Forensic Genetics: Research Projects and Standards Production

Peter M. Vallone, Ph.D.
Leader, Applied Genetics Group
Forensics at NIST 2020
November 5, 2020

Disclaimer

- Points of view in this document are those of the author 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.

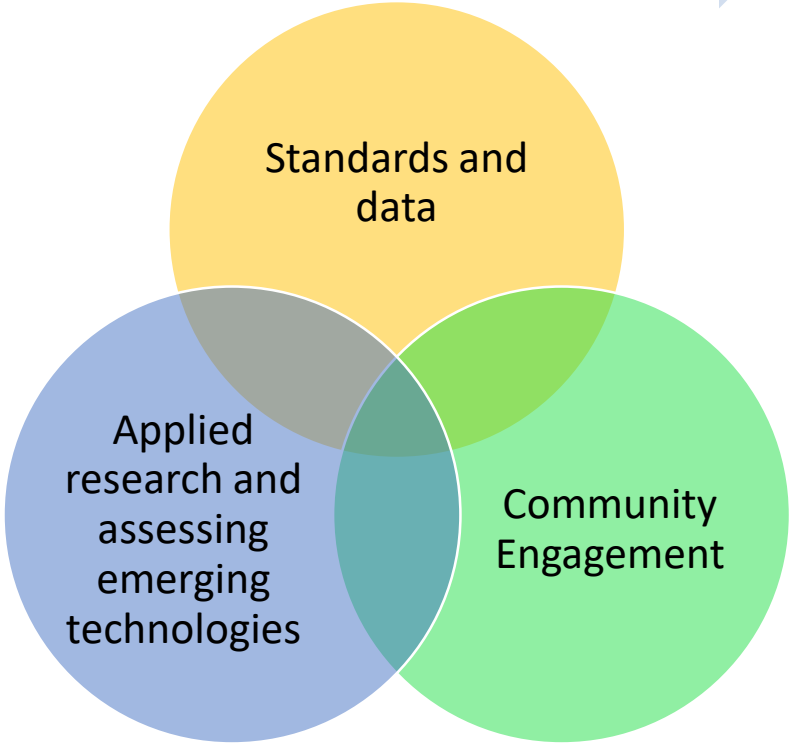
- All work presented has been reviewed and approved by the NIST Research Protections Office.

Forensic Genetics – Forensic DNA Typing Workflow

DNA Extraction → DNA Quantitation → PCR Amplification → Genotyping & Sequencing → Interpretation

Advancing technology and traceability through quality genetic measurements to aid work in Forensic and Clinical Genetics.

Variations upon the polymerase chain reaction (PCR) technique such as **rapid PCR, multiplex PCR, real-time PCR, and digital PCR** are used to **genotype, sequence, and provide quantitative information** pertaining to an organism's genome.



Forensic Genetics Team


Peter Vallone


Becky Steffen


Erica Romsos


Katherine Gettings


Kevin Kiesler


Margaret Kline


Lisa Borsuk


Sarah Riman


David Duewer


Hari Iyer

Statistical Support

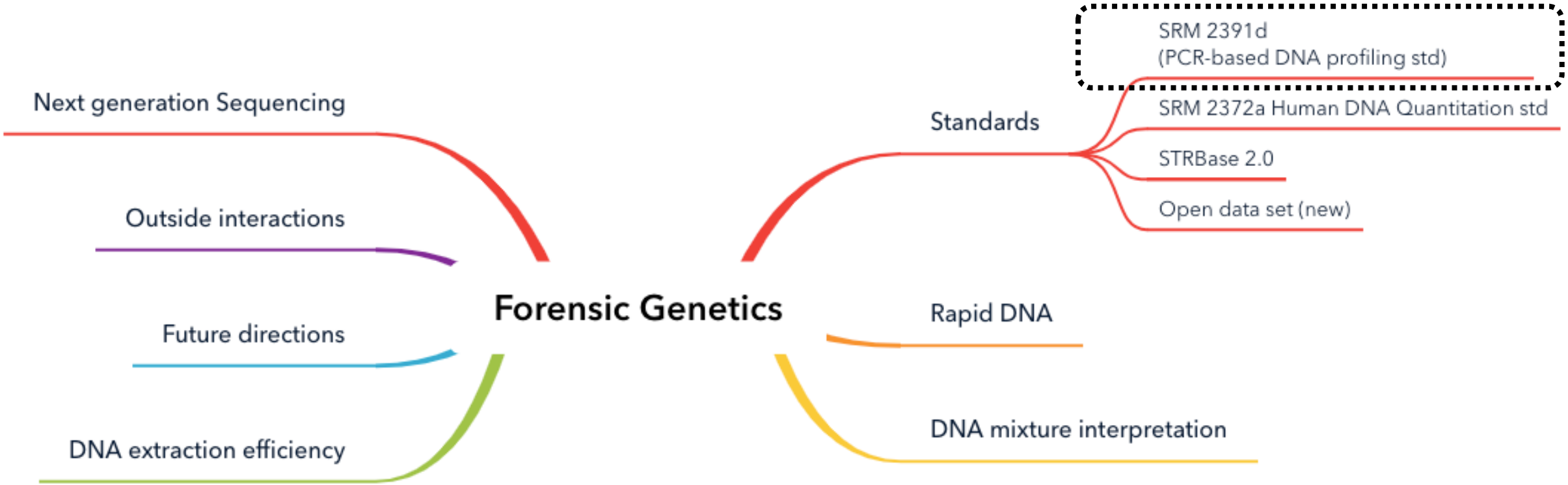

Tunde Huszar PostDoc

Margaret Kline will be retiring on November 20, 2020
Congratulations Margaret on a 35-year career at NIST
We'll miss you!

DNA Workshop Agenda
November 09, 2020

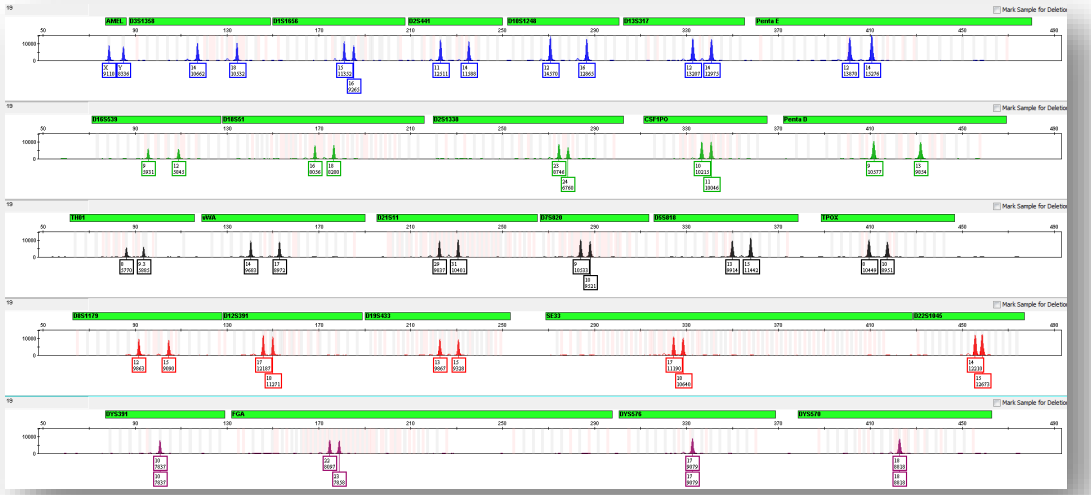
Time Slot	Title	Speaker	Time (mins)
10 – 10:15 am	Welcome and Introduction to the Applied Genetics Group	Peter Vallone	15
10:20 – 10:50 am	Not your standard Standard: Using SRM 2391d: PCR-Based DNA Profiling Standard in Your Lab	Becky Steffen	30
10:55 – 11:25 am	Making the best use of all of your curves: SRM 2372a Human DNA Quantitation Standard	Erica Romsos	30
11:30 – 12:00 pm	A Tour of STRBase 2.0	Lisa Borsuk	30
12:00 – 1:00 pm	Lunch Break		60
1:00 – 1:30 pm	Exploring DNA Interpretation Software Using the PROVEDIt Dataset	Sarah Riman	30
1:35 – 2:05 pm	Sequencing Workflows for Forensics	Peter Vallone	30
2:10 – 2:40 pm	SNPchat: the forensic marker that could be your BFF	Katherine Gettings	30
2:45 – 3:15 pm	Mitochondrial DNA Sequencing: the Next Generation	Kevin Kiesler	30
3:20 – 4:00 pm	AGG Panel: Q & A	all	40

