Forensic DNA Standards for Next Generation Sequencing Platforms

7th Annual
Sequencing, Finishing, and Analysis in the Future Meeting
Santa Fe, New Mexico
June 5, 2012

Dr. Peter M. Vallone, Carolyn R. Hill, Erica L.R. Butts, David L. Duewer, John M. Butler, and Margaret C. Kline

Biochemical Science Division
National Institute of Standards and Technology

National Institute of Standards and Technology

Main campus
Gaithersburg, MD
Boulder, CO
Charleston, SC
≈3000 employees

Material Measurement Laboratory — 6 labs

Biochemical Science Division — 10 divisions

- Applied Genetics
- Multiplex Biomolecular Science
- 6 groups

NIST supplies over 1,300 Standard Reference Materials (SRMs) for industry, academia, and government ***use in calibration of measurements***

Forensic DNA Typing Process

Single source, mixtures, degraded, inhibitors, low copy

Blood, buccal cells, casework evidence → *automated* DNA Extraction → *Real-time PCR* DNA Quantitation (casework only)

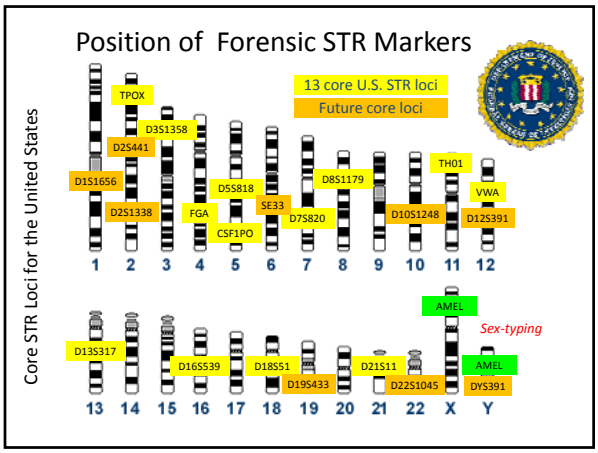
Multiplex PCR (STR Typing Kits) → Separation and detection (Capillary Electrophoresis) → Data interpretation

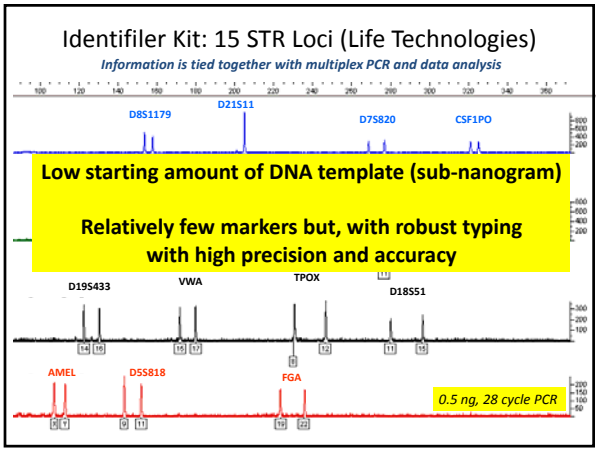
0.5 – 1 ng of genomic DNA
Typically 13-15 markers, plus Amelogenin
Short tandem repeats

1 - 96 capillaries
fluorescent dye color detection

On Windows PC

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_SFAF_June_5.pdf





STR Typing Properties

Fragment analysis

- Typically tetra nucleotide repeats
- PCR products range in size from ≈100-450 bp
- Non-coding region, no phenotype/ancestry information
- No linkage, allele frequencies can be multiplied for random match probabilities ('1 in a trillion')
- 13 CODIS markers are typed in multiplex PCR kits from Promega, Qiagen, Life Technologies
  - 5plex to 23plex range (0.5 ng of genomic DNA)
  - Fluorescently labeled PCR primers

