Application of Emerging Technologies at NIST

GenomeID Forum - Emerging Forensic Genomic Applications
Center of Advanced Forensic DNA Analysis
September 9, 2014 - Greenville, NC

Peter M. Vallone, Ph.D.
Applied Genetics Group
Biomolecular Measurement Division

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 it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Information presented does not necessarily represent the official position of the National Institute of Standards and Technology or the U.S. Department of Justice.

Our group receives or has received funding from the FBI Laboratory and the National Institute of Justice.

Outline

- NIST forensic SRMs
- Digital PCR
- Next-generation sequencing

NIST Forensic SRM Timeline

Current Characterization of Forensic SRMs

- 2391c PCR Based DNA profiling standard
  - 68 STR markers (51 autosomal + 17 Y chromosome)
  - STR repeat lengths (alleles) were certified using multiple (unique) PCR primer sets
  - Sanger sequencing was only performed for loci without multiple PCR primer sets (only 10%)
- 2392 & 2392-I Mitochondrial DNA sequencing standard
  - Entire mtGenome (=16,569 bp) was certified by Sanger sequencing
- 2372 Human DNA Quantitation Standard
  - UV absorbance (decadic attenuation) measurement

Goal: Characterize Existing Forensic SRMs with New and Emerging Technologies

- SRM 2391c: Certify sequence information for STR loci
  - Sanger and NGS methods
  - Supports adoption of NGS in forensic community
  - Understand bias inherent to specific NGS platforms: chemistry and bioinformatics
- SRMs 2392 and 2392-I: confirm Sanger data with high coverage NGS methods
  - Detect lower level heteroplasmies (<20 %)
- SRM 2372: certify concentration with an absolute PCR-based method
  - Digital PCR provides this capability
Certified, Reference & Information Values

**Certified Value**
- NIST has highest confidence in accuracy
- All known/suspected sources of bias investigated/taken into account
  - Two or more methods e.g. Sanger sequencing AND genotyping with multiple primer sets

**Reference Value**
- Best estimate of true value
- All possible sources of bias NOT fully investigated by NIST
  - Genotyping with only two sets of primers

**Information Value**
- Of interest and use to SRM user
- Insufficient information available to assess uncertainty of value
  - Genotyping with only one set of primers

Outline

- NIST forensic SRMs
- Digital PCR
- Next-generation sequencing

SRM 2372 DNA Quantitation Standard

- Used for calibrating DNA quantitation standards
  - (qPCR kits)
  - Current stock: 31 month supply
- In the process of preparing SRM 2372a
- Characterize with dPCR versus UV absorbance

Digital PCR (dPCR) Overview

- A sample is partitioned so that individual nucleic acid targets within the sample are localized
  - Microfluidic (Fluidigm BioMark)
  - Emulsion/droplet PCR (Bio-Rad QX100, RainDance)
- Each partition will contain a negative or positive PCR reaction
- Nucleic acid targets may be quantified by counting the regions that contain PCR end-product
  - A standard curve is not required

Fluidigm BioMark

- Fluidic module transfers PCR mastermix onto chip
- ‘Reader’ performs thermal cycling and fluorescence detection (real-time PCR)
- Fluidigm Digital Arrays
  - 12,765 = 765 chambers × 12 panels (samples)
  - 48,770 = 770 chambers × 48 panels (samples)

- Well volumes
  - 6 nL (12 sample)
  - 0.85 nL (48 samples)
- TagMan compatible chemistry
- FAM-VIC dye detection

Fluorescent signal as a function of amplification cycle in 765 dPCR reactions

- Majority of the wells amplify within a narrow range of C\textsubscript{T} values
- Later amplification may be due to:
  - Damaged target
  - Partially blocked target
  - Secondary binding sites
- Grey lines are no amplification

Concentration (copies per microliter) = \frac{\text{total number of wells} \times \text{total number of negative wells}}{\text{total number of wells} \times \text{volume of all PCR reactions (microliters)}}
PCR master mix and DNA template are partitioned into droplets
- 8 strip tubes - up to 96 samples/run
- Thermal cycling is performed on a standard cycler (9700, Veriti)
- Fluorescence from up to 20,000 droplets are detected in the reader (3.5 h)
- Fluorescence intensity for positive and negative droplets are plotted