Topics for today



SRM 2391 series

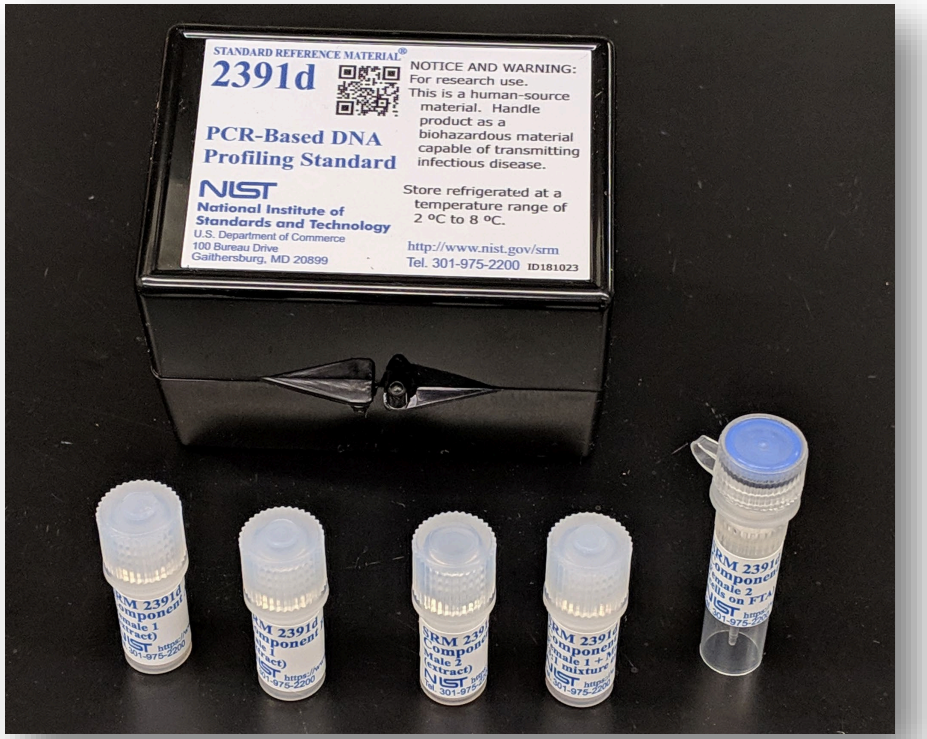
- SRM 2391c and 2391d – PCR-based DNA standard
- Used by the community to calibrate STR typing methods
- Purchased by practitioners and companies performing validations
- Further characterization of the samples and updates to the certificate



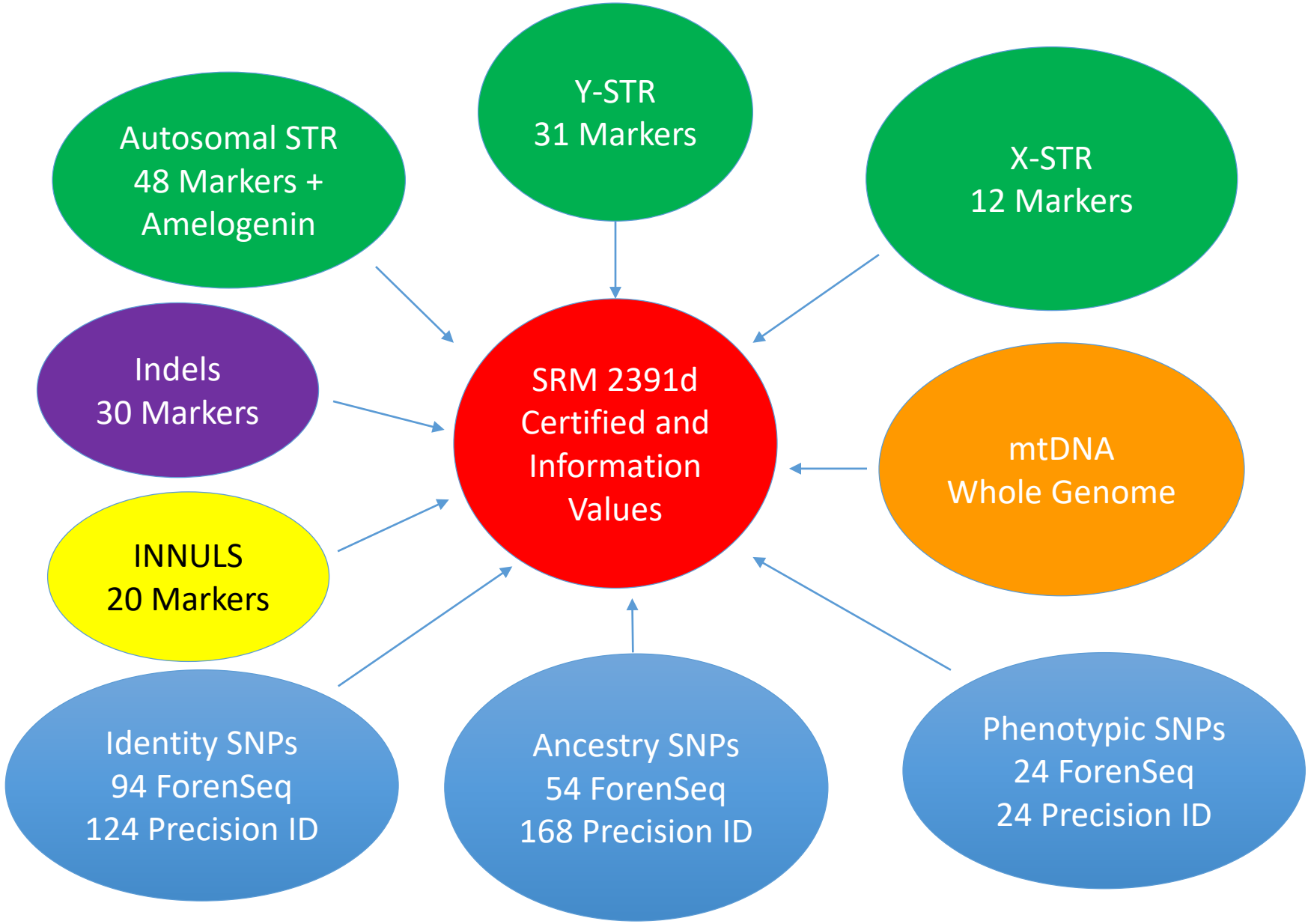
STR profile – certified allele calls

Table 1. Description of Components in SRM 2391d

Component	Description	Volume	Concentration[a]
A	Anonymous single-source female genomic DNA in TE^{-4} buffer	55 µL	1.6 ± 0.5 ng/µL
B	Anonymous single-source male genomic DNA in TE^{-4} buffer	55 µL	1.7 ± 0.5 ng/µL
C	Anonymous single-source male genomic DNA in TE^{-4} buffer	55 µL	1.6 ± 0.2 ng/µL
D	Mixed-source, 3:1 (3 parts Component A and 1 part Component C) genomic DNA in TE^{-4} buffer	55 µL	1.5 ± 0.4 ng/µL
E	Anonymous single-source female cells spotted on FTA paper[b]	Two 6 mm punches	7.5×10^4 cells per punch

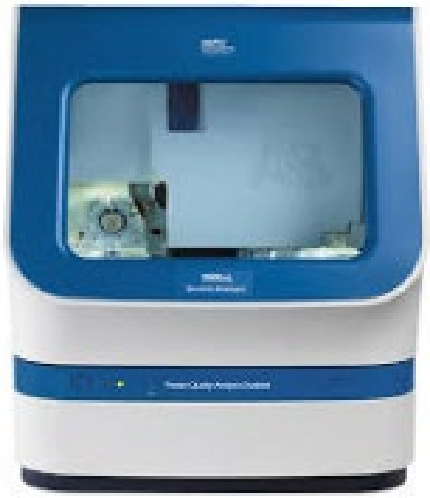


Markers included in the Certificate of Analysis



Platforms used for characterization

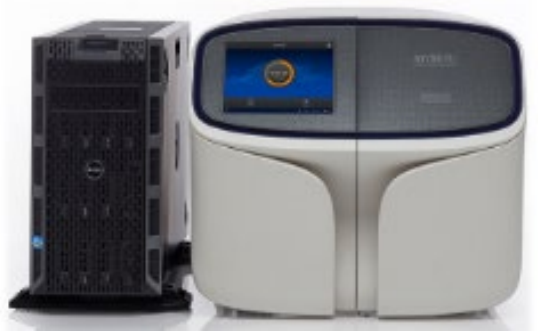
- **Capillary Electrophoresis (CE)** was performed with one instrument:
  - 3500xL Genetic Analyzer (Thermo Fisher)