Quality Assurance Standards for Forensic DNA Testing Laboratories

- The standards describe the quality assurance requirements that laboratories performing forensic DNA testing or utilizing the Combined DNA Index System (CODIS) shall follow to ensure the quality and integrity of the data generated by the laboratory.
- **STANDARD 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.
- **STANDARD 9.5.1** Where quantitation is used, **quantitation standards** shall be used.

http://www.fbi.gov/about-us/lab/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011

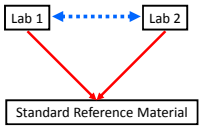
NIST Standard Reference Materials

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

Traceable standards to ensure accurate measurements in our nation's crime laboratories

Helps meet QAS Std. 9.5.5 and ISO 17025

Just released 2391c with expanded CODIS core loci and Y-STRs



SRM 2391c – CODIS core STRs
SRM 2392 & 2392-I – mtDNA
SRM 2395 – Y-STRs
SRM 2372 – Human DNA quantitation

Calibration with SRMs enables confidence in comparisons of results between laboratories

Description of Components in SRM 2391c

Component	Description
A	50 µL of anonymous **female** genomic DNA
B	50 µL of anonymous **male** genomic DNA
C	50 µL of anonymous **male** genomic DNA
D	50 µL of **mixed-source** (Components A and C)
E	Two 6 mm punches of CRL-1486 cells spotted on **903 paper**
F	Two 6 mm punches of HTB-157 cells spotted on **FTA paper**

Liquid components ≈ 2 ng/µL
Paper 75,00 cells per spot

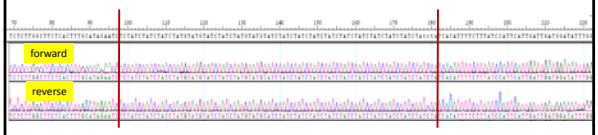
2391c contains confirmed genotypes for:

51 autosomal STR loci (includes FBI core loci, European core loci, others)

17 Y-STR loci

Allele calls (repeats) are confirmed by Sanger sequencing performed at NIST

SRM 2391c, Component C, DYS635



- Repeat motif is confirmed by Sanger sequencing
- Loci amplified (heterozygous samples separated on a gel) and sequenced

Component C (21) =
$[TCTA]_4[TGTA]_2[TCTA]_2[TGTA]_2[TCTA]_{11}$

Table from SRM 2391c

Locus	Component A	B	C	D	E	F
D1S1656	17.3,17.3	11,14	11,15	11,15		
D2S1338	18,23	17,17	19,19	18,19		
D2S441	10,10	10,14	10,10	10		
D3S1358	15,16	15,19	16,18	15,16		
D5S818	11,12	12,13	10,11	10,11		
D7S820	11,11	10,10	10,12	10,11		
D8S1179	13,14	10,13	10,17	10,13		
D8S1115	15,16	15,17	9,9	9,15		
D10S1248	15,16	13,13	12,16	12,15		
D12S391	18.3,22	19,24	19,23	18.3,19		
D13S317	8,8	9,12	11,11	8,1		
D16S539	10,11	10,13	10,10	10,		
D18S51	12,15	13,16	16,19	12,15,		
D19S433	13,14	16.16.2	13.2,15.2	13,13.2,		
D21S11	28,32.2	32,32.2	29,30	28,29.3		
D22S1045	15,15	15,17	16,16	15,		
CSF1PO	10,10	10,11	10,12	10,		
FGA	21,23	20,23	24,26	21,23,		
Penta D	9,13	8,12	10,11	9,10,		
Penta E	5,10	7,15	12,13	5,10,		
SE33	16,18	17,18	28.2,31.2	16,18,28		
TH01	8,9.3	6,9.3	6,8	6,8,		
TPOX	8,8	8,11	11,11	8,1		
vWA	18,19	17,18	16,18	16,18		
Amelogenin	X,X	X,Y	X,Y	X,		
DYS19		14	15	15		
DYS385a		13	13	13		
DYS385b		17	15	15		
DYS389I		13	12	12		
DYS389II		31	27	27		

