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 National Institute of Standards and Technology
U.S. Department of Commerce

Disclaimer

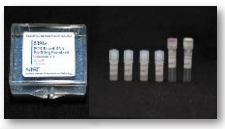
NIST Disclaimer: Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or it imply that any of the materials, instruments or equipment identified are necessarily 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.

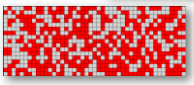
Our group receives or has received funding from the FBI Laboratory and the National Institute of Justice.

Outline

- NIST forensic SRMs

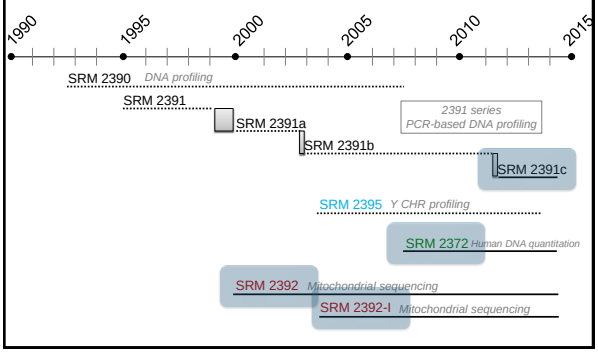


- Digital PCR



- 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 attenuance) 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

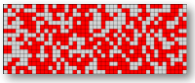
Information Value
- Of interest and use to SRM user
- Insufficient information available to assess uncertainty of value

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 ktis)
- 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

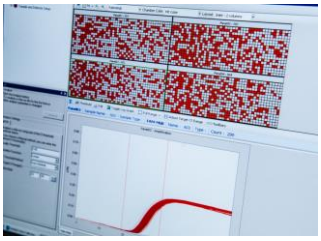
- Sykes, PJ et al. AA (1992) "Quantitation of targets for PCR by use of limiting dilution". Biotechniques 13 (3): 444–449
- Kalinina, O et al. (1997) "Nanoliter scale PCR with TaqMan detection". Nucleic Acids Research 25 (10): 1999–2004
- Vogelstein and Kinzler (1999) "Digital PCR". Proc Natl Acad Sci 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)

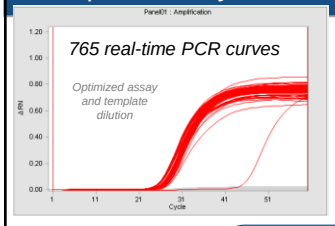
Fluidigm Digital Arrays

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)
- TaqMan compatible chemistry
- FAM-VIC dye detection

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

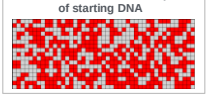


Majority of the wells amplify within a narrow range of C_T values

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

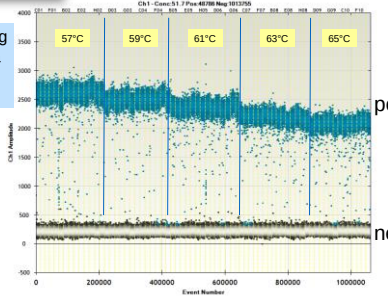


$$\text{Concentration (copies per microliter)} = \frac{\text{total number of wells} \cdot \ln\left(\frac{\text{total number of wells}}{\text{total number of negative wells}}\right)}{\text{volume of all PCR reactions (microliters)}}$$

BioRad QX100

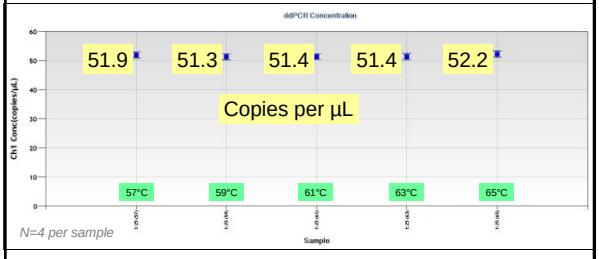
- PCR master mix and DNA template are partitioned into droplets
- 8 strip tubes - up to 96 samples/run
- Thermal cycling is performed on a standard cycler (9700, Veriti)
- Fluorescence from up to 20,000 droplets are detected in the reader (3.5 h)
- Fluorescence intensity for positive and negative droplets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