3500xl

- **Next Generation Sequencing (NGS)** was performed with two different instruments:
  - MiSeq FGx (Verogen)
  - Ion S5 XL (Thermo Fisher)

MiSeq FGx



Ion S5 XL

Autosomal STR Markers

Materials characterized using various commercial kits

Autosomal STR Markers
Thermo Fisher CE STR kits
Promega CE STR kits
Qiagen Investigator CE STR kits
Verogen NGS kit
Thermo Fisher NGS kits
Promega NGS kits
CODIS 20/ESS 12

35 Certified Autosomal STR Markers
13 Information Autosomal STR Markers

Autosomal STR Marker List	MiniFiler	Identifiler	Identifiler Plus	Identifiler Direct	NGM	NGM SElect	NGM Detect	Verifiler Plus	Verifiler Express	GlobalFiler	GlobalFiler Express	PP S5	PP CS7	PP 16	PP 16 HS	PP 18D	PP 21	PP ESX 17	PP ESX 17 Fast	PP ESI 17 Pro	PP ESI 17 Fast	PP Fusion	PP Fusion 6C	PP VersaPlex 27PY	ESSplex SE Plus	HDplex	24plex GO!	24plex QS	ForenSeq	Precision ID GF	PowerSeq 46GY	CODIS 20	European Standard Set	Certified Value	Information Value
D1S1656																																	X	
D1S1677																																	X	
D2S1338																																	X	
D2S441																																	X	
D2S1360																																		X
D2S1776																																	X	
D3S1358																																	X	
D3S1744																																		X
D3S4529																																	X	
D4S2366																																		X
D4S2408																																	X	
D5S818																																	X	
D5S2500																																		X
D5S2800																																	X	
D6S474																																	X	
D6S1043																																	X	
D7S820																																	X	
D7S1517																																		X
D8S1132																																		X
D8S1179																																	X	
D9S1122																																	X	
D10S1248																																	X	
D10S2325																																		X
D12S391																																	X	
D12ATA63																																	X	
D13S317																																	X	
D14S1434																																	X	
D16S539																																	X	
D17S1301																																	X	
D18S51																																	X	
D19S433																																	X	
D20S482																																	X	
D21S11																																	X	
D21S2055																																		X
D22S1045																																	X	
CSF1PO																																	X	
F13A01																																		X
F13B																																		X
FESFPS																																		X
FGA																																	X	
LPL																																		X
Penta C																																		X
Penta D																																	X	
Penta E																																	X	
SE33																																	X	
TH01																																	X	
TPOX																																	X	
vWA																																	X	

Autosomal STR Markers

Typed by CE methods *NGS methods*

Autosomal STR Marker List	MiniFiler	Identifiler	Identifiler Plus	Identifiler Direct	NGM	NGM SElect	NGM Detect	Verifiler Plus	Verifiler Express	GlobalFiler	GlobalFiler Express	PP S5	PP CS7	PP 16	PP 16 HS	PP 18D	PP 21	PP ESX 17	PP ESX 17 Fast	PP ESI 17 Pro	PP ESI 17 Fast	PP Fusion	PP Fusion 6C	PP VersaPlex 27PY	ESSplex SE Plus	HDplex	24plex GO!	24plex QS	ForenSeq	Precision ID GF	PowerSeq 46GY	CODIS 20	European Standard Set	Certified Value	Information Value
D1S1656																																		X	
D1S1677																																		X	
D2S1338																																		X	
D2S441																																		X	
D2S1360																																			X

35 Certified Autosomal STR Markers
13 Information Autosomal STR Markers

Y-STR Markers

Y-STR Markers
Thermo Fisher CE STR kits
Promega CE STR kits
Qiagen Investigator CE STR kits
Verogen NGS kit
Thermo Fisher NGS kits
Promega NGS kits

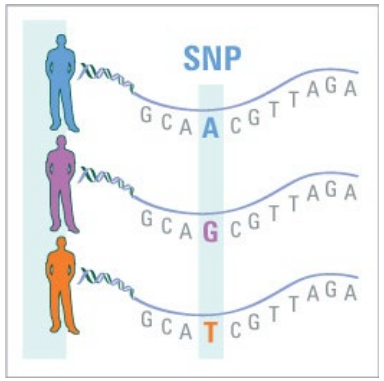
28 Certified Y-STR Markers
3 Information Y-STR Markers

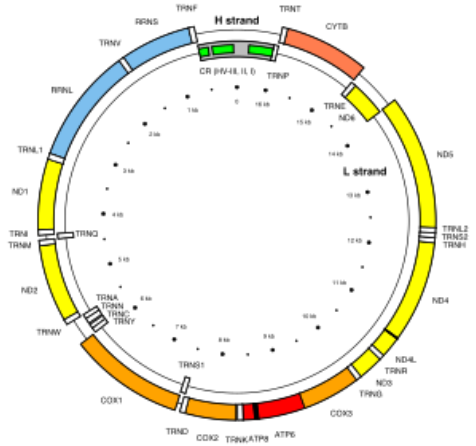
Y-STR Marker List	GlobalFiler	GlobalFiler Express	Yfiler	Yfiler Plus	PP Fusion	PP Fusion 6C	PP VersaPlex 27PY	PowerPlex Y23	24plex GO!	24plex QS	ForenSeq	Precision ID GF	PowerSeq 46GY	Certified Value	Information Value
DYS19														X	
DYS385a/b														X	
DYS389I/II														X	
DYS390														X	
DYS391														X	
DYS392														X	
DYS393														X	
DYS437														X	
DYS438														X	
DYS439														X	
DYS448														X	
DYS449															X
DYS456														X	
DYS458														X	
DYS460														X	
DYS461														X	
DYS481														X	
DYS505														X	
DYS518															X
DYS522														X	
DYS533														X	
DYS549														X	
DYS570														X	
DYS576														X	
DYS612														X	
DYS627															X
DYS635														X	
DYS643														X	
Y-GATA-H4														X	
DYS387S1														X	

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





SNP allele calls for all components

Identity SNP markers

- 101 autosomal SNPs reported
  - ForenSeq (94)
  - Precision ID Identity Panel (90) + 34 Y-SNPs
- Forward strand genotype reported

83 identity autosomal SNPs in common

Ancestry and Phenotype SNP markers

- Ancestry/Phenotype SNPs 188 total
  - ForenSeq (78)
  - Precision ID Ancestry (165) and Phenotype Panel (24)
- Forward strand genotype reported

77 SNPs in common

Component	ForenSeq			Precision ID				Mito	Y SNP
	Ancestry	*Hair*	*Eye*	*Ancestry*	*Hair*	*Hair*	*Eye*		
A	European	0.68	0.66	European	0.66	1.00 light	0.67	T2b3	-
B	African	0.69	0.86	African	0.66	0.93 light	0.85	L1c1a	E
C	African	0.84	1.00	African	0.68	1.00 dark	1.00	L1b1a	E
E	European	0.61	0.71	European/SW Asian	0.69	0.72 light	0.72	T2a3	-

Predictions made using vendor tools for autosomal SNP markers

How can SRM 2391d be used in YOUR lab?

- To meet the FBI Quality Assurance Standards: QAS 8.4

> STANDARD 8.4 Newly validated DNA methods (from amplification through characterization), typing test kit, or platform instrument model shall be checked against an appropriate and available certified reference material (or sample made traceable to the certified reference material) prior to the implementation of the method for forensic analysis.