Validating annealing temperatures for a digital PCR assay (HBB1)

BioRad QX100

Quantitation of a Control DNA Sample Using 4 Unique dPCR Assays

Validating a qPCR assay for dPCR

Design and validate multiple dPCR assays for certification of SRM 2372a

Convert copies/µL and calculate the DNA concentration as ng/µL:

<table>
<thead>
<tr>
<th>Assay</th>
<th>Chromosome</th>
<th>Average ng/µL</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6S474</td>
<td>6</td>
<td>56.3</td>
<td>1.2</td>
</tr>
<tr>
<td>D9S2157</td>
<td>9</td>
<td>55.2</td>
<td>0.7</td>
</tr>
<tr>
<td>HBB1</td>
<td>11</td>
<td>51.7</td>
<td>1.5</td>
</tr>
<tr>
<td>D5S2500</td>
<td>5</td>
<td>51.7</td>
<td>0.9</td>
</tr>
<tr>
<td>D14S1434</td>
<td>14</td>
<td>52.1</td>
<td>0.4</td>
</tr>
<tr>
<td>2PR4</td>
<td>2</td>
<td>50.6</td>
<td>0.9</td>
</tr>
<tr>
<td>22C3</td>
<td>22</td>
<td>50.0</td>
<td>1.2</td>
</tr>
<tr>
<td>EIF5</td>
<td>2</td>
<td>49.0</td>
<td>0.3</td>
</tr>
<tr>
<td>D1P32.3</td>
<td>1</td>
<td>39.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Average of all (except D1P32.3) 52.1 2.5

We are testing and validating multiple dPCR assays for efficiency.
A combination will be used to assign to a concentration for SRM 2372a.

Copy numbers are plotted against annealing temperature.

We are testing and validating multiple dPCR assays for efficiency.
A combination will be used to assign to a concentration for SRM 2372a.
Comparing ddPCR-cdPCR results

### Platforms

- Illumina
  - MiSeq/FGx
  - HiSeq 2000/2500
  - GAIIx
  - NextSeq 500
  - HiSeq X Ten
- Life Technologies
  - SOLiD (5500 series)
  - Ion Torrent PGM
  - Ion Torrent Proton
- Pacific Biosciences
  - PACBIO RS II
- 454 Roche
  - GS p
  - GS FLX+

**Oxford Nanopore**

Others on the horizon

### Multiple NGS Platforms

- Use of multiple platforms to obtain a consensus sequence for the SRMs
- Identify and reduce false positives and negatives
- Identify and control for bias in a specific chemistry and/or informatics pipeline

**PGM**

**HiSeq**

**MiSeq**

**SOLiD 5500**

**consensus**

High confidence sequence information

### NGS Support for mtDNA Analysis

- PGM & MiSeq analysis
- 5% SNP calling threshold
- Concordance across platforms
- Two heteroplasmies in two components, not in previous certificates
- Planned FSIG short communication
- Certificate update Late 2014

Site 1,393 (G/A)

Characterization of SRM 2392 and 2392-I

- Mitochondrial genome sequencing standard

- Sequence the entire mitochondrial genome
  - Two, three and twelve amplicon strategies
- PGM, MiSeq, HiSeq, and SOLiD platforms
- Check concordance with Sanger results
- Detection of heteroplasmy (< 20%)
- Issues with homopolymers
  - Chemistry and informatics related

Outline

- NIST forensic SRMs
- Digital PCR
- Next-generation sequencing
NGS Support for mtDNA Analysis

AmpliSeq Identity Panel

90 autosomal SNPs
- Pakstis 2010
- Kidd 2012
- SNPforID

30 Y-SNPs
- Upper clade branches

SNPforID S2 AmpliSeq Identity Panel

SNP Markers

HID-Ion AmpliSeq™ Identity Panel

Degraded DNA / Sensitivity Study

<table>
<thead>
<tr>
<th>Assays</th>
<th>PGM SNPs</th>
<th>MiniFinder</th>
<th>Identifiler Plus</th>
<th>GlobalFiler</th>
<th>DiPplex</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEGRADED DNA SAMPLES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-250</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>50-200</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>50-150</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>50-100</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>50-75</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SENSITIVITY SAMPLES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 ng</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>0.1 ng</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>0.05 ng</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Degraded DNA Study