Validating annealing temperatures for a digital PCR assay (HBB1)



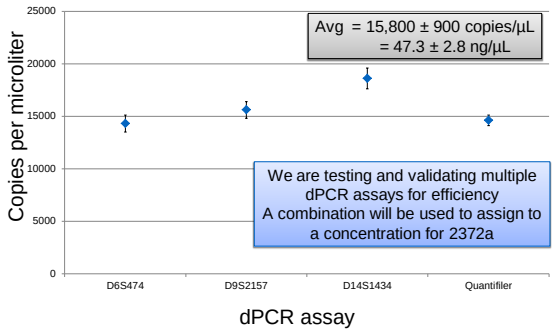
BioRad QX100

Effect of annealing temperature on copy number estimate

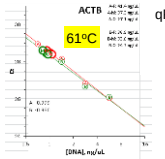


N=4 per sample

Quantitation of a Control DNA Sample Using 4 Unique dPCR Assays

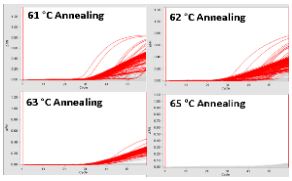


Validating a qPCR assay for dPCR

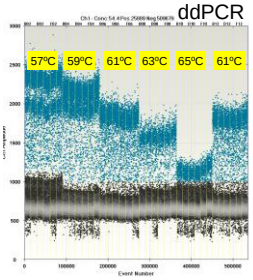


qPCR assay for target ACTB
Linear
$R^2 = 0.99$

cdPCR



ddPCR

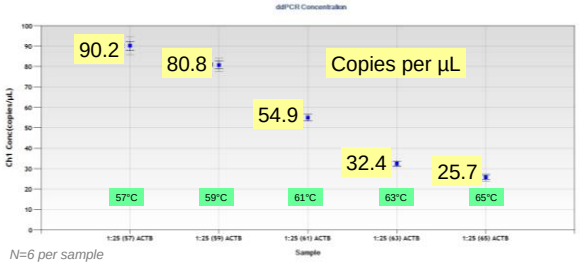


Poor separation between negatives/positives

Assay Optimization

BioRad QX100

Effect of annealing temperature on copy number estimate



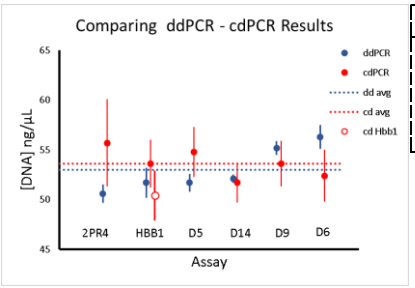
N=6 per sample

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

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

Assay	Chromosome	Average ng/µL	sd
D6S474	6	56.3	1.2
D9S2157	9	55.2	0.7
HBB1	11	51.7	1.5
D5S2500	5	51.7	0.9
D14S1434	14	52.1	0.4
2PR4	2	50.6	0.9
22C3	22	50.0	1.2
EIF5	2	49.0	0.3
D1P32.3	1	39.7	0.8
Average of all (except D1P32.3)		52.1	2.5

Comparing ddPCR-cdPCR results

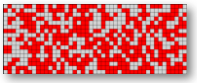


	ddPCR		cdPCR	
Assay	Avg	sd	Avg	sd
2PR4	50.6	0.9	55.7	4.4
HBB1	51.7	1.5	53.6	2.4
D5S2500	51.7	0.9	54.8	2.5
D14S1434	52.1	0.4	51.7	2
D9S2157	55.2	0.7	53.6	2.3
D6S474	56.3	1.2	52.4	2.6
HBB1			50.4	2.5