- Validation Studies: instrument, commercial kit, and software
  - Developmental and Internal Validations
  - Known, ***well-characterized*** samples for forensic marker systems

- Make NIST traceable materials: http://ts.nist.gov/traceability/

Establishing Traceability to NIST SRM 2391d

- Traceability requires the establishment of an unbroken chain of comparisons to stated references: http://ts.nist.gov/traceability/
- In the case of DNA testing with STR markers, the reference material is SRM 2391d
- Materials deemed traceable to NIST-created materials must have a record associated with them

'Lot' of DNA material

Run against SRM 2391d

Everyday use calibrant

Contact becky.steffen@nist.gov for traceability questions

Notes for SRM 2391c and Mitochondrial SRMs

- For those still using SRM 2391c (no longer being sold) the certificate expiration date has been extended through February 3, 2022
- This will be the final extension, after that SRM 2391d must be used
- SRMs 2392 and 2392-I (mitochondrial DNA sequencing) will not be replaced – use SRM 2391d

National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2391c

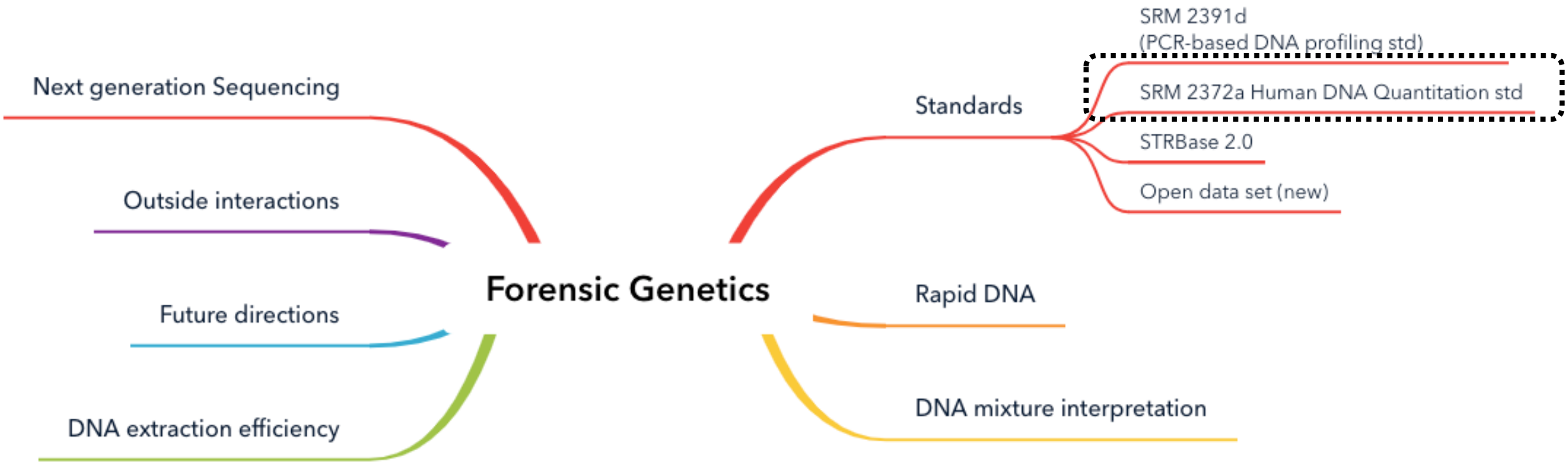
PCR-Based DNA Profiling Standard

Expiration of Certification: The certification of **SRM 2391c** is valid, within the measurement uncertainties specified until **03 February 2022**, provided the SRM is handled and stored in accordance with the instructions given in this certification (see "Instructions for Use"). The certification is nullified if the SRM is damaged, contaminated, or otherwise modified.

Details	
Description:	Mitochondrial DNA Sequencing (Human HL-60 DNA)
Lot:	N/A
Expiration Date:	3/31/2023
Unit Price * :	$622.00
Unit of Issue:	each
Status:	Now Selling See 'Additional Information' for details.
Certificate Date:	2/2/2018
Certificate Revision Date:	02 February 2018 (Change of certification period; editorial changes).
MSDS Date:	2/2/2018
Technical Contact:	Peter Vallone
Additional Information:	SRM 2392-I will not be replaced when the current stock is depleted. SRM 2391d PCR-Based DNA Profiling Standard now contains mitochondrial sequence information and should be considered as a replacement.

SRM 2392-I will not be replaced when the current stock is depleted. SRM 2391d PCR-Based DNA Profiling Standard now contains mitochondrial sequence information and should be considered as a replacement.

Topics for today



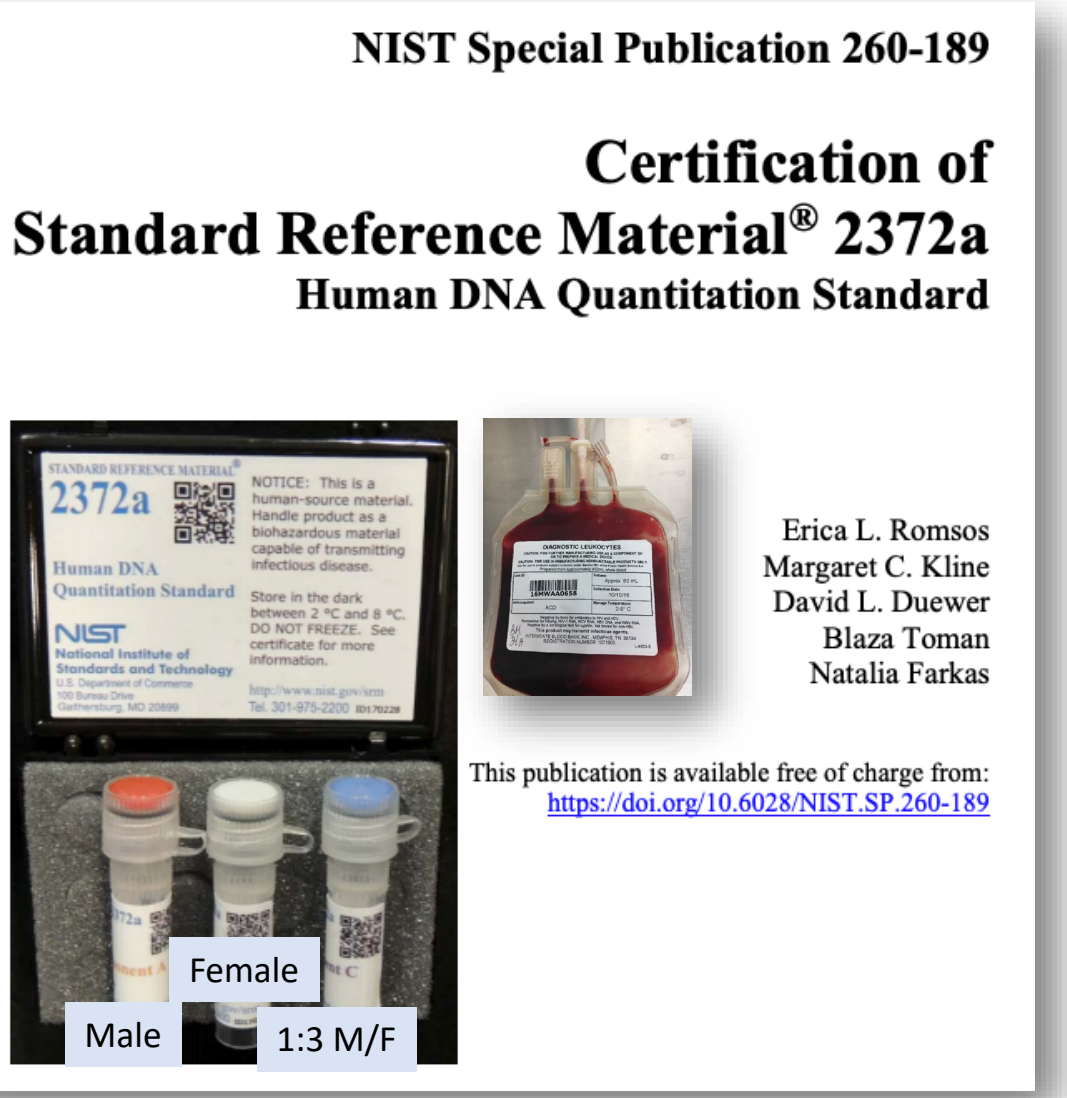
SRM 2372a (Human DNA Quantitation Std)

- SRM 2372a – Human DNA Quantitation Standard
- Used to calibrate qPCR methods and commercial DNA standards
- Digital PCR used for value assignment
- Mitochondrial to nuclear DNA ratio information included

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.