Sheared DNA was fractionated by size range
Size selected sheared DNA
- 50 bp to 200 bp
- 50 bp to 150 bp
- 50 bp to 100 bp
- 50 bp to 75 bp
- 25 bp to 50 bp
1ng DNA Input
- AmpliSeq Library Prep
- PGM Sequencing
- HID SNP Genotyper Analysis

Degraded DNA Study
Identity SNP Panel (amplicon size range 80-220 bp)

Degraded DNA Study
Minifiler STR Kit

Degraded DNA Study
DIPplex Indel Kit

Degraded DNA Study

HID SNP Panel Sensitivity Study

Thresholds: analytical = 50 RFU, stochastic = 300 RFU, PHR = 0.5
HID SNP Panel Sensitivity Study

- 4 samples: 10 ng, 1 ng, 100 pg
- 10 ng & 1 ng: All loci all present, 1 sample at 1 locus imbalanced
- 100 pg: 99% present, 97% balanced

HID-Ion Ampliseq Ancestry Panel

- Ion Torrent PGM

Ancestry SNP NGS Assay

- Sample = S277 (GWU)
- West African Self-ID (Cote d’Ivoire)

RMP of profile in 70 world populations

SNPs and Mixtures

3:1 Mixtures

- Two-person mixtures have 9 possible genotype combinations:
- 3 genotypes (Person 1) × 3 genotypes (Person 2)

SNPs in 3:1 Mixtures

- Single source sample and 3:1 mixed sample, major allele frequency plotted for 90 HID SNPs (in ascending order)
Two-person mixture in 3:1 ratio should have frequencies at: 50%, 62.5%, 75%, 87.5% and 100%

SNPs in 3:1 Mixtures

Sequencing STRs on the MiSeq

- Beta version of PowerSeq Auto System
  - Promega – 24plex STR kit (Doug Storts)
  - NIST - Promega - Battelle collaboration
- Designed for use on NGS platforms
  - Primers redesigned for NGS read lengths
  - Protocol developed for Illumina MiSeq
- Ran SRM 2391c + 188 NIST pop samples
- Data analysis with STRait Razor
  - Further parsing of data with custom Java tools

SRM 2391c Component E

- D2S441
  - Length-based (CE) homozygote 10,10
- Sanger and NGS Sequence data
  - (STRait Razor data parsed with Java tools)
    TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA
  - Length-based homozygote, but sequence-based heterozygote

SRM 2391c Component E

- D13S317
  - Length-based (CE) heterozygote 8,12
  - Sanger heterozygote 8, '13'
  - NGS Sequence data heterozygote 8,12
    - (STRait Razor data parsed with Java tools)
      AATC

Further Parsing STRait Razor Output with Java

- Parse sequence output from STRaitRazor
- Goals
  - Master genotype table
  - Tables of: coverage, PHR, stutter, strand bias, etc
  - Confirm expected repeat structure
  - Evaluate error types and frequency

Genotypes concordant with Sanger and CE typing

A→T SNP results in another TATC repeat
Forensic Certified values


Parsing STRait Razor Output with Java

Forensic STR Sequence Diversity
NIST – Battelle Collaboration

Additional Alleles Obtained with NGS

SRM 2391c:
PCR-Based DNA Profiling Standard

- Components A through D: DNA extracts in liquid form
- Components E and F: DNA spotted on 903 / FTA paper
- Certified values for STR alleles based on CE length polymorphisms

