Update from the NIST Applied Genetics group

Peter M. Vallone, Ph.D.
Leader, Applied Genetics Group NIST
April 4, 2018
17th Annual Forensic DNA Conference – Bode 2018

Applied Genetics Group – Forensic & Clinical Genetics

Peter Vallone (Group Leader), Becky Steffen, Erica Romsos, Katherine Gettings, Kevin Kiesler, Margaret Kline, Lisa Borsuk, Sarah Riman, David Duewer, Megan Cleveland

Topics
- NIST Standard Reference Materials (SRMs)
  - 2372a (Human DNA Quantitation), 2391d (PCR-based Profiling)
- STRBase 2.0
- Rapid DNA Study
- Sequencing projects

SRM 2372a - Human DNA Quantitation Standard

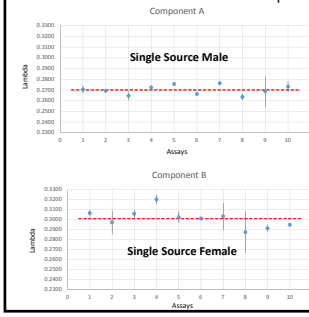
- **On sale March 26, 2018**
- https://www-s.nist.gov/srmors/view_detail.cfm?srm=2372a
- Certified by dPCR measurements

Table 1. Certified Values of Number and Mass Concentration for SRM 2372a[a]
The copy number values are metrologically traceable to the natural units count 1 and ratio 1 and International System of Units (SI) derived units of volume. The DNA mass concentration values are metrologically traceable to the natural units count and ratio 1 and SI derived units of mass and volume.

Component	Copy Number[b] (per nL)	DNA[c] (ng/µL)
A (red cap)	15.1 ± 1.5	49.8 ± 5.0
B (white cap)	17.5 ± 1.8	57.8 ± 5.8
C (blue cap)	14.5 ± 1.5	47.9 ± 4.8

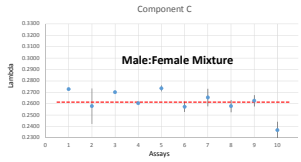
To be used as a qPCR calibrant
OR to assign a value to in house or commercial DNAs

Quantification by dPCR
a method to count DNA copies



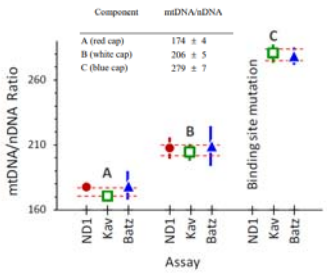
BIO-RAD QX200

Values assigned by 10 human custom genomic assays
Probing single copy targets in the human genome



All samples behave as expected with commercial qPCR assays (11 tested)

SRM 2372a includes the ratio of mitochondrial to nuclear haploid genomes



mtDNA/nDNA ratio for three mitochondrial quantification assays optimized for dPCR.

SRM 2372a provides the ratio of mtDNA to gDNA, which bridges the gap between well characterized mtDNA quantification assays and availability of a commercial standard.

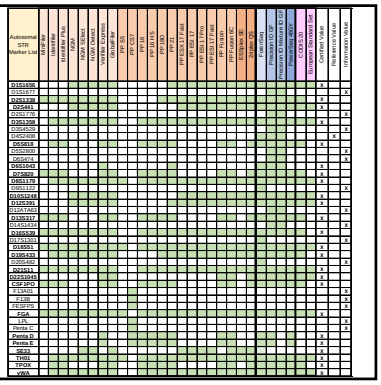
Candidate SRM 2391d

- Successor to SRM 2391c
- Similar format – six tubes
  - A-C three single source components
  - D one mixture; approximately 3:1 (F:M)
  - E-F two components: cells spotted on FTA paper (from cell lines)
- Components A-D are DNA extracted from blood (not cell lines)
- Certified allele calls for core STR loci
- Characterized by CE- and NGS-based methods

Autosomal STR Markers

Autosomal STR Markers
ThermoFisher CE STR kits
Promega CE STR kits
Qiagen Investigator CE STR kits
Illumina NGS kit
ThermoFisher NGS kits
Promega NGS kits
CODIS 20/ESS 12