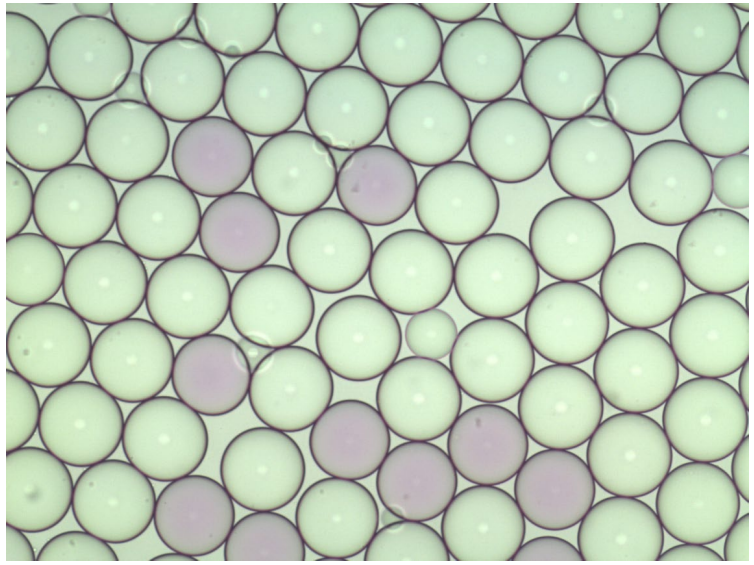
Component	Copy Number[b] (per nL)	DNA[c] (ng/µL)
A (red cap)	15.1 ± 1.5	49.8 ± 5.0
B (white cap)	17.5 ± 1.8	57.8 ± 5.8
C (blue cap)	14.5 ± 1.5	47.9 ± 4.8

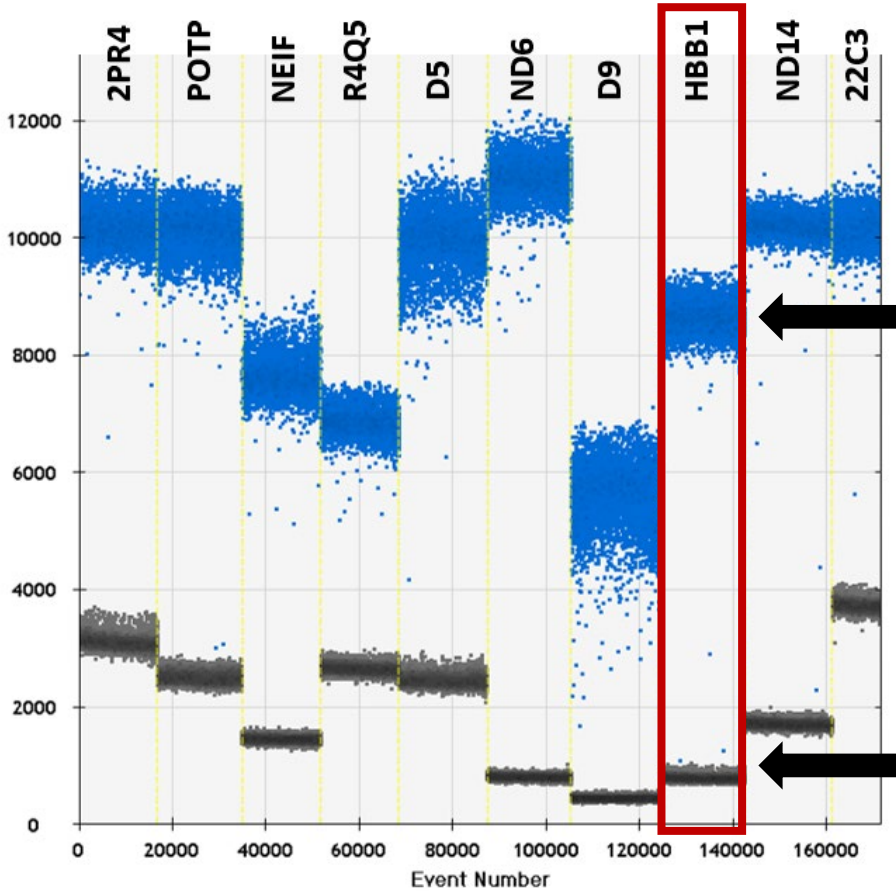


https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.260-189.pdf

Digital PCR

Partitioning of DNA targets into individual chambers or droplets





A standard curve is not needed

Positives
(1 copy of the target)

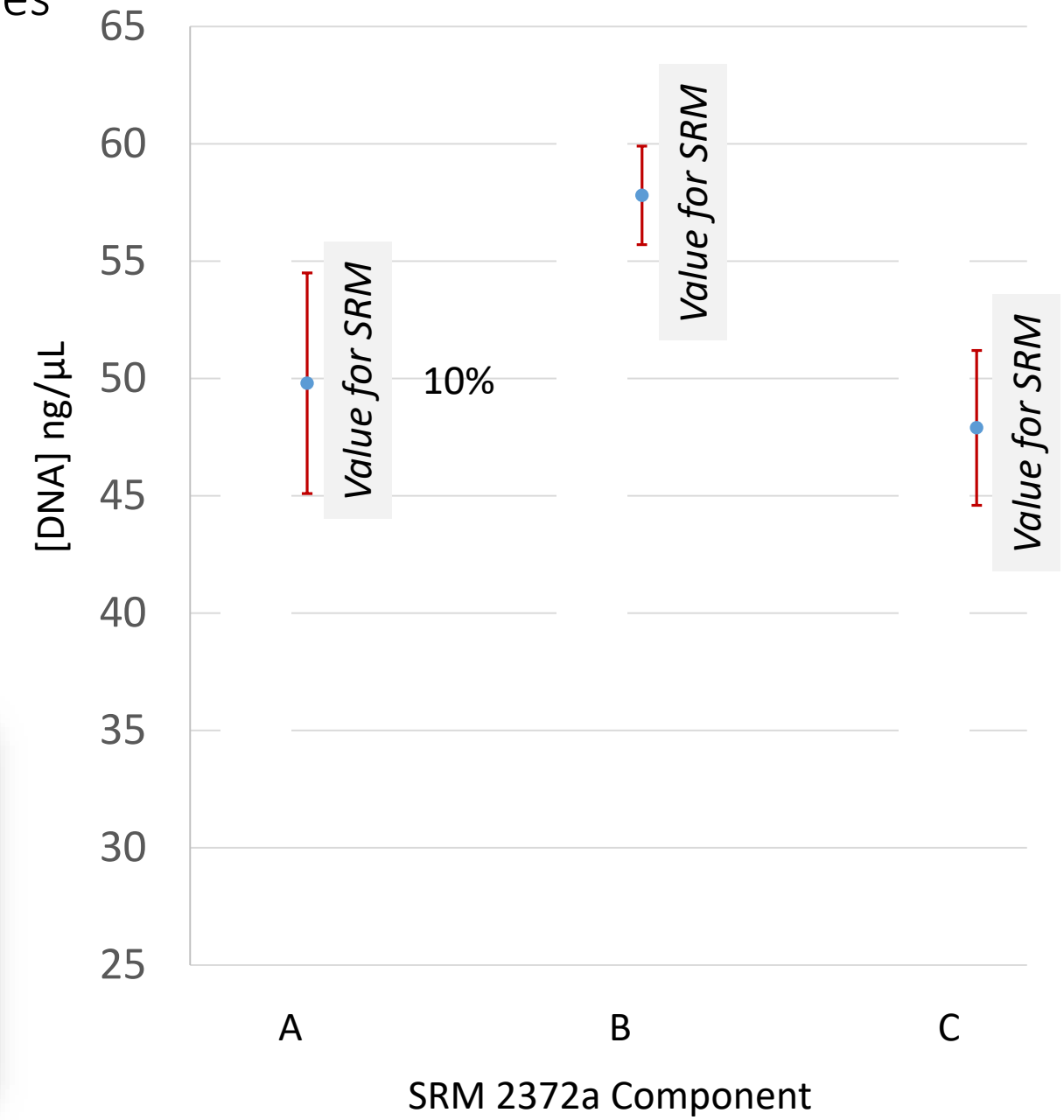
Negatives
(0 copies of the target)