Outline

- NIST forensic SRMs

- Digital PCR



- **Next-generation sequencing**

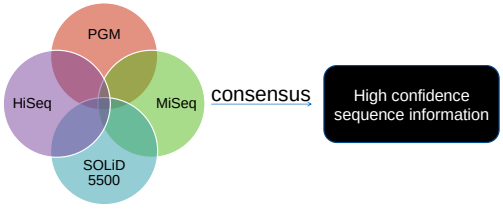
Platforms

- 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 jr
  - GS FLX+

Oxford Nanopore
Qiagen
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



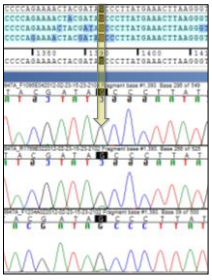
Characterization of SRM 2392 and 2392-I

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

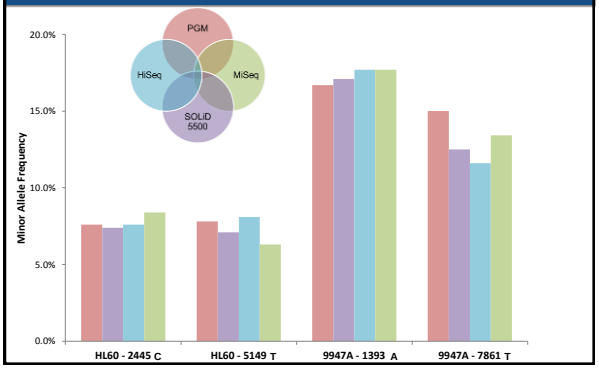
NGS Support for mtDNA Analysis

- 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

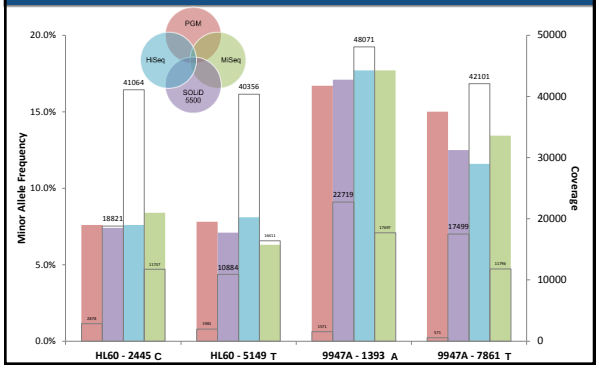


Site 1,393 (G/A)

NGS Support for mtDNA Analysis



NGS Support for mtDNA Analysis

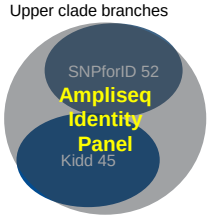


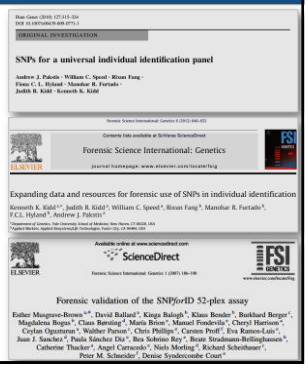
AmpliSeq Identity Panel

Ion Torrent PGM

90 autosomal SNPs
- Pakstis 2010
- Kidd 2012
- SNPforID

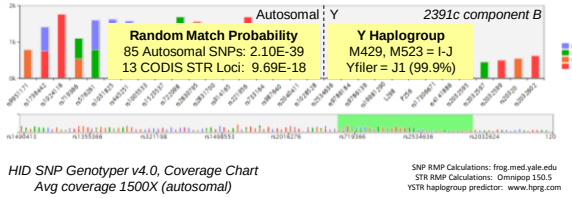
30 Y-SNPs
- Upper clade branches





SNP Markers

HID-Ion AmpliSeq™ Identity Panel



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

Degraded DNA / Sensitivity Study

Assays	PGM IISNPs	MiniFiler	IdentiFiler Plus	GlobalFiler	DIPplex
DEGRADED DNA SAMPLES					
50-250	X	X		X	X
50-200	X	X		X	X
50-150	X	X		X	X
50-100	X	X		X	X
50-75	X	X		X	X
SENSITIVITY SAMPLES					
0.5 ng	X		X	X	X
0.1 ng	X		X	X	X
0.05 ng	X		X	X	X