Genotypes are provided for each component

Forensic labs using commercial STR typing kits should obtain the same results

These materials allow forensic labs to validate/certify:

- DNA extraction method
- PCR (kits and cycler)
- Electrophoretic separation and detection
- Allele calling software

Next Generation Sequencing

Forensic Applications

- Going in depth into STR loci and beyond
  - STRs are useful for legacy (databases)
  - SNPs within STRs identify 'sub-alleles'
  - Millions of bases of sequence variants (SNPs)
- Opens up new human identity applications: biogeographical ancestry, externally visible traits, complex kinship, degraded samples, mixtures, low template, and other applications

Applications are currently being addressed by the forensic genetics community (*Kayser and deKnijff 2011)*

Next Generation Sequencing

- Challenges
  - Repeating sequences (STRs) and read lengths
  - Sample amount requirements (> 1 ng)
  - **Cost** and **time** per unit of information
  - Data analysis (storage, assembly, interpretation)
  - Policy, privacy, disease related markers
  - Validation
  - Standards/reference materials
    - Nomenclature
    - Accuracy of sequence information
    - Errors, platform and bioinformatics-based bias

Requirements for a NGS forensic SRM?

Information gathering stage

- Materials:
  - Genomic DNAs?
  - Cell line DNAs?
  - Use current forensic SRMs?
  - PCR amplicons of forensically relevant markers?
- How many components?
  - Family samples – paternity trio?
  - Mixtures?
- How much material is needed?
- Certify additional markers beyond core STR loci

- Is a full genome standard needed for forensic applications?
- Is it enough to fully characterize 'core' loci and have good/high confidence in other non-core loci?

Ongoing discussion with Multiplex Biomolecular Science group 'Genome in a bottle' consortium – Justin Zook and Marc Salit

Characterization of a NGS forensic SRM?

- Involve multiple platforms/technologies?
  - Short and long read technologies
  - Genomic versus targeted (PCR product) sequencing
  - Interlaboratory studies (pilot prior to SRM production)
- For the core forensic markers (STRs, mito)
  - Continue to include Sanger confirmation
  - Develop specific primer sets for core loci?
- Call SNPs based on **consensus** from multiple platforms and assembly algorithms

Vision for the new SRM

Certified for

- Core autosomal STRs
- X and Y STRs
- Mitochondrial genome
- 'Forensic' SNPs (identification, ancestry, kinship, and phenotype)
- Other: InDels, additional SNPs

- Sufficient material for NGS experiments
- STRs characterized by Sanger sequencing and multiple NGS platforms
- 'Forensic' SNPs characterized with assays (if available) and consensus between multiple NGS platforms
- Uses: test algorithms, library preparation, validation of platforms, enable consistency between labs/platforms

Current Focus - Characterization of SRMs 2392 and 2392-I

- Used for mitochondrial DNA sequencing
- The mtGenome was amplified using an overlapping primer pair strategy
- The amplicons were purified, quantitated and pooled in an equimolar mixture (100 ng)
- Platforms
  - **Ion Torrent (Edge Biosystems)**
  - 5500xl (in house, to be completed this month)
  - Illumina (Beckman, results soon)
  - Hope to have multiple systems in house (2013)

Human HL-60

- Sole component of NIST SRM 2392-I
- NGS accurately called all sequence variants compared to the rCRS (33 variants)

- 315.1 C insertion (homopolymer stretch)
- Heteroplasmy at 12,071 (correctly called)
- Heteroplasmy at 16,069? (artifact of PCR)

Thank you for your attention!

Questions?

peter.vallone@nist.gov
301-975-4872

Acknowledgements
Kevin Kiesler and Mike Coble

Outside funding agencies:
FBI - Evaluation of Forensic DNA Typing as a Biometric Tool
NIJ – Interagency Agreement with the Office of Law Enforcement Standards

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_SFAF_June_5.pdf