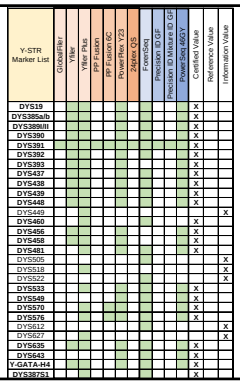
24 Certified Autosomal STR Markers
1 Reference Autosomal STR Marker
15 Information Autosomal STR Markers



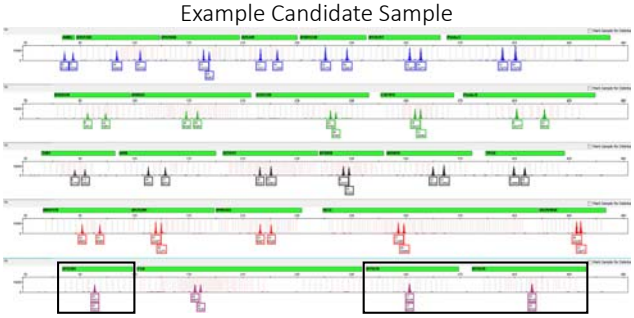
Y-STR Markers

Y-STR Markers
ThermoFisher CE STR kits
Promega CE STR kits
Qiagen Investigator CE STR kits
Illumina NGS kit
ThermoFisher NGS kits
Promega NGS kits

23 Certified Y-STR Markers
0 Reference Y-STR Markers
6 Information Y-STR Markers

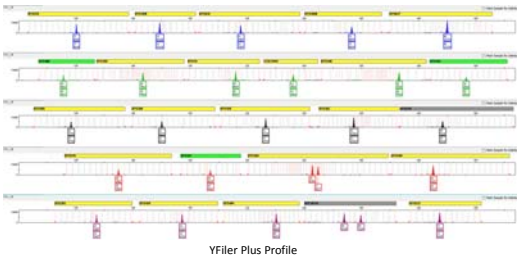


Data Collection for Sample Screening: Autosomal STR Example Candidate Sample



Fully Heterozygous with PowerPlex Fusion 6C *Y-STR Markers

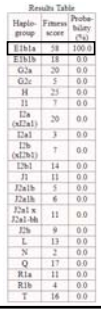
Data Collection for Sample Screening: Y-STR



YFiler Plus Profile

YHRD: No matches in 188,209 Haplotypes (using Minimal Haplotype)
https://yhrd.org

Whit Athey's Haplogroup Predictor: E1b1a
http://www.hprg.com/hapest5/hapest5a/hapest5.htm?order=num



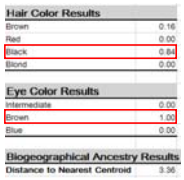
Information for additional marker systems

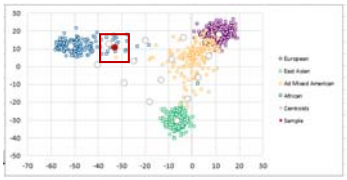
Support the adoption of new markers and technology platforms

- Mitochondrial genome sequence
- Identity SNPs – for degraded samples
- Ancestry SNPs – biogeographical ancestry prediction
- Phenotype SNPs – eye and hair color prediction

Data Collection for Sample Screening: SNPs

ForenSeq SNP Phenotype and Ancestry Estimation





Additional markers to be characterized:
X-STRs, Indels, INNULS, other SNP Panels, and Microhaplotypes

Data Collection for Sample Screening: **mtDNA**

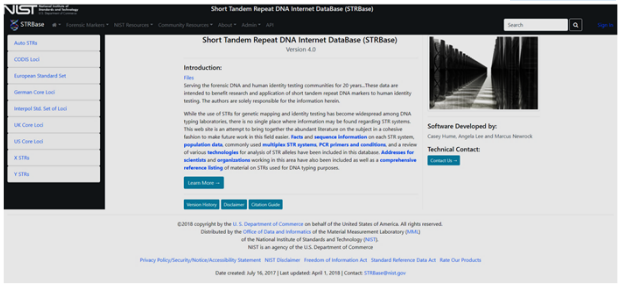
Illumina mtDNA Whole Genome Sequencing protocol with Nextera XT Sample Prep Kit