Degraded DNA Study

Sheared DNA was fractionated by size range

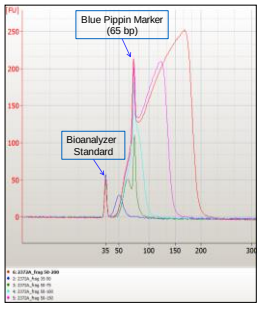
Size selected sheared DNA
- 50 bp to 200 bp
- 50 bp to 150 bp
- 50 bp to 100 bp
- 50 bp to 75 bp
- 35 bp to 50 bp

1ng DNA Input

AmpliSeq Library Prep

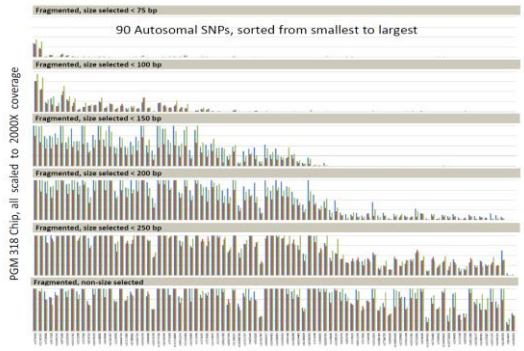
PGM Sequencing

HID SNP Genotyper Analysis



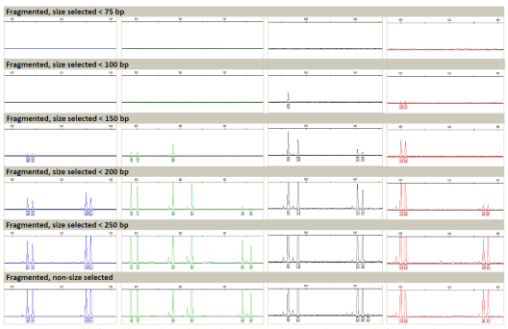
Degraded DNA Study

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



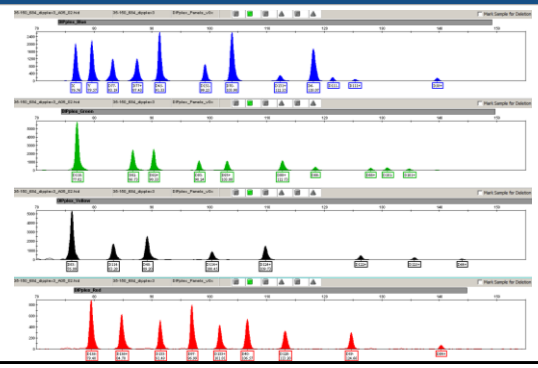
Degraded DNA Study

Minifiler STR Kit

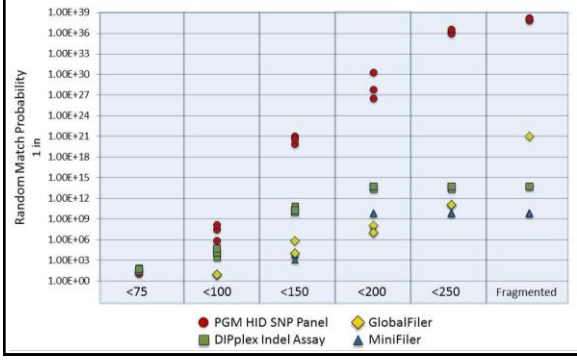


Degraded DNA Study

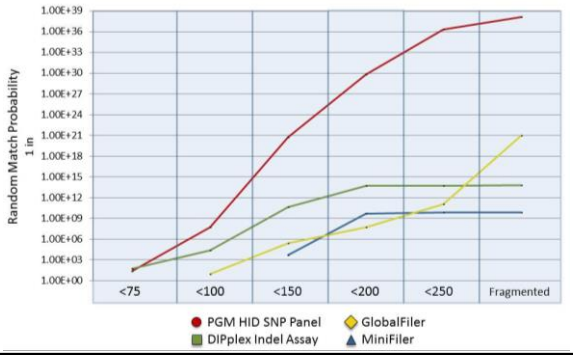
DIPplex Indel Kit



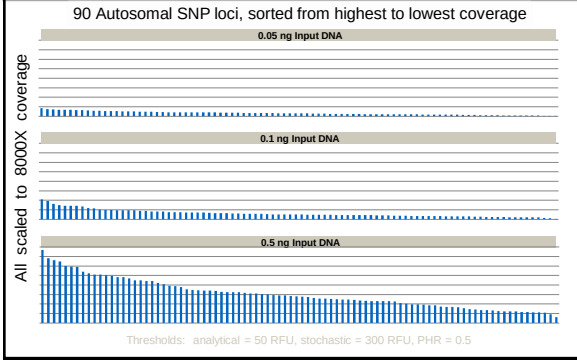
Degraded DNA Study



