NIST Update

Peter Vallone, Michael Coble, Becky Hill,
Erica Butts, Kevin Kiesler, Margaret Kline
Applied Genetics Group
U.S. National Institute of Standards and Technology

SWGDAM

July 18, 2013
Dumfries, VA

NIST Human Identity Project Team

within the Applied Genetics Group

Forensic DNA Team

Funding from the **National Institute of Justice (NIJ)** through NIST Office of Law Enforcement Standards


Pete Vallone
Group Leader


Mike Coble


Becky Hill


Margaret Kline

DNA Biometrics Team

Funding from the **FBI** through NIST Information Access Division


Erica Butts


Kevin Kiesler

Data Analysis Support


Dave Duewer

As of April 1, John Butler has moved into the Office of Special Programs and is working on Forensic Science efforts across NIST

Sources of external funding

Topics

- Status of SRM 2372
- Rapid DNA instrumentation
- Casework expert systems
- New Y STR loci
- Next-generation sequencing
- Update on new STR loci and typing kits
- Sequencing of variant alleles (SRM 2391c)
- Completion of PLEX-ID mass spectrometry work
- Assessing DNA extraction efficiency

Status of SRM 2372

- NIST SRM 2372 Human DNA Quantitation Standard – **returned to sale** (as of January 8, 2013)
- Certified based on absorbance value

Why Was SRM 2372 Taken Off the Market?

- During measurement of the DNA samples to verify stability of certified values we observed **that the UV absorbance values for the samples had changed significantly**
  - Not due to degradation of the DNA but rather unraveling or opening up of the DNA strands in the TE^{-4} buffer (**single-stranded DNA absorbs more UV light than double-stranded DNA**)
  - SRM 2372 is certified for UV absorbance
- **The sample changes over time that impact UV absorbance do not appear to affect qPCR sample performance**

How did we re-certify SRM 2372?

- **Force the material to an all ssDNA conformation**
- Measurements were made using a modification of ISO 21571 Annex B "Methods for the quantitation of the extracted DNA"
  - Combine equal volumes of the DNA extract and freshly prepared 0.4 mol/L NaOH
  - Measure against a reference of equal volumes of TE^{-4} buffer and the 0.4 mol/L NaOH
    - Microvolume spectrometers may have issues with NaOH solutions
- Apparent Absorbance is $D_{10\ (260\ nm)} - D_{10\ (320\ nm)}$

Component A	Component B	Component C
0.777 (0.725 – 0.829)	**0.821** (0.739 – 0.903)	**0.804** (0.753 – 0.855)

Convert Apparent Absorbance to ng/µL

- Conventional concentration values are derived from the assertion that a solution of ssDNA with an absorbance of 1.0 at 260 nm and a pathlength of 1.0 cm has a DNA mass concentration of 37 µg/mL (37 ng/µL)

Parameter	A	B	C
2012 DNA Mass Concentration	57	61	59
2007 DNA Mass Concentration	52.4	53.6	54.3
Theoretical difference, %	9 %	14 %	9 %
Theoretical difference, Ct	0.12 cycle	0.19 cycle	0.12 cycle

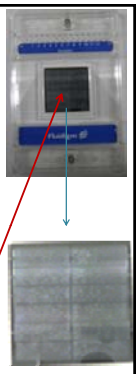
Difference between the original and re-certified values is within the noise of the assay

Digital PCR is Planned as the Next Certification Method

- The next generation of SRM 2372 will be certified for "copy/target number" not UV absorbance
  - dPCR assays require optimization to improve measurement accuracy and reproducibility
- It is important to realize that there is no one human genomic material that will have the same "target number" for all assays; **lots of variability is being discovered at the genome level in terms of copy number variants and chromosomal rearrangements**