dPCR is counting *accessible* amplifiable targets

Assigned concentration values

These are the assigned values (with error) for the Components in SRM 2372a

Value and uncertainty based on ten unique digital PCR assays



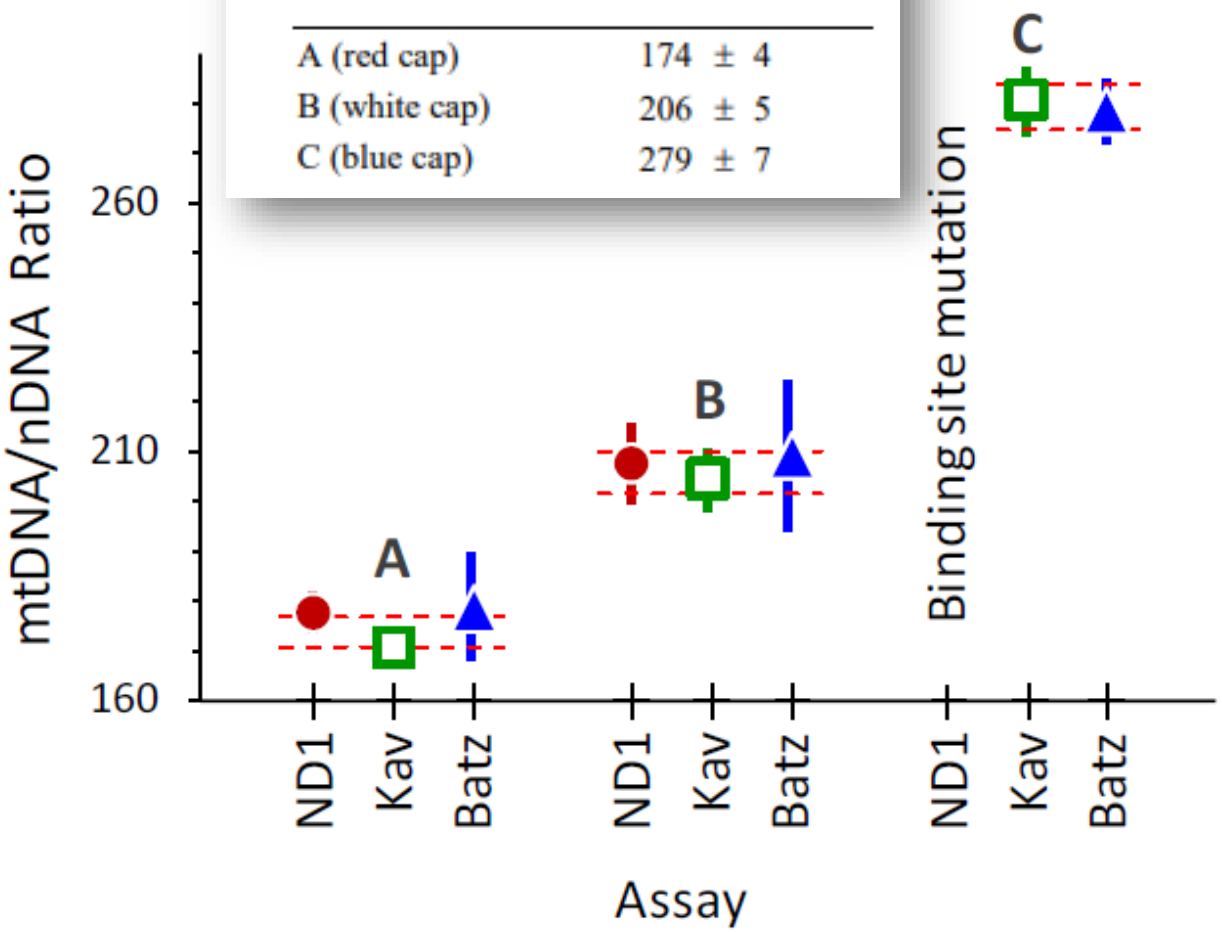
Published in final edited form as:
Anal Bioanal Chem. 2018 May ; 410(12): 2879–2887. doi:10.1007/s00216-018-0982-1.

Evaluating droplet digital PCR for the quantification of human genomic DNA: converting copies per nanoliter to nanograms nuclear DNA per microliter

David L. Duewer[1], **Margaret C. Kline**[2], **Erica L. Romsos**[2], and **Blaza Toman**[3]
[1]Chemical Sciences Division, Material Measurement Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8390, Gaithersburg, MD 20899-8390, USA

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

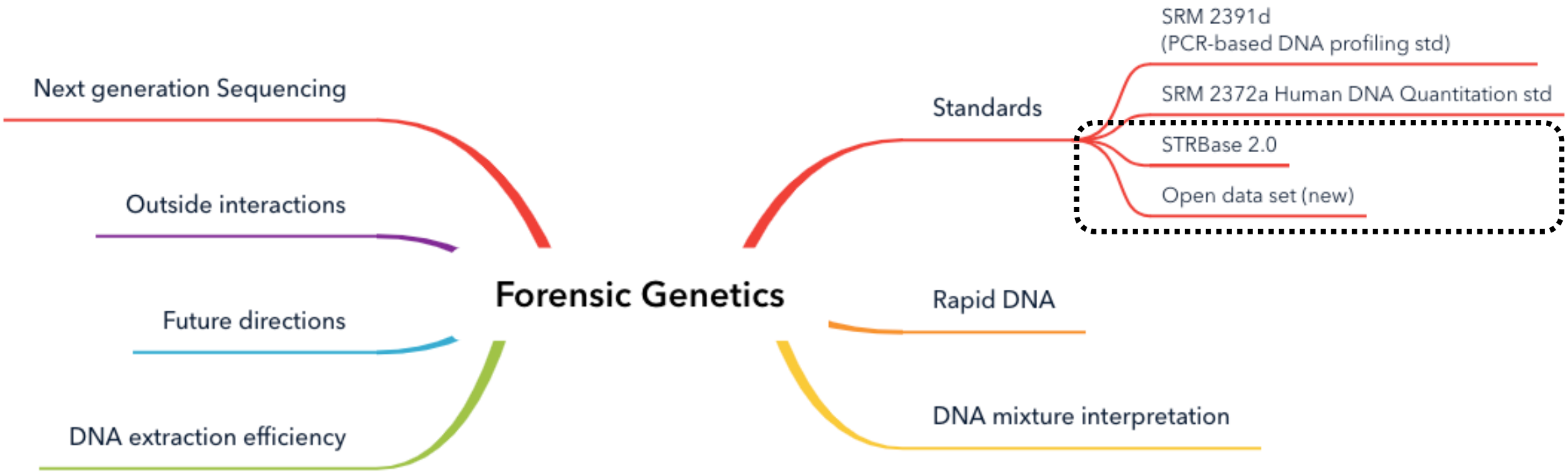
Component	mtDNA/nDNA
A (red cap)	174 ± 4
B (white cap)	206 ± 5
C (blue cap)	279 ± 7



