Beyond the STRs: A Comprehensive View of Current Forensic DNA Markers Characterized in the PCR-Based DNA Profiling Standard SRM 2391d

Carolyn (Becky) Steffen, Kevin Kiesler, Lisa Borsuk, and Peter Vallone
National Institute of Standards and Technology
September 2, 2017
Development of the Next PCR-Based DNA Profiling Standard

• As a successor to SRM 2391c
  – Inventory may be depleted by late 2018
  – Develop SRM 2391d now to ensure availability when needed

• Next Generation Sequencing will be used for certification in addition to Capillary Electrophoresis testing
  – Length- and sequence-based genotypes will be provided
  – Include information values for all commercially available forensic markers beyond the STRs

Goal: SRM 2391d will be the most comprehensive NIST forensic SRM to date
NIST Forensic DNA SRMs
Historical Perspective: Past, Present, Future

**Past**

- **RFLP Testing & DNA Probes (1990)**
- **PCR-Based Testing (1995)**
  - VNTR, Dot Blot
  - STR typing (updated 1998)
  - Focus on STR typing
  - VNTR, Dot Blot
- **PCR-Based Testing (2003)**
  - Autosomal STR loci
  - More STR loci added (updated 2008)
- **PCR-Based Y-STR Testing (2003)**
  - Y-STR loci
  - More Y-STR loci added (updated 2008)

**Present**

- **PCR-Based STR Testing (2011)**
  - Autosomal and Y-STR loci
  - More autosomal and Y-STR loci, X-STR loci, and Indels added (updated 2015)
  - Identity and Ancestry SNPs, and Y-Indel added (updated 2017)

**Future...SRM 2391d**

- More Forensic Markers
- Beyond the STRs
How will SRM 2391d values be assigned?

- **NIST Certified Values** will be assigned when multiple CE primer sets **AND** sequencing results are compared
  
  *Highest confidence*; all sources of uncertainty and bias examined

- **Reference Values** will be assigned when multiple CE primer sets **OR** sequencing results are compared
  
  *Fit for purpose*; not all sources of uncertainty have been examined

- **Information Values** will be assigned when only one primer set is used from either CE or sequencing
  
  *For informational purposes*; no guarantees for uncertainty
SRM 2391d: Forensic Markers Planned to be Included

- **Certified Values**
- **Reference Values**
- **Information Values**

- **Autosomal STR**
  - 40 Markers + Amelogenin

- **Y-STR**
  - 29 Markers

- **X-STR**
  - 12 Markers

- **Indels**
  - 30 Markers

- **INNULS**
  - 20 Markers

- **Microhaplotypes**
  - 36 Markers

- **mtDNA Whole Genome**

- **Phenotypic SNPs**
  - 24 ForenSeq
  - 24 Precision ID

- **Identity SNPs**
  - 95 ForenSeq
  - 124 Precision ID

- **Ancestry SNPs**
  - 54 ForenSeq
  - 168 Precision ID
## Which Autosomal STR Markers will have Certified Values?

| Autosomal STR Marker List | MiniFiler | Identifier | NGM | NGM SElect | NGM Detect | Verifier Express | GlobalFiler | PP S5 | PP CS7 | PP 16 | PP 16 HS | PP 18D | PP 21 | PP ESX 17 Fast | PP ESI 17 | PP ESI 17 Pro | PP Fusion | PP Fusion 6C | ESSplex SE | 24plex QS | ForenSeq | Precision ID GF | Precision ID Mixture ID GF | PowerSeq 46GY | CODIS 20 | European Standard Set | Certified Value | Reference Value | Information Value |
|--------------------------|-----------|------------|-----|------------|------------|-----------------|-------------|-------|-------|-------|--------|--------|-------|----------------|----------|----------------|----------|----------------|-------------|------------|-----------------|---------------|------------------|-------------------|
| D1S1656                  |           |            |     |            |            |                 |             |       |       |       |        |        |      |                |           |                |          |                |              |            |                 |               |                 |                  |
| D1S1677                  |           |            |     |            |            |                 |             |       |       |       |        |        |      |                |           |                |          |                |              |            |                 |               |                 |                  |
| D4S2408                  |           |            |     |            |            |                 |             |       |       |       |        |        |      |                |           |                |          |                |              |            |                 |               |                 |                  |

24 Certified Autosomal STR Markers  
1 Reference Autosomal STR Marker  
15 Information Autosomal STR Markers
Which Y-STR Markers will have Certified Values?

