NIST Update: What's new? What's going on?

Erica Romsos
Peter Vallone, Becky Steffen, Lisa Borsuk, Kevin Kiesler, Sarah Riman, Katherine Gettings

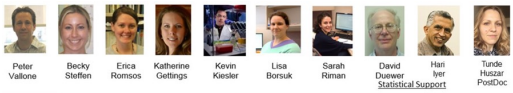
NIST APPLIED GENETICS

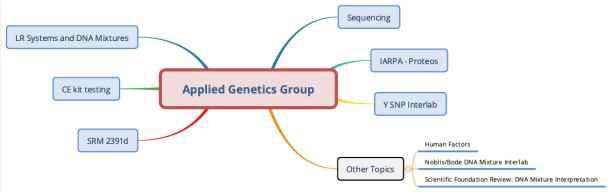
NIST National Institute of Standards and Technology U.S. Department of Commerce

CODIS State Administrators Meeting
Virtually
November 2021

1

Applied Genetics Group – Forensic & Clinical Genetics





2

Updates to SRM 2391d: PCR-Based DNA Profiling Standard

- Performing further analysis of the components and updating the Certificate of Analysis
- CE kits and NGS panels released since 2019 will be added

CE Kits		
Promega	Thermo Fisher	Qiagen (Investigator)
N/A	GlobalFiler IQC	26plex QS
	Y Filer Direct	Argus Y-28 QS
	NGM SElect Express	IDplex Plus
	*Y Indel from GlobalFiler	IDplex GO!
		ESSplex SE QS
		Argus X-12 QS

*Y Indel will be added as an Information Value

NGS Panels			
Promega	Qiagen (Investigator)	Thermo Fisher (Precision ID)	Verogen (ForenSeq)
PowerSeq 46GY	Identity I	Mito Control Region	Kintelligence
	Identity II	Ion Torrent Ampliseq Custom Y-SNP (859)	MainstAY
	Ancestry I		mtDNA Whole Genome
	Ancestry II		mtDNA Control Region
	Microhaplotype STR		

3

SRM 2391d - Summary

- The SRM 2391 series will continue to support the FBI-QAS and the validation and implementation of forensic marker systems

Marker Type	Number of Certified loci	Number of loci with Information values
Autosomal STR	35	13
Y-STR	28	3
X-STR	7	5
Mitochondrial DNA	-	Full mtGenome
Indel/Innuls	-	50 + Y indel
SNPs	-	323 + (10,230 + 859) = 11,412

Certified allele calls supported by sequence data and CE-length based measurements

4

CE Concordance Studies

- VeriFiler Plus PCR Amplification Kit (Thermo Fisher)
  - 595 NIST U.S. population samples tested
  - Completed August 2020

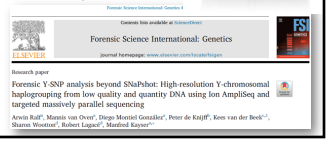
- PowerPlex 35GY System (Promega)
  - 661 U.S. population samples tested
  - Samples were amplified at NIST and sent to Promega to run on the Spectrum CE System
  - Completed March 2021

- Investigator Argus Y-28 Kit (Qiagen)
  - ~1032 male U.S. population samples will be tested July 2021
  - Data to be be includes in Y-STR sequence paper (Steffen, et al.)

5

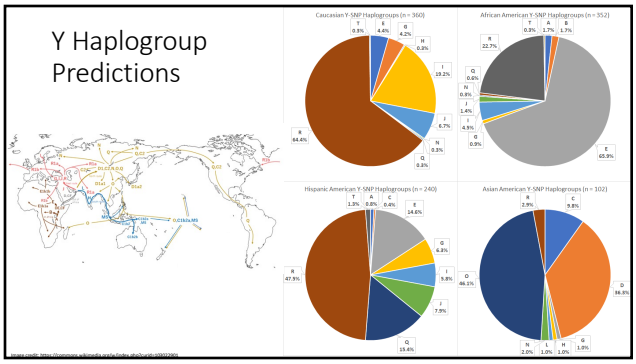
Y SNP Typing Interlaboratory Study

- International collaborative population study
- Erasmus Medical College - Rotterdam, Netherlands
  - Arwin Ralf, Manfred Kayser
- Ion Torrent AmpliSeq Custom Assay
  - 859 Y-chromosome SNPs
  - Resolves > 640 Y haplogroups
  - Run on Ion S5xl sequencer (Thermo Fisher Scientific)
- NIST population samples
  - 1055 samples sequenced
  - 941 contributed to study
    - 351 African American
    - 355 Caucasian
    - 239 Hispanic



6

Y Haplogroup Predictions



7

Support work for Genetically Variant Peptides (GVPs) – IARPA Proteos Project

- Using proteins for human identification (from skin)
- ***Our role:*** Comparison of performer developed extraction protocol (Signature Science) and commercial chemistry (QIAGEN)
  - Dual extraction process for protein and DNA
- Offering support and review of matching statistic calculations