EMPOP results:
https://empop.online/haplotypes#matches_details

Haplogroup	Ancestry	Match
L1b1a12	African	unique

STRBase 2.0

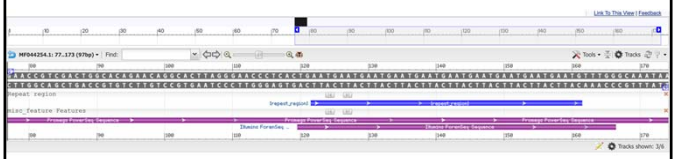
Draft image



STRBase 2.0

- First round of development
  - STR fact sheets (for 24 loci)
  - Variant allele reporting
  - Goal is to have a beta site up this summer
- Provide search, sort, and download functionalities
- Automated submission of variant alleles
- Embedded viewer for STR sequence and presentations
- Other ideas? – let us know

STR Fact sheet: example D1S1656 Visualization of the Sequence



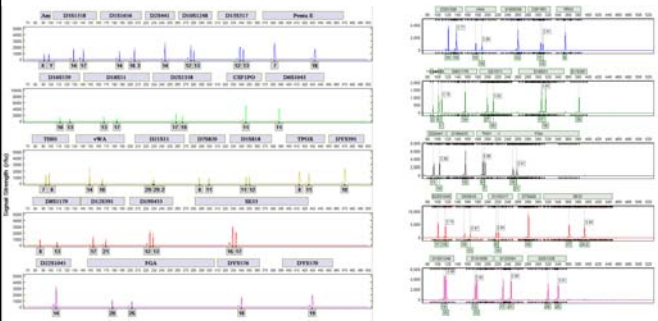
Identify surrounding sequence and provide observed SNPs

Viewer will link to NCBI Bioproject

Rapid DNA Assessment III

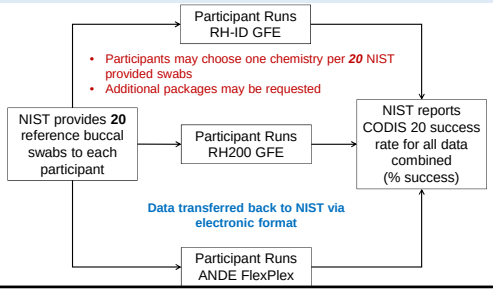
- Spring 2018
- Core **20** STR markers
- **Projected: 8 labs and 2 vendors**
- ANDE and IXI platforms (**new kits, configurations**)
- **20** samples per lab (single source swab)
- Currently collecting single source swabs

Profiles from ***high quality single source*** samples generated by Rapid DNA instruments



NIST Applied Genetics Group
Rapid DNA Maturity Assessment 2018

To measure the status of rapid DNA typing technology for the **20 CODIS core loci**
In support of lab and booking station Rapid DNA implementation



Requirements for Participation

Participants

- Contact Erica Romsos of desired participation in the 2nd Rapid DNA Maturity Assessment
- Establish Material Transfer Agreement (MTA) with NIST for transfer of samples to each participating lab
  - This can take place while samples are being prepared
- Participants are responsible for purchasing kits/cartridges
  - Chemistry **must** contain 20 CODIS Core Loci

NIST

- Collect and characterize buccal reference samples
- Shipment of samples to all participants
- Collection, retention, analysis of data from all participants
  - Through a NIST provided FTP site
- Reporting success metrics to Rapid DNA community

Questions? Contact Erica Romsos (NIST): erica.romsos@nist.gov

Sequencing Projects

- FGx and S5 platforms
- **1036 population samples**
- **Highly polymorphic locus SE33**
- **STRSeq resource**
- **Nomenclature support**
- Sensitivity studies
- Concordance projects

Applications for Sequencing STRs

- Targeted sequencing of STRs
  - STR motif sequence variation; flanking region variation
  - Further understand simple versus complex repeat motifs
  - Characterize stutter
- Greater degree of multiplexing
  - Not confined by dye colors; smaller PCR amplicons (degraded samples)
  - PCR for sample enrichment
  - ***Still using PCR*** – stochastic effects, stutter
- Applications
  - One to one matching?
  - With the new U.S. core loci we are already quite high (>10^{-20})
  - Partial profiles
  - Kinship
- Mixtures
  - Resolve alleles identical by length, but differ by sequence
  - Separate stutter from low level contributors (based on sequence)
  - A sequenced allele *may* have a lower frequency (higher RMP or LR)

Allele frequencies of sequenced STR alleles are needed to formally apply this gain in information -> **Generate population data!**

Sequencing Forensic STRs in Population Samples

When a match is made in a forensic case, allele frequencies are used to calculate how common or rare the DNA profile is in a given population