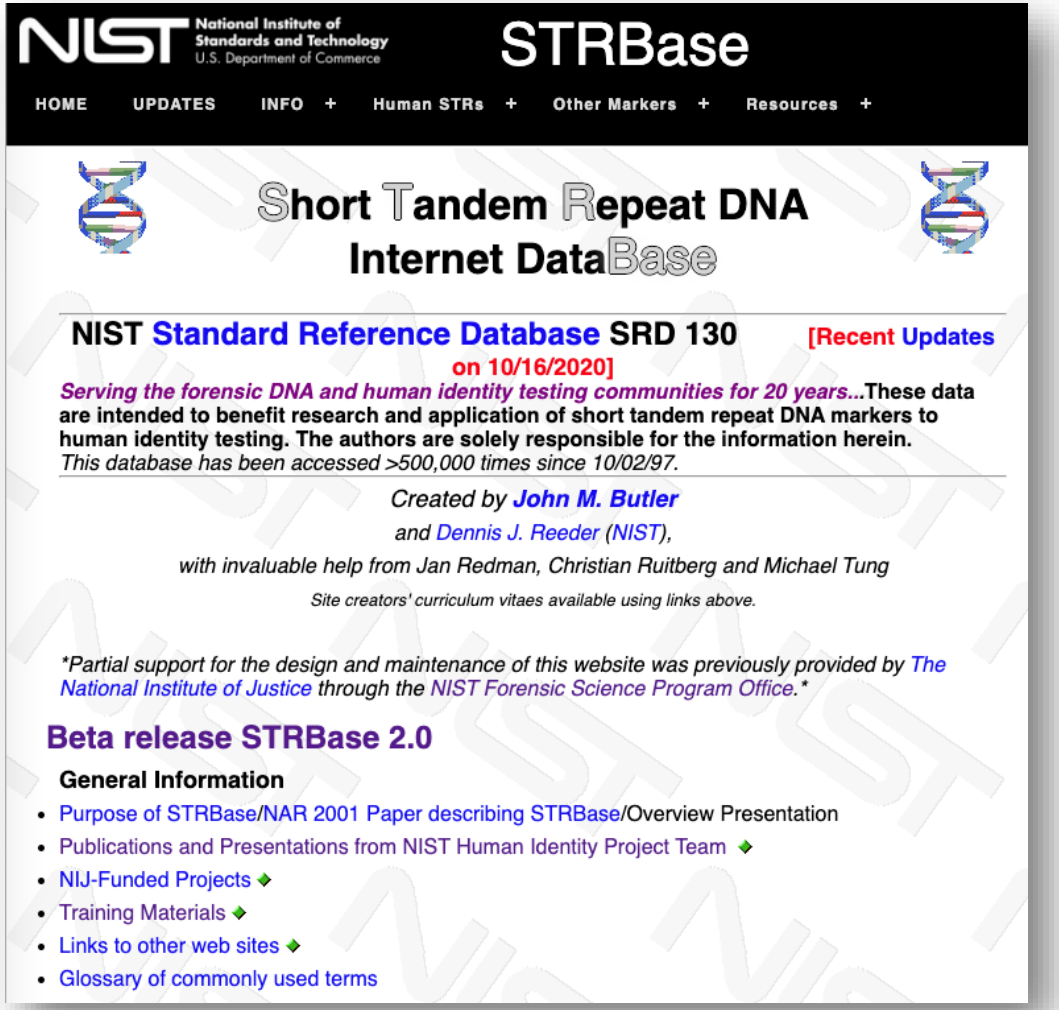
mtDNA/nDNA ratio for three mitochondrial quantification assays optimized for dPCR