8

Extraction methods examined

Sample → **Extraction** → **Quantification**

ESM (Skin Material) Provided by LLNL
SRM 2372a Component A DNA

1. Qiagen AllPrep DNA/RNA/Protein Kit
2. Signature Science Dual Extraction Protocol
3. Qiagen DNA Investigator without addition of Pro K

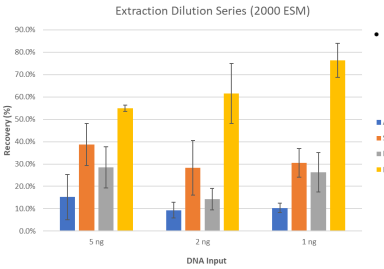
qPCR-Quant Trio (Thermo Fisher) Digital PCR

Samples were extracted and total DNA recovered was calculated
ng/µL * Elution Volume = Total DNA (ng)

Examining the difference in total DNA recovery between all methods

9

DNA Recovery



- Forensic methods exhibited the highest recovery (~20% greater) at all concentrations of DNA input
- SigSci recovery was ~2x that of the AllPrep kit at all concentrations

	Mean	SD
AllPrep	11.6%	3.1%
SigSci	32.5%	5.5%
DI no Pro K	23.0%	7.6%
DI With Pro K	64.3%	11.0%

10

Riman et al. manuscript accepted in PLOS ONE

PLOS ONE

Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset

An independent study to examine the **discrimination performance** and **understand similarities/differences of the assigned LR values**, using two well-cited fully continuous PROBGEN models, **STRmix (proprietary) and EuroForMix (open-source)**, and publicly available ground truth known mixture data (**PROVEDIt database**).

bioRχiv

Examining Discrimination Performance and Likelihood Ratio Values for Two Different Likelihood Ratio Systems Using the Provedit Dataset

- Evaluation of the ability of each software to discriminate between contributor and non-contributor scenarios.
- Assessment of profile log10(LR) values of H1-true tests and H2-true tests on a case-by-case basis.
- Study the distribution of differences in log10(LR) values between the two software.
- Discussion of cases of LR < 1 for H1-true tests and LR > 1 for H2-true tests.
- Evaluation of apparent differences in log10(LR) values.
- Verbal classifications of the numeric LR values assigned by STRmix and EFM.

https://www.biorxiv.org/content/10.1101/2021.05.26.445891v1

11

Summary of different parameters and modelling assumptions of both software

STRmix v2.6	EuroForMix (EFM) v2.1.0
▪ Bayesian approach	▪ Maximum likelihood estimation (MLE) approach
▪ Peak height distribution (Log-normal)	▪ Peak height distribution (Gamma)
▪ Calibration of single source samples	▪ Possible
▪ N-1, N-2 and N+1 stutter peaks modeled	▪ N-1 stutter peaks modeled
▪ Drop-in and degradation models	▪ Drop-in and degradation models were jointly turned on
▪ STRmix reports contain summary statistics (diagnostics)	▪ NA

- Same/fixed mixture EPG features
- Same defined pair of propositions
- True NOC
- Same combination of comparisons (mixture vs POI) per each analysis
- NIST 1036-Caucasian allele frequencies
- Coancestry coefficient correction (Fst or θ) = 0.01

12

Dataset from GlobalFiler PROVEDIt database used in our study

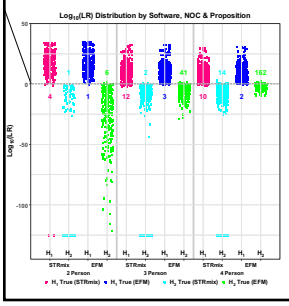
Number of contributors	Number of mixtures of varying DNA quality*	Propositions	Number of H1-true tests	Number of H2-true tests	Mixture ratios	Template amount (pg) (minor)
2	154	H1: POI + U1 H2: U1 + U2	308	308	1:1|1:2|1:4|1:9	15 – 125
3	147	H1: POI + U1 + U2 H2: U1 + U2 + U3	441	441	1:1:1|1:2:1|1:2:2|1:4:1 1:4:4|1:9:1|1:9:9	15 – 125
4	127	H1: POI + U1 + U2 + U3 H2: U1 + U2 + U3 + U4	508	508	1:1:1:1|1:1:2:1|1:1:4:1 1:1:9:1|1:2:2:1|1:4:4:1 1:4:4:4|1:9:9:1	15 – 125

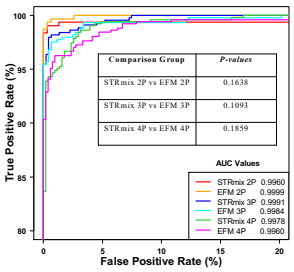
*Pristine DNA or Compromised DNA (enzymatic degradation, UV irradiation, sonication-induced degradation, and PCR inhibition with humic acid).