Current Values for STR Loci

STR Typing kits and Primer Mixes

100% Concordance with all kits

Updated Typing

### Updated Values for STR Loci

#### Certified Values

<table>
<thead>
<tr>
<th>Locus</th>
<th>Reference Value</th>
<th>Certified Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1S1656</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DYS19</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D2S1338</td>
<td>DYS385a</td>
<td>DYS385a</td>
</tr>
<tr>
<td>D2S441</td>
<td>DYS385b</td>
<td>DYS385b</td>
</tr>
<tr>
<td>D3S1358</td>
<td>DYS389I</td>
<td>DYS389I</td>
</tr>
<tr>
<td>D5S818</td>
<td>DYS389II</td>
<td>DYS389II</td>
</tr>
<tr>
<td>D6S1043</td>
<td>DYS390</td>
<td>DYS390</td>
</tr>
<tr>
<td>D7S820</td>
<td>DYS391</td>
<td>DYS391</td>
</tr>
<tr>
<td>D8S1179</td>
<td>DYS392</td>
<td>DYS392</td>
</tr>
<tr>
<td>D8S1115</td>
<td>DYS393</td>
<td>DYS393</td>
</tr>
<tr>
<td>D10S1248</td>
<td>DYS437</td>
<td>DYS437</td>
</tr>
<tr>
<td>D12S391</td>
<td>DYS438</td>
<td>DYS438</td>
</tr>
<tr>
<td>D13S317</td>
<td>DYS439</td>
<td>DYS439</td>
</tr>
<tr>
<td>D16S539</td>
<td>DYS448</td>
<td>DYS448</td>
</tr>
<tr>
<td>D18S51</td>
<td>DYS456</td>
<td>DYS456</td>
</tr>
<tr>
<td>D19S433</td>
<td>DYS458</td>
<td>DYS458</td>
</tr>
<tr>
<td>D21S11</td>
<td>DYS635</td>
<td>DYS635</td>
</tr>
<tr>
<td>D22S1045</td>
<td>Y GATA H4</td>
<td>Y GATA H4</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>DYS449</td>
<td>DYS449</td>
</tr>
<tr>
<td>FGA</td>
<td>DYS460</td>
<td>DYS460</td>
</tr>
<tr>
<td>Penta D</td>
<td>DYS481</td>
<td>DYS481</td>
</tr>
<tr>
<td>Penta E</td>
<td>DYS518</td>
<td>DYS518</td>
</tr>
<tr>
<td>SE33</td>
<td>DYS533</td>
<td>DYS533</td>
</tr>
<tr>
<td>TH01</td>
<td>DYS549</td>
<td>DYS549</td>
</tr>
<tr>
<td>TPOX</td>
<td>DYS570</td>
<td>DYS570</td>
</tr>
<tr>
<td>vWA</td>
<td>DYS576</td>
<td>DYS576</td>
</tr>
<tr>
<td>DYS627</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DYS643</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DYF387S1a</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DYF387S1b</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

#### Reference Values

<table>
<thead>
<tr>
<th>Locus</th>
<th>Reference Value</th>
<th>Certified Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1GATA113</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D1S1627</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D1S1677</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D2S1776</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D3S3053</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D3S4529</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D4S2364</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D4S2408</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D5S2500</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D6S1017</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D6S474</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D9S1122</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D9S2157</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D10S1435</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D17S1301</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D17S974</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D18S853</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D20S1082</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D20S482</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>F13A01</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>F13B</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>FESFPS</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>LPL</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

#### Information Values

<table>
<thead>
<tr>
<th>Locus</th>
<th>Reference Value</th>
<th>Certified Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argus X-12 kit</td>
<td>Penta C</td>
<td>Penta C</td>
</tr>
<tr>
<td>DXS7132</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS7423</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS8378</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10074</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10079</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10101</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10103</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10134</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10135</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10146</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DXS10148</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>HPRTB</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

*Other Information Value Considerations:*
- DIPplex kit loci
- PGM and MiSeq IISNPs

#### New Y-STR loci in commercial kits (Yfiler Plus & PPY23)
- *Update to be completed by Oct. 2014*

---

### Acknowledgements

NIST
- Katherine Gettings
- Erica Butts
- Margaret Kline
- Dave Duewer
- Kevin Kiesler
- Harish Swaminathan
- Jo Lynne Harenza
- Nate Olson
- Battelle
- Seth Faith
- Rich Guerrieri
- Liz Montano
- Brian Young
- Esley Heizer
- Francisco Martinez
- Promega
- Doug Stotts
- Jay Patel

Thermo Fisher
- Nnamdi Ihuegbu
- Rob Legace
- Narasi Rajagopalan
- Wenczi Liao
- Gloria Lam
- Sharon Wootten
- Joseph Chang

Promega
- Doug Stotts
- Jay Patel

Battelle
- Seth Faith
- Rich Guerrieri
- Liz Montano
- Brian Young
- Esley Heizer
- Francisco Martinez

Promega
- Doug Stotts
- Jay Patel

Contact Information: pete.vallone@nist.gov

---