<table>
<thead>
<tr>
<th>Y-STR Markers</th>
<th>Y-STR Marker List</th>
<th>Certified Value</th>
<th>Reference Value</th>
<th>Information Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-STR Markers</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
<tr>
<td>ThermoFisher CE STR kits</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
<tr>
<td>Promega CE STR kits</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
<tr>
<td>Qiagen Investigator CE STR kits</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
<tr>
<td>Illumina NGS kit</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
<tr>
<td>ThermoFisher NGS kits</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
<tr>
<td>Promega NGS kits</td>
<td></td>
<td>Certified Value</td>
<td>Reference Value</td>
<td>Information Value</td>
</tr>
</tbody>
</table>

23 Certified Y-STR Markers
0 Reference Y-STR Markers
6 Information Y-STR Markers
Screening and Planning Phase of Development

- **Sample format:**
  - 4 extracted DNA samples
  - 3 single source and 1 mixed sample at a 3:1 ratio (female:male)
  - 2 cell lines spotted onto FTA and 903 paper (intact cells)

- **Concentration of the samples will be ~1-2 ng/µL DNA for the extracted DNA and 7.5 x 10^4 cells spotted on paper**
  - The concentrations will NOT be certified values – just information values

---

Components A-D will have **different profiles** from SRM 2391c
Components E and F will have **the same profiles** as SRM 2391c
“Beyond the STRs”
How will SRM 2391d be tested “Beyond the STRs”?

- **CE** will be performed with kits from two commercial companies:
  - **Qiagen Inc. (2)**: Investigator DIPplex, Investigator Argus X-12 QS
  - **Innogenomics (1)**: InnoTyper 21

- **NGS** will be performed with the HID available sequencing panels/kits

<table>
<thead>
<tr>
<th>Kit Provider</th>
<th>Thermo Fisher (4)</th>
<th>Illumina (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Qiagen Inc. (2)</strong></td>
<td>Precision ID Ancestry Panel</td>
<td>ForenSeq DNA Signature Prep Kit</td>
</tr>
<tr>
<td>Investigator DIPplex</td>
<td>Precision ID Identity Panel</td>
<td>Nextera XT Sample Prep Kit (mtDNA)</td>
</tr>
<tr>
<td>Investigator Argus X-12 QS</td>
<td>Precision ID mtDNA Whole Genome Panel</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ion Ampliseq DNA Phenotyping Panel</td>
<td></td>
</tr>
</tbody>
</table>
How will SRM 2391d be tested?

- **Capillary Electrophoresis (CE)** will be performed with three different instruments:
  - 3130xL and 3500xL Genetic Analyzer (ThermoFisher)
  - Spectrum CE System (Promega) – when available

- **Next Generation Sequencing (NGS)** will be performed with two different instruments:
  - MiSeq FGx (Illumina)
  - Ion S5 XL (ThermoFisher)
What about Sanger Sequencing?

- There are currently NOT plans to Sanger Sequence every certified type for the components of SRM 2391d
  - Sanger and NGS methods were used in parallel to characterize all STR alleles for SRM 2391c
    - All results were fully concordant

- However, if there are any issues, concerns or questionable results:
  - Discordant results between kits
  - Null alleles
  - Any other ambiguities that are observed

**Sanger Sequencing will be used to confirm results**
Other Forensic Markers Under Consideration

• Rapidly-Mutating Y-STR markers

• Non-CODIS STR markers beyond commercial kits

• Others??
  • New kits, instruments and/or software yet to be released
Candidate Screening
Example Data
Data Collection for Sample Screening: Autosomal STR

Example Candidate Sample

Fully Heterozygous with PowerPlex Fusion 6C

*Y-STR Markers
Data Collection for Sample Screening: Y-STR

YFiler Plus Profile

YHRD: No matches in 188,209 Haplotypes  
(Using Minimal Haplotype)
https://yhrd.org

Whit Athey's Haplogroup Predictor: E1b1a
http://www.hprg.com/hapest5/hapest5a/hapest5.htm?order=num

Results Table

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>Fitness</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1b1a</td>
<td>58</td>
<td>100.0</td>
</tr>
<tr>
<td>E1b1b</td>
<td>18</td>
<td>0.0</td>
</tr>
<tr>
<td>G2a</td>
<td>20</td>
<td>0.0</td>
</tr>
<tr>
<td>G2c</td>
<td>5</td>
<td>0.0</td>
</tr>
<tr>
<td>H</td>
<td>25</td>
<td>0.0</td>
</tr>
<tr>
<td>I1</td>
<td>7</td>
<td>0.0</td>
</tr>
<tr>
<td>I2a</td>
<td>20</td>
<td>0.0</td>
</tr>
<tr>
<td>I2a (xI2a1)</td>
<td>7</td>
<td>0.0</td>
</tr>
<tr>
<td>I2b</td>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>I2b (xI2b1)</td>
<td>14</td>
<td>0.0</td>
</tr>
<tr>
<td>J1</td>
<td>11</td>
<td>0.0</td>
</tr>
<tr>
<td>J2a1b</td>
<td>5</td>
<td>0.0</td>
</tr>
<tr>
<td>J2a1h</td>
<td>6</td>
<td>0.0</td>
</tr>
<tr>
<td>J2a1 x J2a1-bh</td>
<td>11</td>
<td>0.0</td>
</tr>
<tr>
<td>J2b</td>
<td>9</td>
<td>0.0</td>
</tr>
<tr>
<td>L</td>
<td>13</td>
<td>0.0</td>
</tr>
<tr>
<td>N</td>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>Q</td>
<td>17</td>
<td>0.0</td>
</tr>
<tr>
<td>R1a</td>
<td>11</td>
<td>0.0</td>
</tr>
<tr>
<td>R1b</td>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td>T</td>
<td>16</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Data Collection for Sample Screening: mtDNA