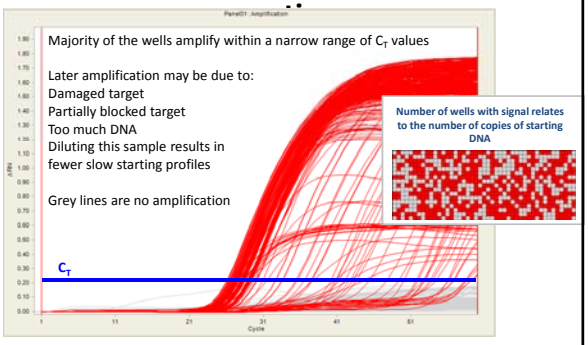
Digital PCR (dPCR) Overview

- Estimates the number of *accessible amplifiable* targets without an external calibrant
- Samples are split into 100s to 1000s of reaction chambers
  - Fluidigm 12.765 Digital Array
  - 765 chambers × 12 panels = **9180 dPCR reactions**
- The count of the number of chambers containing at least 1 target can be used to estimate the total number of targets in a sample

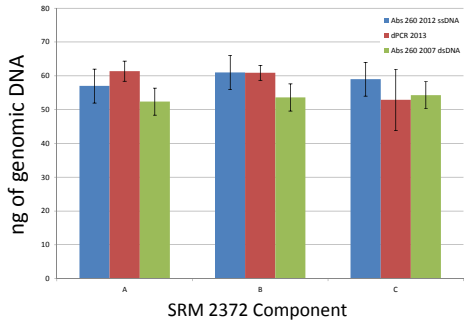


BioRad QX100 Droplet Digital PCR System
≈20,000 PCR droplets

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



dPCR DNA Concentration Estimates Comparison to Absorbance



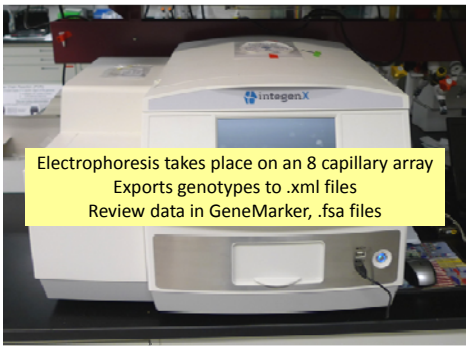
Rapid DNA Prototype Assessment

- Carrying out testing on IntegenX and NetBio R-DNA prototype STR typing instruments

- Over 250 samples (buccal cells on swabs) have been run on each platform
- In the process of assessing genotyping success and providing early feedback for improvement to the vendors

Improvements and optimization are being made to cartridge manufacturing, expert system software, chemistry, and hardware robustness

RapidHit200 IntegenX



Electrophoresis takes place on an 8 capillary array
Exports genotypes to .xml files
Review data in GeneMarker, .fsa files

ANDE NetBio

Electrophoresis takes place on chip
Exports genotypes to .xml files
View egram as bitmap
.fsa files are exported (can be opened in GeneMapper/Genemarker)

Defining Success

- A complete and correct CODIS core 13 STR profile (as called by the expert system software)
  - If any of the 13 loci allele calls were incorrect or absent this was deemed a lane failure
  - Comparing correct genotypes (lab generated) to the types exported to cmf
- Note: we are not including chips that failed due to hardware issues in success calculations

DNA Mixture Interpretation

- Currently exploring the various mixture software programs that use probabilistic approaches to interpretation.
- An oral presentation of this work will be presented at the ISFG in September.

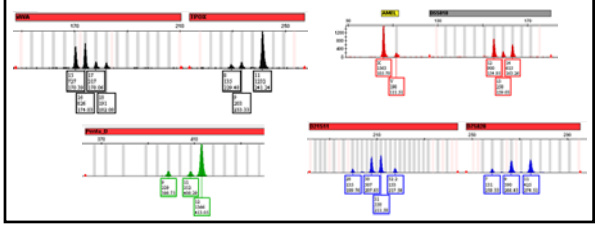
http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm

Lab Retriever

likeLTD (likelihoods for low-template DNA profiles)

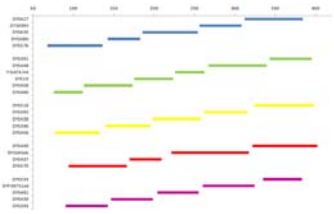
MIX13 – Inter-Laboratory Study

- Currently in the planning stages for an inter-laboratory study (MIX13) with several mixture examples.
- This is to assess how well labs are applying the 2010 SWGDAM recommendations.
- The study will be an interpretational challenge with case scenarios and .fsa files from the 3130 Genetic Analyzer provided (ID Plus and PP16HS).
- Target release is late July. More information is coming soon!