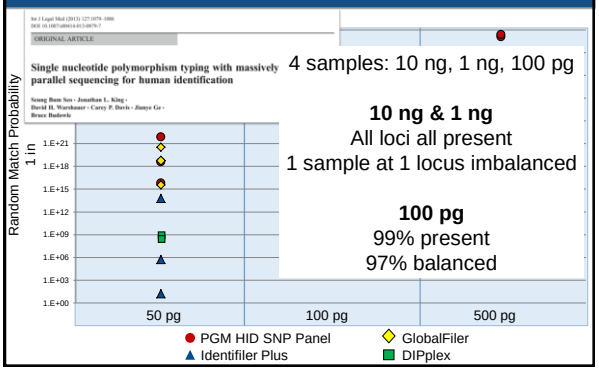
Degraded DNA Study



HID SNP Panel Sensitivity Study



HID SNP Panel Sensitivity Study



4 samples: 10 ng, 1 ng, 100 pg

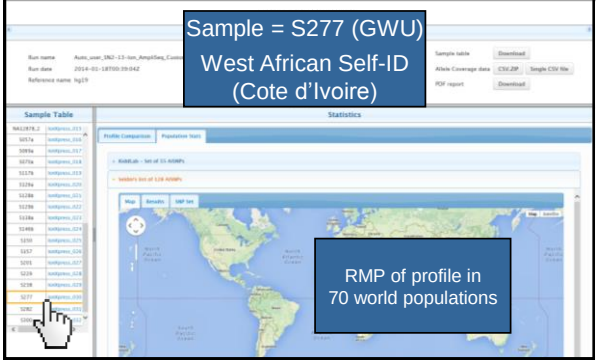
10 ng & 1 ng
All loci all present
1 sample at 1 locus imbalanced

100 pg
99% present
97% balanced

HID-Ion Ampliseq Ancestry Panel

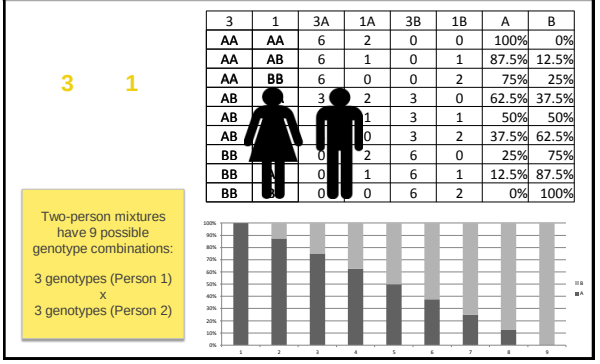
Ion Torrent PGM

Ancestry SNP NGS Assay



SNPs and Mixtures

3 1

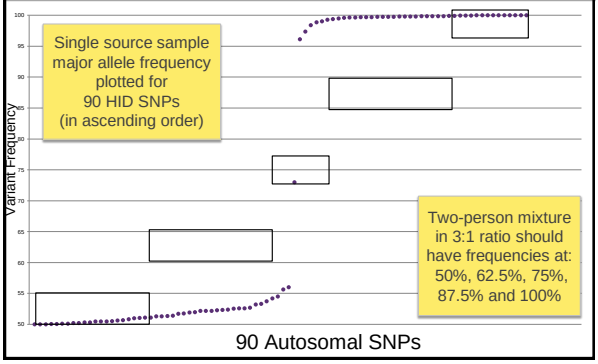


3	1	3A	1A	3B	1B	A	B
AA	AA	6	2	0	0	100%	0%
AA	AB	6	1	0	1	87.5%	12.5%
AA	BB	6	0	0	2	75%	25%
AB	AA	3	2	3	0	62.5%	37.5%
AB	AB	3	1	3	1	50%	50%
AB	BB	3	0	3	2	37.5%	62.5%
BB	AA	0	2	6	0	25%	75%
BB	AB	0	1	6	1	12.5%	87.5%
BB	BB	0	0	6	2	0%	100%

