Characterization of DNA-Based Certified Reference Materials with New and Emerging Technologies

Peter M. Vallone, PhD*, Margaret C. Kline, MS, Kevin M. Kiesler, MS, Erica L. R. Butts, MFS, David L. Duewer, PhD, Carolyn R. Hill, MS

National Institute of Standards and Technology

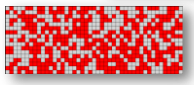
66th Annual AAFS Meeting
February 21, 2014
Seattle, WA

final version can be found at
http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_AAFS_2014.pdf

Disclaimer

- Forensic DNA research conducted at NIST is supported by an interagency agreement between the National Institute of Justice and the NIST Law Enforcement Standards Office.
- 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.

Outline

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

NIST Standard Reference Materials

http://www.nist.gov/srm/

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

Current Human Identity SRMs
SRM 2391c – PCR-Based DNA Profiling
SRM 2392 & 2392-I – mitochondrial DNA
SRM 2372 – Human DNA quantitation Std

Genomic DNAs characterized for the expanded CODIS core loci and Y-STRs

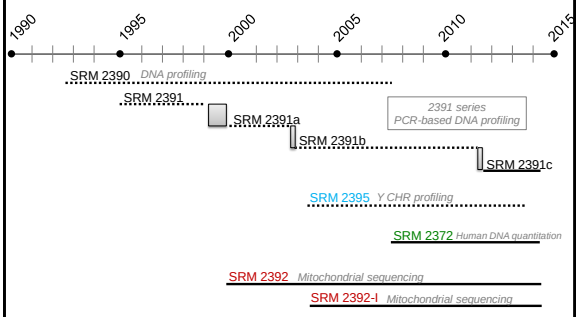
Lab 1 ⟷ Lab 2

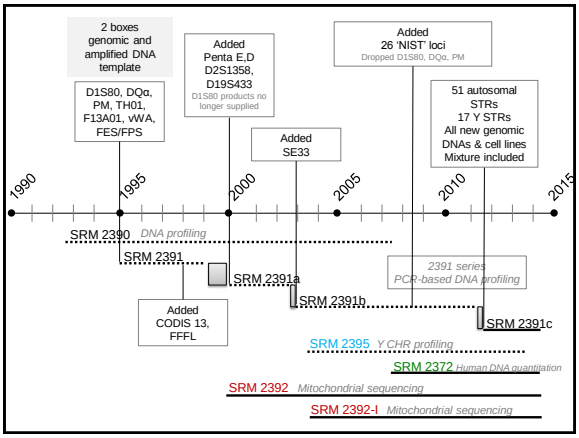
Helps meet QAS Std. 9.5.5 and ISO 17025

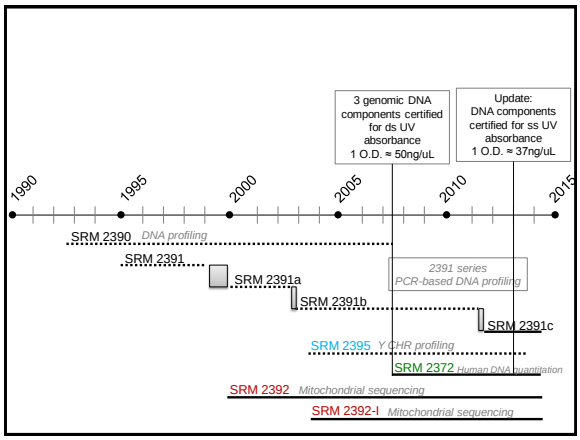
Standard Reference Material

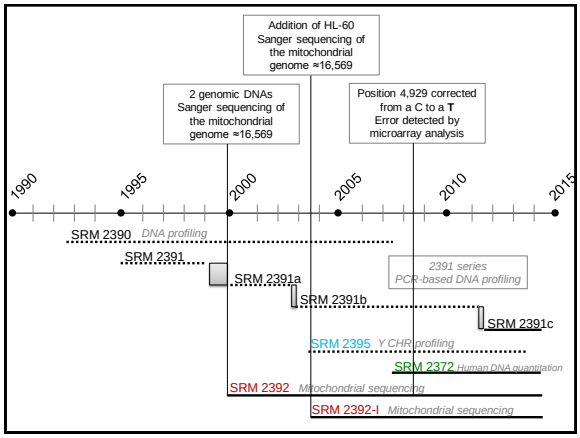
Calibration with SRMs enables confidence in comparisons of results between laboratories