Y STRs

- Currently working on a MTA to test the new 27 Y-STR multiplex kit from Life Technologies (Yfiler Plus). The kit uses 6 fluorescent dye technology.
- Includes all 17 Yfiler loci, 7 Rapidly Mutating (RM) loci, and 3 single copy Y-STR loci.
- Mutation rate information will be needed to assist in interpreting differences among close male relatives.



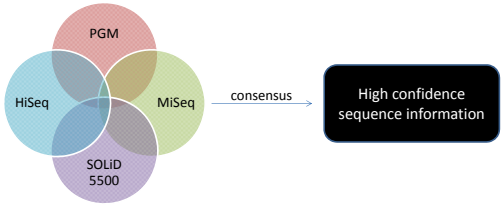
N = 948 males	PowerPlex Y	Yfiler	PowerPlex Y23	Yfiler Plus*
# haplotypes	816	930	945	946
discrimination capacity	0.8608	0.9810	0.9968	0.9979
# times haplotype observed	PPY (12 loci)	Yfiler (17 loci)	PPY23 (23 loci)	Yfiler Plus (27 loci)
1	751	916	942	944
2	42	11	3	2
3	12	2	.	.
4	4	1	.	.
5	2	.	.	.
6	2	.	.	.
7	.	.	.	.
8	1	.	.	.
9	.	.	.	.
10	.	.	.	.
11	1	.	.	.
12	.	.	.	.
13	.	.	.	.
14	.	.	.	.
15	.	.	.	.
16	.	.	.	.
17	.	.	.	.
18	.	.	.	.
19	.	.	.	.
20	1	.	.	.

Next-Generation Sequencing

- Multiple platforms used
  - Illumina
    - HiSeq
    - MiSeq
  - Life Technologies
    - SOLiD4
    - Ion Torrent PGM
- Pilot study sequencing
  - NIST Standard Reference Materials 2392 & 2392-I
    - **For mitochondrial DNA sequencing**
  - Deep sequence coverage
    - 100x to 60,000x
    - Further Characterization
      - Heteroplasmy

Multiple NGS Platforms

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



NIST SRM 2392 & 2392-I

- Mitochondrial DNA sequencing Standard Reference Materials
  - Characterized for mtDNA genome sequence composition
  - Reference used to validate measurement techniques
  - Recommended by FBI as positive control for sequencing labs
- SRM 2392
  - Contains 3 components (extracted DNA)
    - 2392 A – From cell line CHR
    - 2392 B – From cell line 9947A
    - 2392 C – Cloned region of heteroplasmy
- SRM 2392–I
  - From cell line HL-60

False Positives and False Negatives

Using platform specific informatics pipeline

		PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
9947A	FP	1	5	3	21	9	11
	FN	3	4	3	3	3	3
CHR	FP	2	6	10	21	9	10
	FN	3	5	4	3	3	4
HL-60	FP	1	8	8	20	9	8
	FN	1	2	1	1	1	1
Avg Coverage		280	6,500	9,000	49,000	41,000	29,000

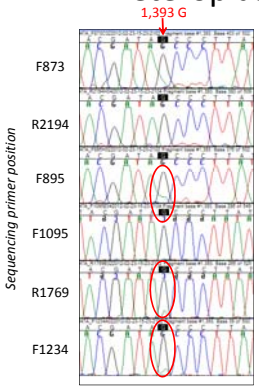
Calls made to the rCRS
On average 99.94 % agreement with Sanger sequencing

Heteroplasmy at Position 1,393

SRM 2392 Component B (9947A)