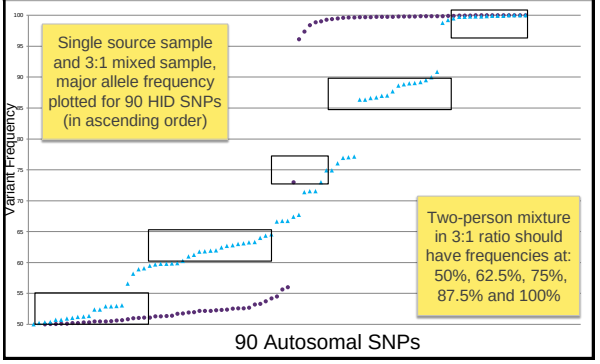
Two-person mixtures have 9 possible genotype combinations:

3 genotypes (Person 1)
x
3 genotypes (Person 2)

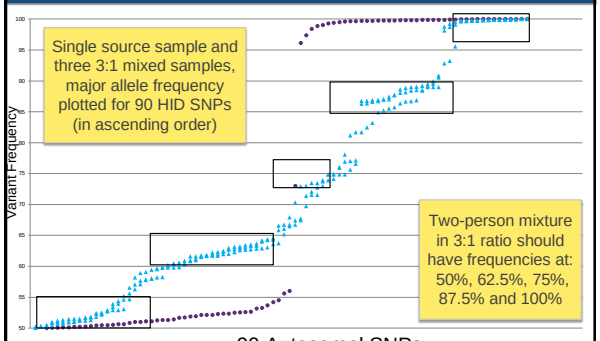
SNPs in 3:1 Mixtures



SNPs in 3:1 Mixtures



SNPs in 3:1 Mixtures

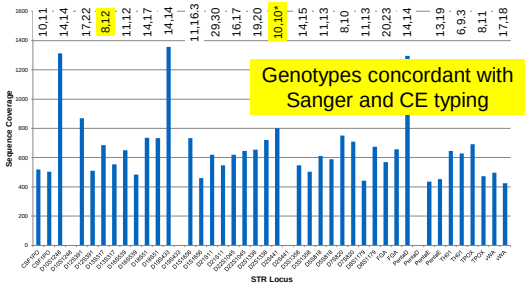


Sequencing STRs on the MiSeq

- 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

STRait Razor: A length-based forensic STR allele-calling tool for use with second generation sequencing data

SRM 2391c Component E



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 TCTG 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)

Flanking/recognition sequence

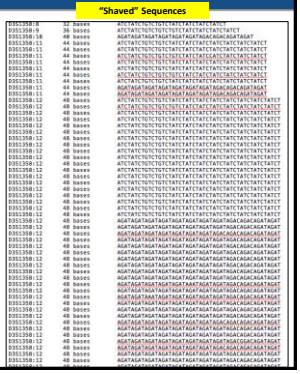
[8] TATC TATC TATC TATC TATC TATC TATC TATC ---- ---- ---- ---- AATCAATCATCTATC