13

Qualitative and quantitative assessment of the discrimination performance

Overall distribution plots | *Receiver Operating Characteristic (ROC) plots*





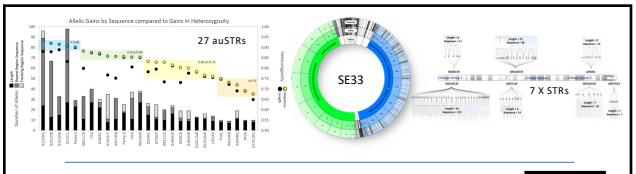
14

Sequencing Projects
(brief updates)

- SNP and Y STR manuscripts
- Lebanese population samples
- Informatics review article
- SWGDAM NGS mixture plate

Future updates to come
- Bode microhaplotypes
- STRSeq
- STRAND

15



Target journal: FSI Reports

Papers in development for 2021:
Y STR sequences and haplogroups
IISNP frequencies and LD analysis with autosomal STRs

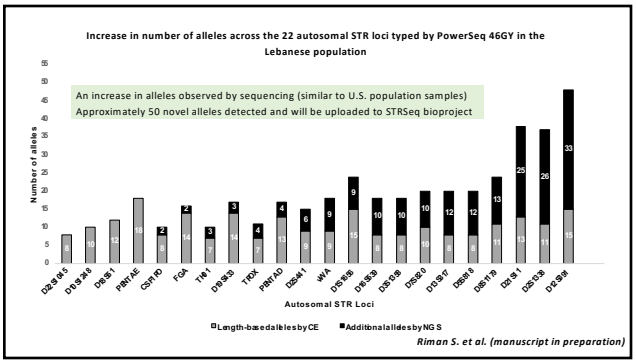
Finalize and disseminate the information derived from the 1036 dataset (ForenSeq marker systems)

16

Characterization of autosomal STR sequence variations in Lebanese population

- Collaborative project with Susan Walsh's laboratory at IUPUI
- Collection of 194 saliva samples from unrelated individuals (85 males and 109 females) of self-reported Lebanese ancestry
- Autosomal STR loci typed by both PowerPlex Fusion 6C (CE-based method) and PowerSeq 46GY System Prototype (NGS-based method) - 22 autosomal STR loci in common
- Identify and characterize sequence variations
- Conduct population genetics data analysis for both the length-based and sequence-based allele calls
- Compare frequencies to geographically neighboring Levantine and Mediterranean countries

17



18

The analysis for NGS/MPS data with open source and commercial software/algorithms

Snapshot of current state of software options for sequencing data analysis of forensic markers.

A one-stop review from integrated, commercial to free and open-source solutions.

Peter Vallone

Katherine Gettings

Tunde Huszar

Papers in development for 2021

Journal: **Genes**

- *Recently submitted*
- Review article

GCAT TACG GCAT

MDPI genes

19

SWGDAM NGS Mixture Plate

- Designed based on survey feedback from 10 organizations (companies and academics)
  - Included developers of PG software, and developers of NGS analysis tools
- Combinations of individuals was chosen to be middle of the road (not hard, not easy)
- To be typed with ForenSeq, TF Precision ID, and Promega PowerSeq 46GY

20

4 combinations of 4P mixtures

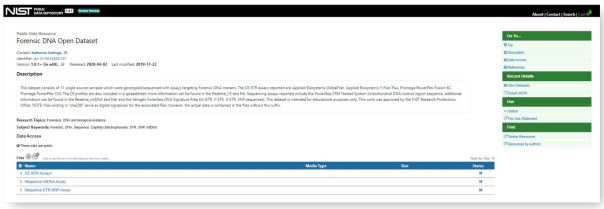
3 combinations of 3P mixtures

5 combinations of 5P mixtures

Columns 1, 5, and 9 are replicates of the same 3P mixture sensitivity series
Ranges in quantity from 0.25 ng/µL to 4 ng/µL
Will be included in every run – even if running sets of 32 samples

21

All data files will be uploaded to the NIST MIDAS Data sharing site for public access and download



Questions about this dataset? Contact Katherine Gettings (Katherine.gettings@nist.gov)

22

NIJ/NIST Expert Working Group on Human Factors in Forensic DNA Interpretation

Slides provided by Niki Osborne and Melissa Taylor

23

Charge

The Expert Working Group on Human Factors in Forensic DNA Interpretation is charged with conducting a scientific assessment on the effects of human factors in forensic DNA examination with the goal of recommending approaches to improve its practice and reduce the likelihood of errors. The Working Group will evaluate relevant bodies of scientific literature and technical knowledge to develop its recommendations and will publish a report of its findings.

24

Key Topics

Each subgroup will take the lead on writing about their topics.

Research needs are being covered by each subgroup.

- QA/QC
- Testimony and Reporting
- Research, Education, and Training
- Interpretation and Technology
- Management
- Work Environment
- Research Needs

25

Updates

- First convened in-person in February 2020 with 24 members
- 15 x Full Group webinars since June 2020
- Currently have 27 members spanning academia, practitioners, researchers, legal community members
- 200+ pages of draft report written
- Report is in the first round of internal reviews and edits

Currently in the first round of internal review of the draft report.

26

Thank you for your attention! Questions?

Contact: Erica.Romsos@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*

- **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. 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.

27