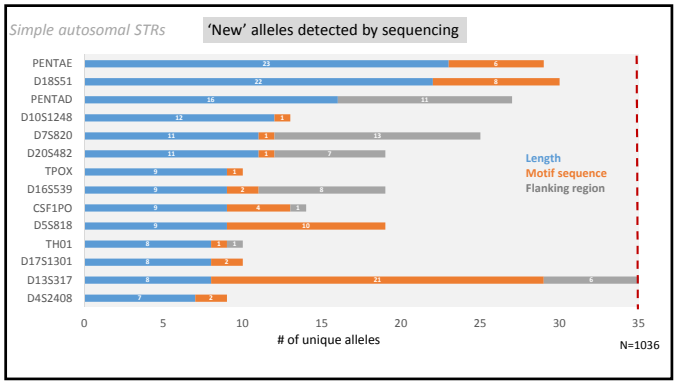
Example of length versus sequence-based frequency calculation:

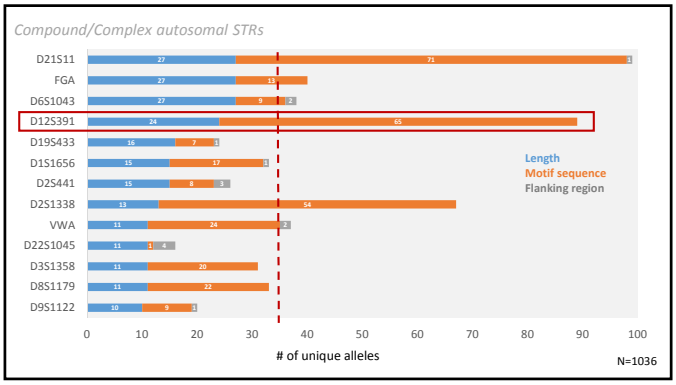
D4S2408

Allele	N	Freq	Sequence Allele	N	Freq	Length	Sequence
7	1	0.6%	[ATCT]7	1	0.6%	8,9	[ATCT]8, [ATCT]9
8	23	14.4%	[ATCT]8	23	14.4%	2pq	2pq
9	60	37.5%	[ATCT]9	18	11.3%	2*0.144*0.375	2*0.144*0.113
			[ATCT] GTCT [ATCT]7	42	26.3%		
10	53	33.1%	[ATCT]10	53	33.1%	0.108	0.033
11	21	13.1%	[ATCT]11	21	13.1%		
12	2	1.3%	[ATCT]12	2	1.3%	1 in 9.3	1in 30.7

Sequencing of 1036 NIST population samples

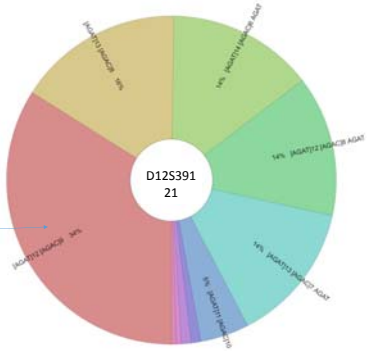
- Work performed on Illumina FGx – ForenSeq kit
- Allele calls were made with Ilumina-UAS and STRait Razor and compared to CE length-based calls (***high confidence***)
- Will include flanking region variation (SNPs, indels)
- Purpose: provide sequence allele frequencies for four U.S. Population groups
  - U.S.: Caucasian, African American, Hispanic, Asian
- The manuscript will be submitted by the end of April
  - Focus on the autosomal loci