[12] TATC TATC TATC TATC TATC TATC TATC TATC TATC TATC TATC TATC TATCAATCATCTATC

A –T SNP results in another TATC repeat

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

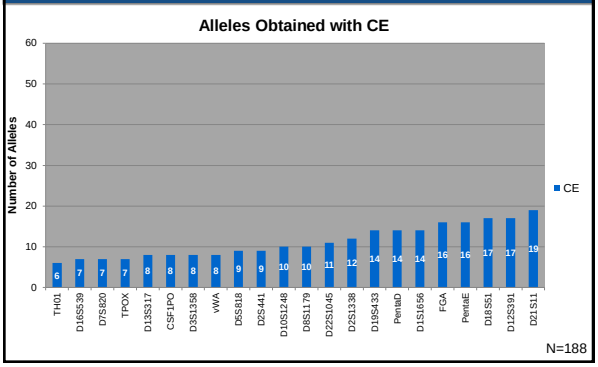


Parsing STRait Razor Output with Java

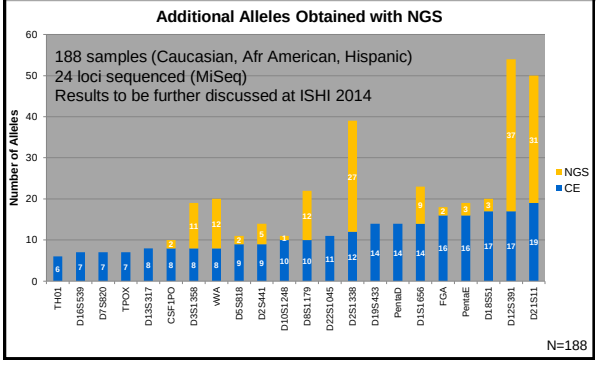
	A	B	C	D
1	Sample name	CSF1PO	CSF1PO	D10S1
2	A10-H-GT37900_S73_L001_R1_001.fastq.STRaitRazor	11	12	
3	A11-H-TT51422_S81_L001_R1_001.fastq.STRaitRazor	10	12	
4	A12-H-ZT80786_S89_L001_R1_001.fastq.STRaitRazor	10	11	
5	A2-C-UT57318_S9_L001_R1_001.fastq.STRaitRazor	11	13	
6	A3-C-WT51362_S17_L001_R1_001.fastq.STRaitRazor	11	12	
7	A4-C-WA29594_S25_L001_R1_001.fastq.STRaitRazor	11	12	
8	A5-AA-JT51471_S33_L001_R1_001.fastq.STRaitRazor	10	13	
9	A6-AA-OT05897_S41_L001_R1_001.fastq.STRaitRazor	12	13	
10	A7-AA-PT84223_S49_L001_R1_001.fastq.STRaitRazor	12	12	
11	A8-AA-PT84232_S57_L001_R1_001.fastq.STRaitRazor	9	10	
12	A9-H-GT37778_S65_L001_R1_001.fastq.STRaitRazor	10	12	
13	B1-2391c-E_S2_L001_R1_001.fastq.STRaitRazor	10	11	
14	B10-H-GT37913_S74_L001_R1_001.fastq.STRaitRazor	11	12	
15	B11-H-TT51435_S82_L001_R1_001.fastq.STRaitRazor	9	12	
16	B12-H-ZT80815_S90_L001_R1_001.fastq.STRaitRazor	10	13	
17	B2-C-WT51342_S10_L001_R1_001.fastq.STRaitRazor	10	12	
18	B3-C-WT51373_S18_L001_R1_001.fastq.STRaitRazor	10	12	
19	B4-C-WA29612_S26_L001_R1_001.fastq.STRaitRazor	11	12	
20	B5-AA-JT51499_S34_L001_R1_001.fastq.STRaitRazor	11	12	

	Locus	height ratio
48		
49	CSF1PO	0.683
50	D10S1248	0.954
51	D12S391	N/A
52	D13S317	0.985
53	D16S539	0.961
54	D18S51	0.799
55	D19S433	0.983
56	D1S1656	0.845
57	D21S11	0.941
58	D22S1045	0.847
59	D2S1338	0.752
60	D2S441	N/A
61	D3S1358	N/A
62	D5S818	N/A
63	D7S820	0.92
64	D8S1179	0.925
65	FGA	0.989
66	PentaD	0.935
67	PentaE	N/A
68	TH01	N/A
69	TPOX	N/A
70	vWA	0.787

Forensic STR Sequence Diversity
NIST – Battelle Collaboration



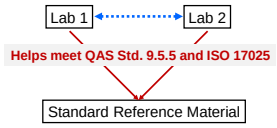
Forensic STR Sequence Diversity
NIST – Battelle Collaboration