NIST Forensic SRM Timeline









Current Characterization of Forensic SRMs

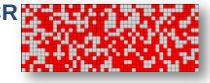
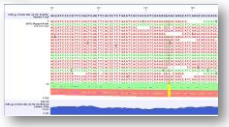
- 2391c **PCR Based DNA profiling standard**
  - 68 STR markers (51 autosomal + 17 Y chromosome)
  - STR repeat lengths (alleles) were certified using multiple (unique) PCR primer sets
  - Sanger sequencing was only performed for loci without multiple PCR primer sets (only 10%)
- 2392 & 2392-I **Mitochondrial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

Goal: Characterize Existing Forensic SRMs with New and Emerging Technologies

- SRM 2391c: Certify sequence information for STR loci (Components A-F)
  - Sanger and NGS methods
- 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

- Supports adoption of NGS in forensic community
- Understand bias inherent to specific NGS platforms: chemistry and bioinformatics

Outline

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

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

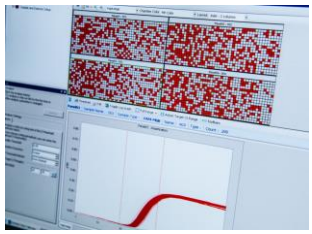
- Sykes, PJ et al. AA (1992) "Quantitation of targets for PCR by use of limiting dilution". Biotechniques 13 (3): 444–449
- Kalinina, O et al. (1997) "Nanoliter scale PCR with TaqMan detection". Nucleic Acids Research 25 (10): 1999–2004
- Vogelstein and Kinzler (1999) "Digital PCR". Proc Natl Acad Sci U S A. 96 (16): 9236–9241
- Pohl and Shih (2004) "Principle and applications of digital PCR". Expert Rev Mol Diagn 4 (1): 41–47
- Dressman et al. (2003). "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations". Proc Natl Acad Sci USA 100 (15): 8817–8822

Fluidigm BioMark

- Fluidic module transfers PCR mastermix onto chip
- 'Reader' performs thermal cycling and fluorescence detection (real-time PCR)

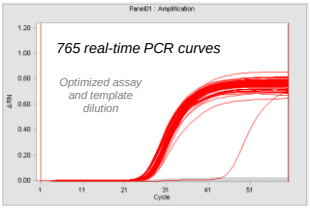
Fluidigm Digital Array

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)
- TaqMan compatible chemistry
- FAM-VIC dye detection

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

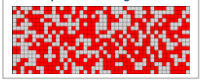


Majority of the wells amplify within a narrow range of C_T values

Later amplification may be due to:
Damaged target
Partially blocked target
Secondary binding sites

Grey lines are no amplification

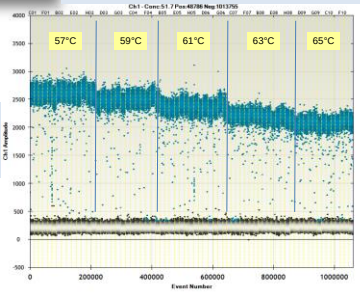
Number of wells with signal relates to the number of copies of starting DNA



$$\text{Concentration (copies per microliter)} = \frac{\text{total number of wells} \cdot \ln\left(\frac{\text{total number of wells}}{\text{total number of negative wells}}\right)}{\text{volume of all PCR reactions (microliters)}}$$

