic enough to be helpful with any potential familial search filter
Determining the gender of the source of forensic DNA evidence is based on the amelogenin test. However, at times the assay may not be indicative of gender assignment, because of deletions at the amelogenin site. …The study herein addresses the validation of primers for the target SRY gene regarding specificity, sensitivity, and robustness.”
Why Consider DYS391?

• **DYS391 is located on the long arm of the Y-chromosome over 7 Mb away from amelogenin.** Thus, it is likely to be detected in the event of an amelogenin Y deletion that could make a male sample falsely appear as a female (X,-).

• **DYS391 is not very polymorphic.** From a data set of 97,575 haplotypes available on the Y-Chromosome Haplotype Reference Database, over half of them possess allele 10. However, only two null alleles have been reported and 0.01% duplication events (11 total) have been seen in over 700 different population groups from around the world. Thus, it is a stable locus with a relatively narrow allele range.

• **DYS391 has a mutation rate of 0.26%,** which is comparable to most autosomal STRs commonly in use. There have been 38 mutations observed so far in the 14,621 meioses reported in the literature and compiled on YHRD.
DYS391 Variability

YHRD (Y-chromosome Haplotype Reference Database) **data from 97,575 samples**

Allele frequencies for locus DYS391

Allele 10 is most common

http://www.yhrd.org/Research/Loci/DYS391
Deletions of the Y-chromosome can encompass >1 Mb around the AMEL Y region (DYS458 is often lost in these situations)

Most commonly seen in males of **Indian subcontinent origin**

12/649 Malaysian males showed no AMEL Y

5/77 Nepal males showed no AMEL Y
Why mixing Y-STRs and autosomal STRs in a single DNA test is a bad idea...

Offender/arrestee reference samples
- Male samples: will work fine
- Female samples: only autosomal STRs will amplify resulting in a waste of reagents compared to match probability produced

Casework samples
- Mixtures: excess of either male or female DNA will result in poor STR typing results

Missing person samples
- Y-STRs will fail to work on female DNA samples

Do females represent a significant portion of the samples being examined in these specimen categories?
Not all DNA samples tested are male…
And if not male, then Y-STRs fail to amplify!

SDIS Offender/Arrestee Data:
• Virginia (371,000): ~22% female*
• California (1.9 million): ~17% female*
• Illinois (463,000): ~16% female*  
  *Determined to be female based on amelogenin results or meta data

Missing Persons:
  – Unidentified persons: 20% female (1699/8438 cases)
  – Missing persons: 36% female (3278/9012 cases)
• **NDIS Statistics** (Aug 2011):
  – Unidentified human remains: 5,324
  – Missing person cases: 1,039

Per NDIS Custodian (11/4/11):
~45% females in MP cases (by amelogenin results)

Data kindly provided by George Li & Brad Jenkins, Ken Konzak, and Taylor Scott
Summary

• It is vital that an expanded set of core loci be carefully considered and implemented to avoid adventitious hits on large and growing DNA databases.

• There is limited “electrophoretic real-estate” in constructing STR multiplex assays that will work in 5-dye instruments and contain PCR products <500 bp – 6-dye kits and instruments will help.

• The number of females in DNA database and missing persons cases make required use of multiple Y-STRs of questionable utility.

• **Data driven decisions are being made** by the CODIS Core Loci Working Group.

• The CODIS community will be involved in the implementation phase of adding new kits.
Thank you for your attention

**Acknowledgments:** Applied Biosystems, Promega, and Qiagen for STR kits used in concordance studies

**Contact Information**

John Butler  
NIST Fellow  
Group Leader of Applied Genetics  
john.butler@nist.gov  
301-975-4049

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

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