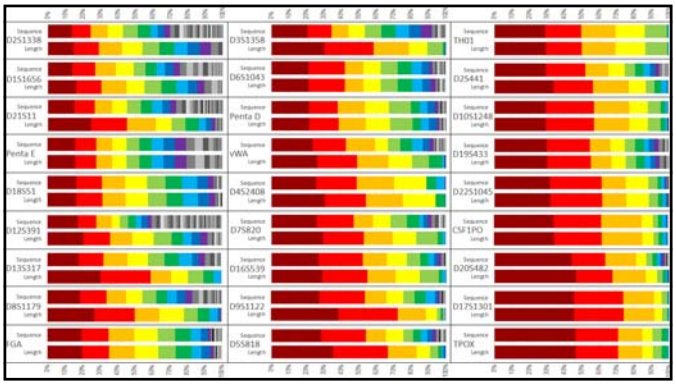




D12S391
Allele "21" frequencies
by sequence 10 "flavors"
Global n=1036

	"Flavors of D12S391 - 21 allele"
1	[AGAT]10 [AGAC]10 AGAT
2	[AGAT]11 [AGAC]10
3	[AGAT]11 [AGAC]9 AGAT
4	[AGAT]12 [AGAC]8 AGAT
5	[AGAT]12 [AGAC]9
6	[AGAT]13 [AGAC]4 AGGC [AGAC]2 AGAT
7	[AGAT]13 [AGAC]7 AGAT
8	[AGAT]13 [AGAC]8
9	[AGAT]14 [AGAC]6 AGAT
10	[AGAT]14 [AGAC]7

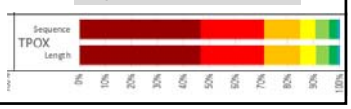




Allele frequencies

D12S391	
Sequence	89 alleles
Length	24 alleles

TH01	
Sequence	10 alleles
Length	8 alleles

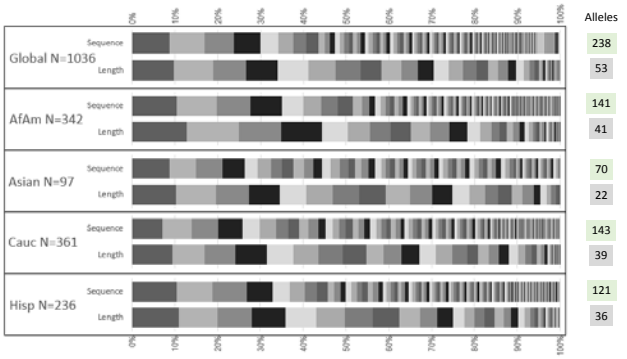


Sequencing of SE33 for 1036 samples

- Separate publication (Electrophoresis Special Issue)
- Focus solely on SE33
- Required further bioinformatic curation of the FGx sequence files

- 53 alleles by length
- 238 allele by sequence
- Data supported by CE-based allele calls

	CE	NGS	CE	NGS
	N		H obs	
All	53	238	0.935	0.965
Cauc	39	143	0.950	0.981
Asian	22	70	0.928	0.959
AfAm	41	141	0.927	0.956
Hisp	36	121	0.928	0.958



STRSeq

Forensic Science International: Genetics

Research paper

STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci

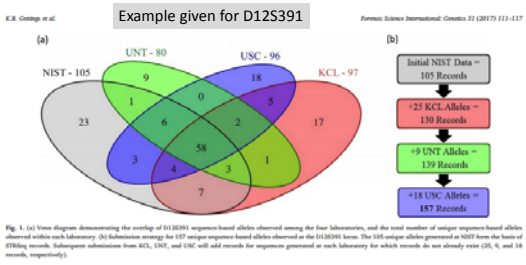
- Bioproject hosted at NCBI to catalog ***unique*** STR alleles
  - https://www.ncbi.nlm.nih.gov/bioproject/380127
- Fully annotated sequence of the STR amplicon
- Each STR allele will have an accession number

https://strider.online/

Collaborative Effort (NIST, KCL, UNT, USC)

1786 + 1043 + 839 + 944 = 4612 samples

Example given for D12S391



STRSeq

- A unique record for each observed allele
- CE-based allele call
- Genomic coordinates
- Platform/kit used to generate sequence

- Bioproject hosted at NCBI to cata
  - https://www.ncbi.nlm.nih.gov/bio
- Fully annotated sequence of the
- Each STR allele will have an acces

Nomenclature Support

Forensic Science International: Genetics

STR allele sequence variation: Current knowledge and future issues

Forensic Science International: Genetics

Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements

https://strider.online

Forensic Science International: Genetics

"The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide

Defining annotated STR reference sequences

Thank you for your attention! Questions?

Contact: Peter.Vallone@nist.gov

- Funding
  - NIST Special Programs Office: *Forensic DNA*
  - FBI Biometrics Center of Excellence: *Forensic DNA Typing as a Biometric tool.*
  - NIJ: *STRSeq and Nomenclature*
  - DHS S&T: *Rapid DNA for Kinship*

- **Disclaimer** - 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 or the Department of Justice. 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.

- All work presented has been reviewed and approved by the NIST Human Subjects Protections Office.