BioRad QX100

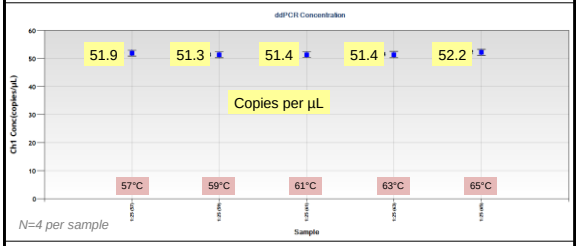
- PCR master mix and DNA template are partitioned into droplets
- 8 strip tubes - up to 96 samples/run
- Thermal cycling is performed on a standard cycler (9700, Veriti)
- Fluorescence from up to 20,000 droplets are detected in the reader (3.5 h)
- Fluorescence intensity for positive and negative droplets are plotted



Validating annealing temperatures for the validation of a digital PCR assay

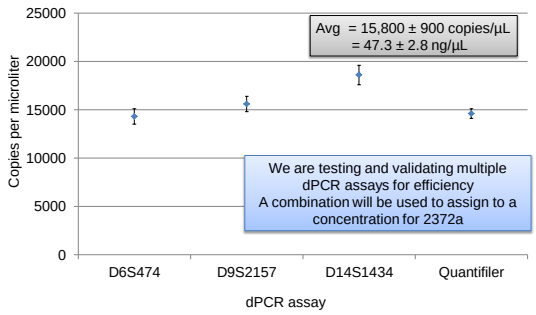
BioRad QX100

Effect of annealing temperature on copy number estimate



N=4 per sample

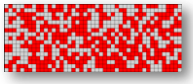
Quantitation of a Control DNA Sample Using 4 Unique Assays Adapted for dPCR



Outline

- NIST forensic SRMs

- Digital PCR



- **Next-generation sequencing**

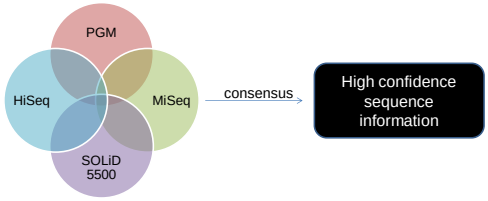
Platforms

- Illumina
  - **MiSeq**
  - 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 jr
  - GS FLX+




Multiple NGS Platforms

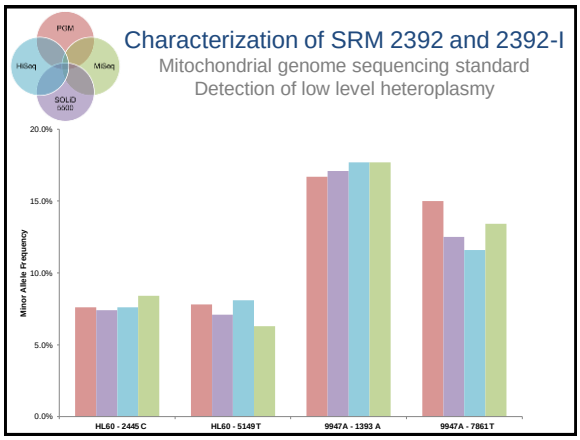
- Use of multiple platforms to obtain a consensus sequence for the SRMs
  - Identify and reduce false positives and negatives
  - Identify and control for bias in a specific chemistry and/or informatics pipeline