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

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



Calibration with SRMs enables confidence in comparisons of results between laboratories

Current Values for STR Loci

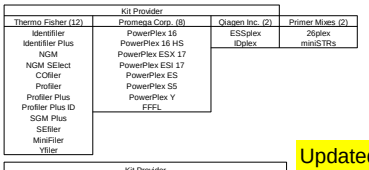
Certified Values		
Autosomal STR (24)	Y STR (17)	X STR (0)
D1S1656	DYS19	None
D2S1338	DYS385a	
D2S441	DYS385b	
D3S1358	DYS389I	
D5S818	DYS389II	
D7S820	DYS390	
D8S1179	DYS391	
D8S1115	DYS392	
D10S1248	DYS393	
D12S391	DYS437	
D13S317	DYS438	
D16S539	DYS439	
D18S51	DYS448	
D19S433	DYS456	
D21S11	DYS458	
D22S1045	DYS635	
CSF1PO	Y GATA H4	
FGA		
Penta D		
Penta E		
SE33		
TH01		
TPOX		
vWA		

Reference Values		
Autosomal STR (23)	Y STR (0)	X STR (0)
D1GATA113	None	None
D1S1627		
D1S1677		
D2S1776		
D3S3053		
D3S4529		
D4S2364		
D4S2408		
D5S2500		
D6S1017		
D6S474		
D9S1122		
D9S2157		
D10S1435		
D17S1301		
D17S974		
D18S853		
D20S1082		
D20S482		
F13A01		
F13B		
FESFPS		
LPL		

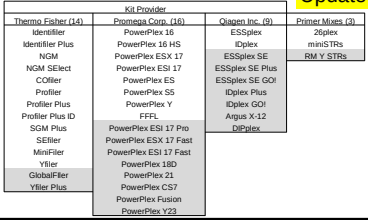
Information Values		
Autosomal STR (1)	Y STR (0)	X STR (0)
Penta C	None	None

STR Typing kits and Primer Mixes

100% Concordance with all kits



Updated Typing



Updated Values for STR Loci

All certified loci have been fully sequenced

Certified Values		
Autosomal STR (25)	Y STR (29)	X STR (0)
D1S1656	DYS19	None
D2S1338	DYS385a	
D2S441	DYS385b	
D3S1358	DYS389I	
D5S818	DYS389II	
D6S1043	DYS390	
D7S820	DYS391	
D8S1179	DYS392	
D8S1115	DYS393	
D10S1248	DYS437	
D12S391	DYS438	
D13S317	DYS439	
D16S539	DYS448	
D18S51	DYS456	
D19S433	DYS458	
D21S11	DYS635	
D22S1045	Y GATA H4	
CSF1PO	DYS449	
FGA	DYS460	
Penta D	DYS481	
Penta E	DYS518	
SE33	DYS533	
TH01	DYS549	
TPOX	DYS570	
vWA	DYS576	
	DYS627	
	DYS643	
	DYF387S1a	
	DYF387S1b	

Reference Values		
Autosomal STR (23)	Y STR (0)	X STR (0)
D1GATA113	None	None
D1S1627		
D1S1677		
D2S1776		
D3S3053		
D3S4529		
D4S2364		
D4S2408		
D5S2500		
D6S1017		
D6S474		
D9S1122		
D9S2157		
D10S1435		
D17S1301		
D17S974		
D18S853		
D20S1082		
D20S482		
F13A01		
F13B		
FESFPS		
LPL		

Information Values		
Autosomal STR (1)	Y STR (0)	X STR (12)
Penta C	None	DXS7132
		DXS7423
		DXS8378
		DXS10074
		DXS10079
		DXS10101
		DXS10103
		DXS10134
		DXS10135
		DXS10146
		DXS10148
		HPRTB

Argus X-12 kit

*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

ThermoFisher
Nnamdi Ihuegbu
Rob Legace
Narasi Rajagopalan
Wenchi Liao
Gloria Lam
Sharon Wootton
Joseph Chang

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

Promega
Doug Storts
Jay Patel

Contact Information
peter.vallone@nist.gov