Illumina mtDNA Whole Genome Sequencing protocol with Nextera XT Sample Prep Kit

EMPOP results: https://empop.online/haplotypes#matches_details

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>Ancestry</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1b1a12</td>
<td>African</td>
<td>unique</td>
</tr>
</tbody>
</table>
Data Collection for Sample Screening: SNPs
ForenSeq SNP Phenotype and Ancestry Estimation

Hair Color Results
- Brown: 0.16
- Red: 0.00
- Black: 0.84
- Blond: 0.00

Eye Color Results
- Intermediate: 0.00
- Brown: 1.00
- Blue: 0.00

Biogeographical Ancestry Results
Distance to Nearest Centroid: 3.36

<table>
<thead>
<tr>
<th>Population/Region, sample Size 2N</th>
<th>Probability of Genotype in each Population</th>
<th>Likelihood Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somali (Africa, 40)</td>
<td>1.57E-15</td>
<td></td>
</tr>
<tr>
<td>African American (ASW) (Africa, 122)</td>
<td>3.04E-16</td>
<td>5.18</td>
</tr>
<tr>
<td>Sandawe (Africa, 80)</td>
<td>1.82E-16</td>
<td>8.64</td>
</tr>
<tr>
<td>Ethiopian Jews (Africa, 64)</td>
<td>1.03E-16</td>
<td>15.3</td>
</tr>
<tr>
<td>African Americans (Africa, 182)</td>
<td>7.11E-17</td>
<td>22.1</td>
</tr>
<tr>
<td>Masai (Africa, 44)</td>
<td>8.17E-18</td>
<td>193.0</td>
</tr>
<tr>
<td>Chagga (Africa, 90)</td>
<td>1.28E-18</td>
<td>1220.0</td>
</tr>
<tr>
<td>Luhya (LWK) (Africa, 198)</td>
<td>4.07E-20</td>
<td>38700.0</td>
</tr>
<tr>
<td>Lusongo (Africa, 16)</td>
<td>3.21E-20</td>
<td>49100.0</td>
</tr>
<tr>
<td>Hansa (Africa, 78)</td>
<td>4.48E-21</td>
<td>351000.0</td>
</tr>
</tbody>
</table>

KiddLab – Set of 55 AISNPs
Population likelihoods based on 55 SNPs and 139 reference populations for this DNA profile
http://frog.med.yale.edu/FrogKB/

Other Markers To Be Determined:
X-STRs, Indels, INNULS, other SNP Panels, and Microhaplotypes
Applications of SRM 2391d
What can you use SRM 2391d for?

• To meet the FBI Quality Assurance Standards: QAS 9.5.5
  
  9.5.5 The laboratory shall check its DNA procedures annually or whenever substantial changes are made to a procedure against an appropriate and available NIST standard reference material or standard traceable to a NIST standard.

• Validation Studies: instrument, commercial kit, and software
  • Developmental and Internal Validations
  • Known, well-characterized samples for all systems commercially available

• Make NIST traceable materials (see http://ts.nist.gov/traceability/)
Support to the Forensic Community

• PCR-Based DNA Profiling Standard Customers
  • U.S. state and local crime laboratories
  • Private sector DNA testing companies
  • Instrument and assay manufacturers
  • Academic institutions
  • National institutes
  • Miscellaneous companies/industry

• Emerging Forensic Technology
  • New Markers
    • CODIS 13 → CODIS 20: January 1, 2017
    • New SNP markers for ancestry and eye/hair color predictions
  • New Methods
    • Next Generation Sequencing (full sequence strings)
    • New CE instruments and STR kits
Summary and Final Thoughts

• The next PCR-Based DNA Profiling Standard is being developed as the most comprehensive forensic SRM yet
  • STR genotypes and haplotypes
  • Information from commercially available forensic markers beyond the STRs

• Capillary Electrophoresis and Next Generation Sequencing will be performed to assign certified, reference, and information values to the final components

• SRM 2391d can be used for validation studies and to support the forensic community as new technologies emerge
Thank you for your attention!

Questions?
becky.steffen@nist.gov
1-301-975-4275

Acknowledgements
Margaret Kline
David Duewer
Michael Coble
Katherine Gettings
Erica Romsos

Disclaimer: Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. 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 NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

All work presented has been reviewed and approved by the NIST Human Subjects Protections Office

Funding: NIST Special Programs Office: Forensic DNA

A copy of this presentation is available at:http://strbase.nist.gov/NISTpub.htm#Presentations