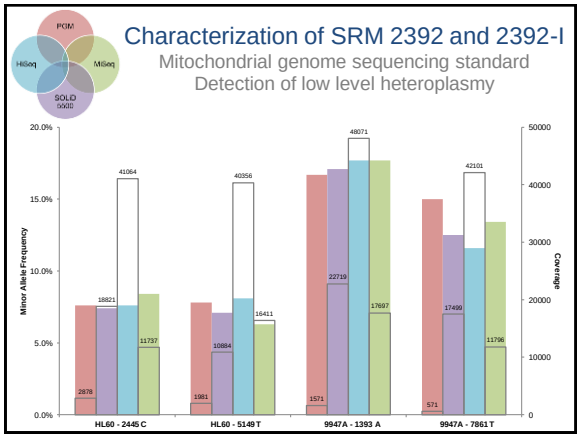


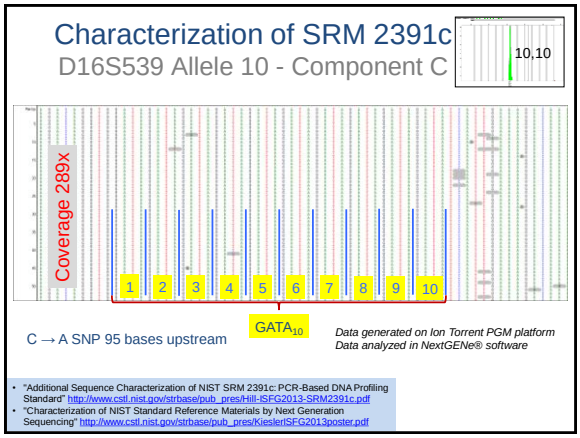
Characterization of SRM 2392 and 2392-I

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

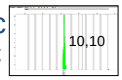

PGM
HiSeq
MiSeq
SOLiD 5500
Characterization of SRM 2392 and 2392-I
Mitochondrial genome sequencing standard
Detection of low level heteroplasmy
Minor Allele Frequency
20.0%
15.0%
10.0%
5.0%
0.0%
HL60 - 2445 C
HL60 - 5149 T
9947A - 1393 A
9947A - 7861 T

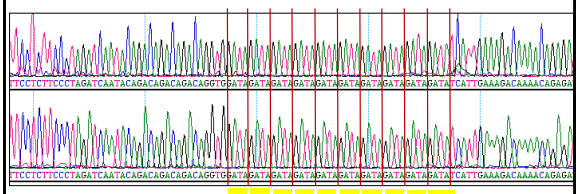

PGM
HiSeq
MiSeq
SOLiD 5500
Characterization of SRM 2392 and 2392-I
Mitochondrial genome sequencing standard
Detection of low level heteroplasmy
Minor Allele Frequency
Coverage
20.0%
15.0%
10.0%
5.0%
0.0%
50000
40000
30000
20000
10000
0
2878
18821
41064
11737
1981
10884
40356
16411
1571
22719
48071
17697
571
17499
42101
11796
HL60 - 2445 C
HL60 - 5149 T
9947A - 1393 A
9947A - 7861 T


Characterization of SRM 2391c
D16S539 Allele 10 - Component C
10,10
Coverage 289x
1
2
3
4
5
6
7
8
9
10
GATA10
C → A SNP 95 bases upstream
Data generated on Ion Torrent PGM platform
Data analyzed in NextGENe® software
"Additional Sequence Characterization of NIST SRM 2391c: PCR-Based DNA Profiling Standard" http://www.cstl.nist.gov/strbase/pub_pres/Hill-ISFG2013-SRM2391c.pdf
"Characterization of NIST Standard Reference Materials by Next Generation Sequencing" http://www.cstl.nist.gov/strbase/pub_pres/KieslerISFG2013poster.pdf

Characterization of SRM 2391c

D16S539 Allele 10 - Component C





Confirming with Sanger

1 2 3 4 5 6 7 8 9 10

- Components A-C have been Sanger sequenced with 24 commercially available autosomal STR markers and 17 Y-STR markers
- Components E and F are currently being sequenced with these same markers