Nucleotide Position	rCRS Reference Sequence	SRM 2392 Component B Sanger Call	EdgeBio PGM	NIST PGM run 1	NIST PGM run 2	EdgeBio Illumina MiSeq	Beckman Genomics Illumina HiSeq	NIST SOLiD
93	A	G	G	G	G	G	G	G
195	T	C	C	C	C	C	C	C
214	A	G	G	G	G	G	G	G
263	A	G	G	G	G	G	G	G
309.1	:	C						
309.2	:	C						
315.1	:	C						
750	A	G	G	G	G	G	G	G
1393	G	G	G/A	G/A	G/A	G/A	G/A	G/A
1438	A	G	G	G	G	G	G	G
4135	T	C	C	C	C	C	C	C
4769	A	G	G	G	G	G	G	G
7645	T	C	C	C	C	C	C	C
7861	T	C	C	C	C	C	C	C
8448	T	C	C	C	C	C	C	C
8860	A	G	G	G	G	G	G	G
9315	T	C	C	C	C	C	C	C
13572	T	C	C	C	C	C	C	C
13759	G	A	A		A	A	A	A
15326	A	G	G	G	G	G	G	G
16311	T	C	C	C	C	C	C	C
16519	T	C	C	C	C	C	C	C

Heteroplasmy at 1,393?



- 6x coverage by Sanger
- 3/6 of reads indicate low-level heteroplasmy
  - Red circles
- Not reproducible in all reads
  - Not always detected by Sanger sequencing

Heteroplasmy detected by NGS at Site 1,393

- Agreement across platforms (high confidence)

≈ 17.6% (± 2.6%) minor component "A"

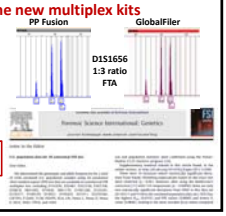
Experiment	Reference "G"	Variant "A"	Coverage
EdgeBio PGM	77.3%	22.7%	97 x
NIST PGM Run 1	82.5%	17.5%	2940 x
NIST PGM Run 2	83.4%	16.6%	3275 x
Illumina MiSeq	83.7%	16.3%	26,234 x
Illumina HiSeq	84.4%	15.6%	62,186 x
NIST SOLiD	82.5%	16.9%	24,226 x

Site 1,393 also confirmed by Niels Morling's lab using 454 technology (Martin Mikkelsen)

Characterizing New STR Loci

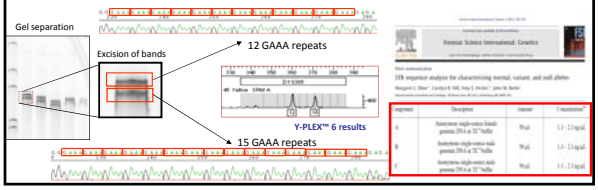
- In April 2011, the FBI announced plans to expand the core loci for the U.S. beyond the current 13 CODIS STRs to 20 total (including DYS391)
- Our group has collected U.S. population data on new loci and characterized them to aid understanding of various marker combinations
- We have recently published the genotypes, allele frequencies and population statistics from these samples at all 29 of these loci in FSI: Genetics
- Recently two commercial kits were released that include the extended core loci GlobalFiler (Life Tech) and PowerPlex Fusion (Promega)
- **We are working closely with the FBI in a consortium validation project to determine how well these new loci perform in the new multiplex kits**

STR Kit or Core Set of Loci	Total N=1036	Caucasians (n=361)	African Am. (n=342)	Hispanics (n=236)	Asians (n=97)
CODIS 13	5.02E-16	2.97E-15	1.14E-15	1.36E-15	1.71E-14
Identifiler	6.18E-19	6.87E-18	1.04E-18	2.73E-18	5.31E-17
PowerPlex 16	2.82E-19	4.24E-18	6.09E-19	1.26E-18	2.55E-17
PowerPlex 18D	3.47E-22	9.82E-21	5.60E-22	2.54E-21	7.92E-20
ESS 12	3.04E-16	9.66E-16	9.25E-16	2.60E-15	3.42E-14
ESI 16 / ESX 16 / NGM	2.80E-20	2.20E-19	6.23E-20	4.03E-19	9.83E-18
ESI 17 / ESX 17 / NGM SElect	1.85E-22	1.74E-21	6.71E-22	3.97E-21	1.87E-19
CODIS 20	9.35E-24	7.32E-23	6.12E-23	8.43E-23	4.22E-21
GlobalFiler	7.73E-28	1.30E-26	3.20E-27	2.27E-26	1.81E-24
PowerPlex Fusion	6.58E-29	2.35E-27	1.59E-28	2.12E-27	1.42E-25
All 29 autosomal STRs	2.24E-37	7.36E-35	3.16E-37	2.93E-35	4.02E-32
29 autoSTRs + DYS391	1.07E-37	3.26E-35	1.77E-37	1.29E-35	2.81E-32