Detection of Repeat Motifs*

Confirmed by NGS and Sanger

Marker	Component	Allele	Allele Repeat Structure
D8S1179	C	17	$[TCTA]_2$ TCTG $[TCTA]_{14}$
D12S391	A	22	$[AGAT]_{13}$ $[AGAC]_8$ AGAT
D12S391	C	19	$[AGAT]_{13}$ $[AGAC]_5$ AGAT
D12S391	C	23	$[AGAT]_{12}$ $[AGAC]_{10}$ AGAT
D21S11	B	32	$[TCTA]_4$ $[TCTG]_6$ {$[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCATA} $[TCTA]_{14}$
SE33	C	31.2	$[AAAG]_2$ AG $[AAAG]_3$ AG $[AAAG]_9$ AAAAAG $[AAAG]_{21}$ G AAGG$[AAAG]_2$ AG
DYS389II	B	31	$[TCTG]_6$ $[TCTA]_{12}$ $[TCTG]_3$ $[TCTA]_{10}$
DYS458	B	17.2	$[GAAA]_{15}$ AA $[GAAA]_2$
DYS635	B	20	$[TCTA]_4$ $[TGTA]_2$ $[TCTA]_2$ $[TGTA]_2$ $[TCTA]_{10}$
DYS635	C	21	$[TCTA]_4$ $[TGTA]_2$ $[TCTA]_2$ $[TGTA]_2$ $[TCTA]_{11}$

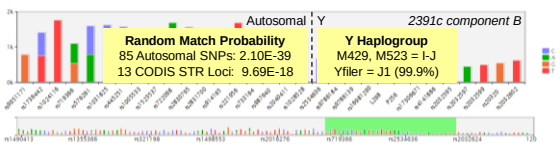
**Sequence variations that were not listed in Butler J.M (2012) or STRBase*

So far: 8 unique flanking region SNPs detected in components A-C

SNP Markers

HID-Ion AmpliSeq™ Identity Panel

- 90 autosomal Identity SNPs (high het, low F_{ST})
- 30 upper Y-clade SNPs
- Goal: to provide informational genotypes for SRM 2391c



HID SNP Genotyper v4.0, Coverage Chart
Avg coverage 1500X (autosomal)

SNP RMP Calculations: frog.med.yale.edu
STR RMP Calculations: Omnipop 150.5
YSTR haplogroup predictor: www.hprg.com

Future Directions

- SRM 2391c
  - Perform STR typing on MiSeq platform
  - Sequence male components for 5 new single copy Y-STR markers and 7 rapidly mutating Y-STR markers
  - Type Qiagen InDel and X-STR kits (information values)
  - Goal: update SRM 2391c by Fall 2014
- Currently working on a candidate material for SRM 2372a to be certified with digital PCR

Thank you for your attention!

Questions?

Peter.Vallone@nist.gov (1-301-975-4872)

Acknowledgements

Dr. Katherine B. Gettings (NIST)

Robert Lagace (Life Tech)

Nnamdi Ihuegbu (Life Tech)

Outside funding agencies:

FBI - Evaluation of Forensic DNA Typing as a Biometric Tool

NIJ – Interagency Agreement with the Office of Law Enforcement Standards

NIST/NRC Postdoc Program

Working in the Applied Genetics Group at NIST

- Current stipend (2014) is **$66,256 per year**
  - Currently a limit of 120 slots per year
  - Congressionally-mandated program for NIST
  - Maximum 2-year appointments
- Awardees **must be U.S. citizens**
- Awardees are chosen through a **national competition** administered by the National Research Council of the National Academy of Sciences.
- Two competitions per year
  - **deadlines of February 1 and August 1**
- **Contact either Dr. Peter Vallone (peter.vallone@nist.gov) or or Dr. Michael Coble (michael.coble@nist.gov)**

Selected Topics
Rapid DNA Typing
DNA Mixture Analysis
Forensic Applications of Next-Gen Sequencing
DNA Extraction efficiency
Forensic SNPs
Y-STRs
Open to suggested topics/projects

http://www.nist.gov/iaao/postdoc.cfm
http://nrc58.nas.edu/RAPLab10/Opportunity/Program.aspx?LabCode=50