Sequencing Variant Alleles and SRM 2391c

- Sequencing of "off-ladder" variant alleles, null alleles or any other "odd" result seen in datasets is a free service funded by the NIJ
- Results are provided to the customer and listed on STRBase:
  http://www.cstl.nist.gov/biotech/strbase/STRseq.htm
- NIST DNA sequencing procedures and all sequencing primers were published in FSI: Genetics in 2011
- **The purpose of sequencing of SRM 2391c, Components A-C, is to further characterize and determine interesting genomic characteristics within STR fragments - this work will support Next Generation Sequencing of Components A-C**



Completion of PLEX-ID Work

NIST Report to the FBI:
Plex-ID Electrospray Time-of-Flight Mass Spectrometer for Mitochondrial DNA Base Composition Profiling

http://www.cstl.nist.gov/strbase/pub_pres/NIST-report-on-PlexID.pdf

Comparison of Base Composition Analysis and Sanger Sequencing of Mitochondrial DNA for 4 U.S. Populations

Kevin M. Kiesler, Michael D. Coble, Thomas Hall, Peter M. Vallone

Abstract

A set of 711 samples from four U.S. population groups was analyzed using a novel mass spectrometry

Submitted to FSI Genetics

Croat Med J 2013, 54: 225-31

Allele frequencies for 40 autosomal SNP loci typed for US population samples using electrospray ionization mass spectrometry

Assessing DNA Extraction Efficiency

- Absolute extraction efficiency is the ratio of the amount of DNA recovered (quantitated) to the original amount of DNA (known) after extraction
  - Knowing the original starting amount of DNA allows for the ability to evaluate the absolute efficiency of the extraction process
- Currently examining the efficiency of three extraction methods: **Organic**, **Chelex**, and **Qiagen EZ1 Advanced XL** robotic platform.
- DNA from three sources tested in varying amounts: **purified DNA**, **human cell lines**, **whole blood**

Erica Butts presentation at the American Academy of Forensic Sciences meeting (Washington, D.C.), February 21, 2013, "Evaluation of DNA Extraction Efficiency"

Current testing shows a **loss of 70-80% of the initial sample** during the extraction process. The loss is independent of extraction method and source of DNA (i.e. purified DNA, human cells, blood, etc)

Coming Up

- ISFG (Sept 2-7, Melbourne, Aus)
  - Workshop: ***Advanced Topics in Forensic DNA Evidence Interpretation*** (Jo-Anne Bright, John Buckleton, John Butler, Michael Coble, Peter Gill, and Duncan Taylor)
  - Talks: **An investigation of software programs using "drop-out" and "continuous" methods for complex mixture interpretation** (Mike Coble)
  - **Assessing Concordance and STR Kit Performance with PowerPlex ESX 17 and ESI 17 Fast Systems** (Becky Hill)
  - 5 posters
- BCC (Sept 17-19, Tampa, FL)
  - **Update on the NIST R-DNA inter-laboratory study** (Pete Vallone)
- Promega (October 7-10, Atlanta, GA)
  - Workshop: **New Autosomal and Y-STR Loci and Kits: Making Data Driven Decisions** (John Butler, Becky Hill, Mike Coble, TBA)

NIST/NRC Postdoc Program

Working in the Applied Genetics Group at NIST

- Current stipend (2013) is **$65,600 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

Thanks for your attention!

Questions?

Peter.Vallone@nist.gov
301-975-4872

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