Developing Y chromosome specific digital PCR assays

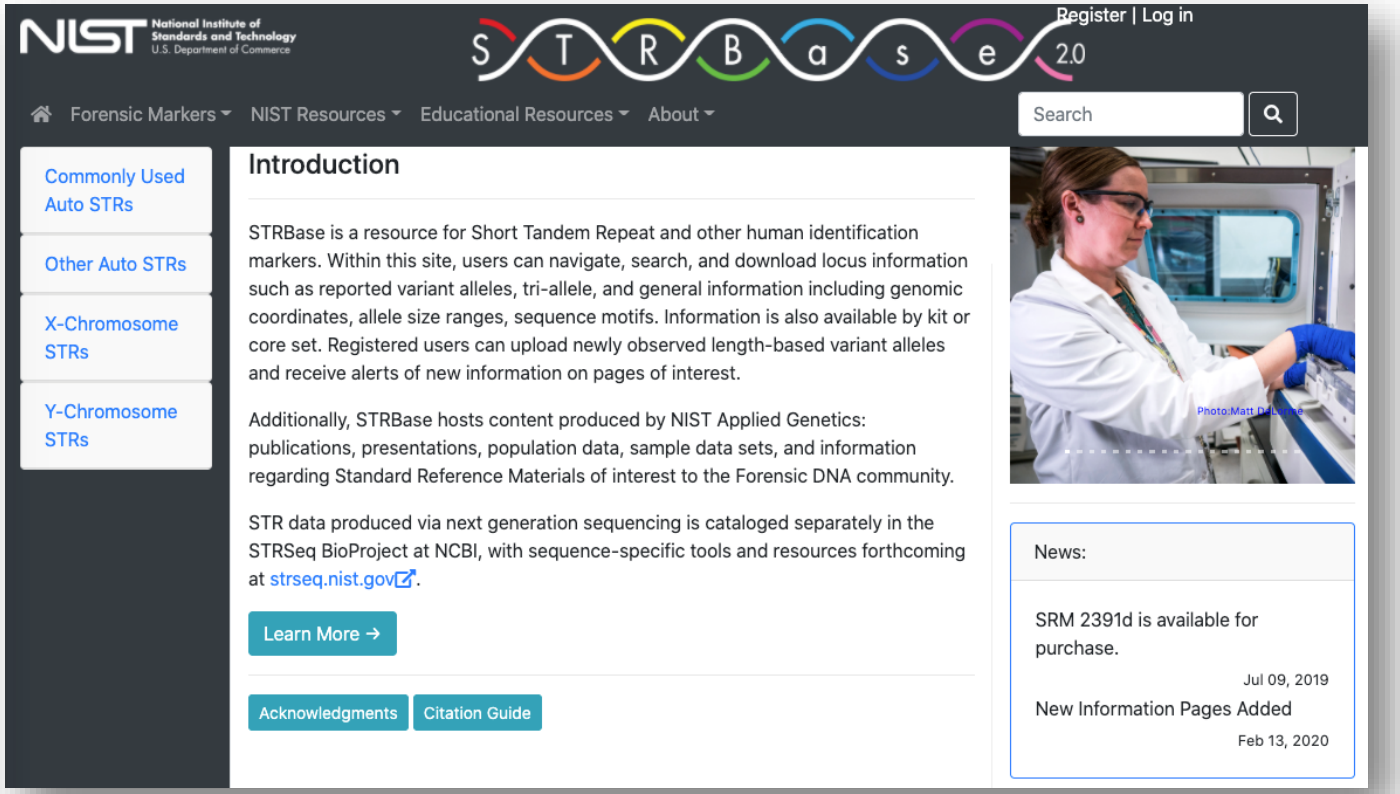
Topics for today



STRBase and STRBase 2.0

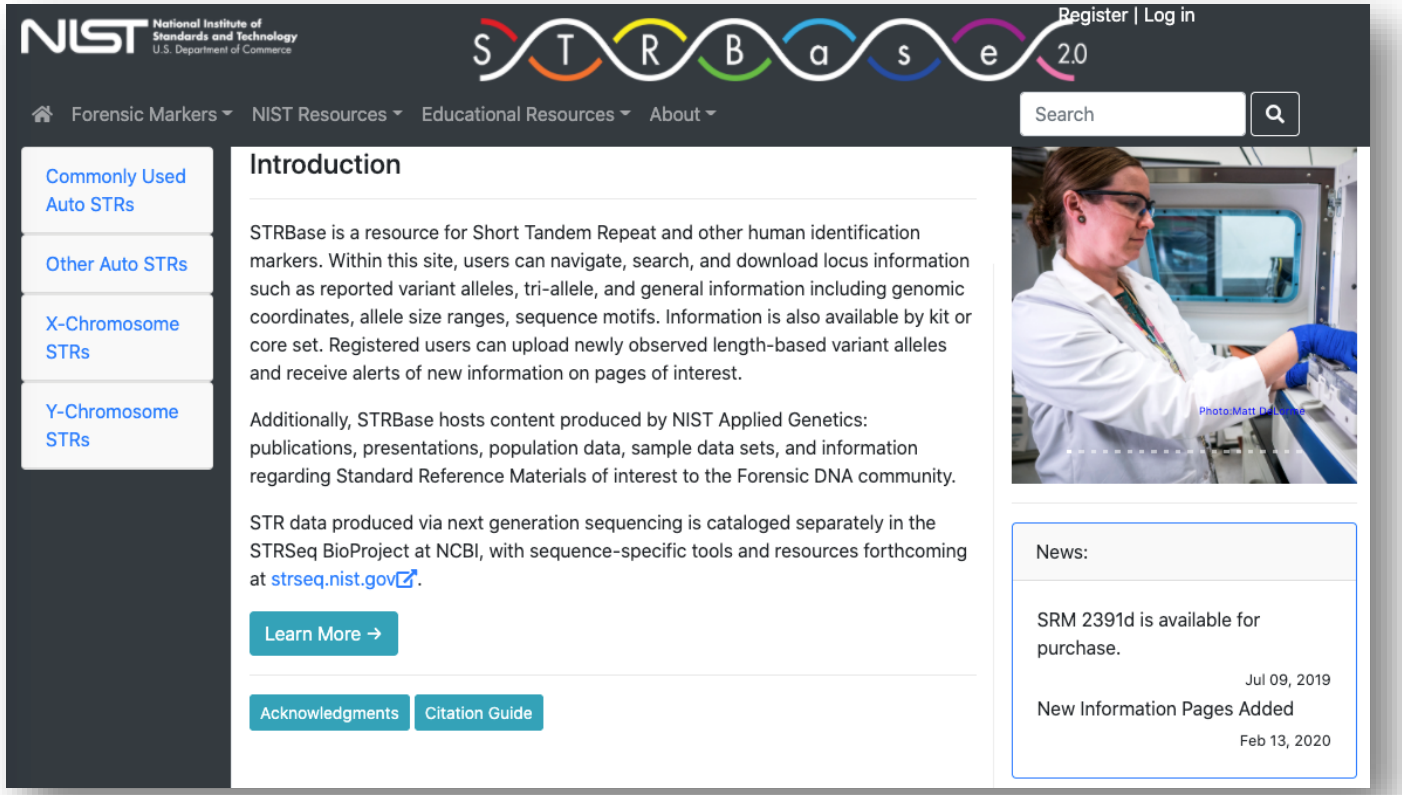


https://strbase.nist.gov/



https://strbase-b.nist.gov/

STRBase 2.0



https://strbase-b.nist.gov/

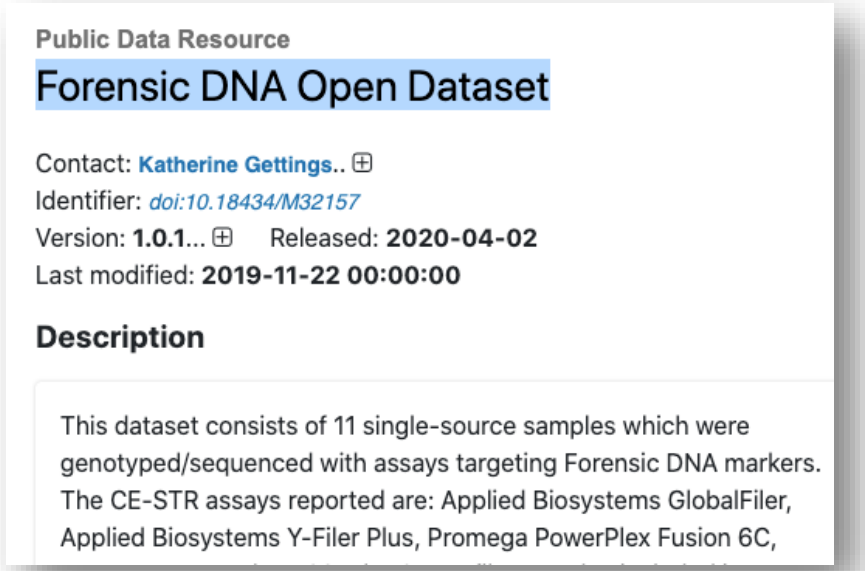
Updated navigation of STR fact sheets and variant alleles

Updated topic pages for:

- SNP markers
- Mitochondrial DNA
- Insertion-Deletion markers
- Forensic SRMs
- NIST Interlaboratory studies
- Population data
- Educational resources

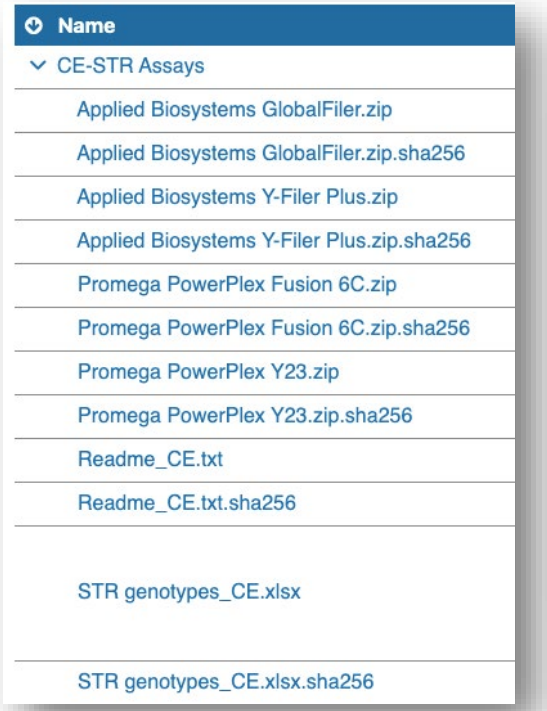
- Please visit and give us any feedback: strbase@nist.gov

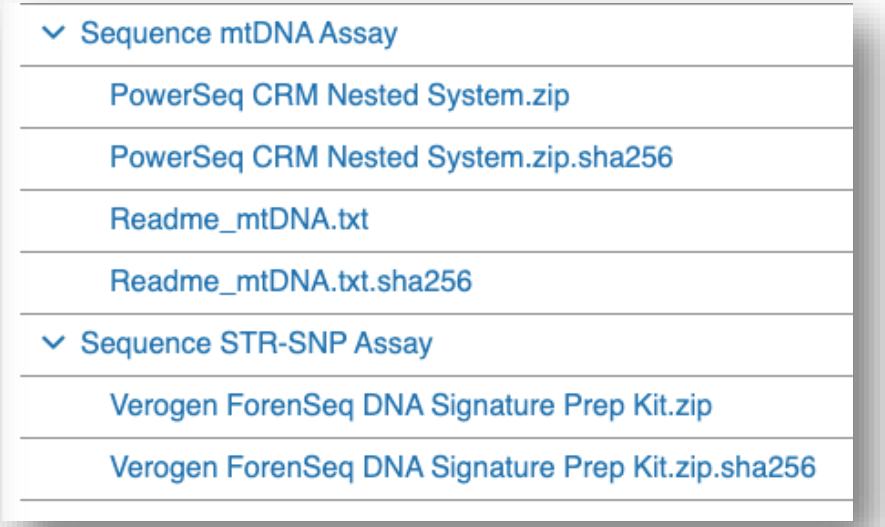
Forensic DNA Open Dataset



https://data.nist.gov/od/id/mds2-2157

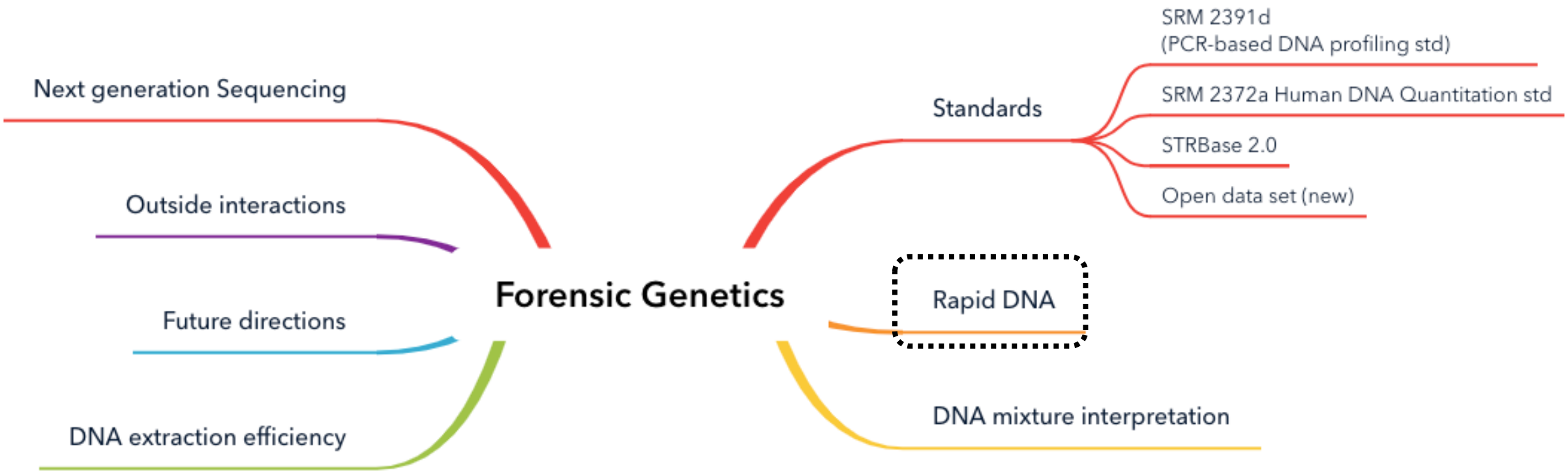
- **Periodic requests for raw CE (.HID) or sequencing files (.FASTQ)**
- From academics, researchers, software vendors, etc
- Goal: make data freely available for learning/training purposes
- Source: Eleven single source samples previously screened as SRM candidates (not NIST population samples or SRM samples)



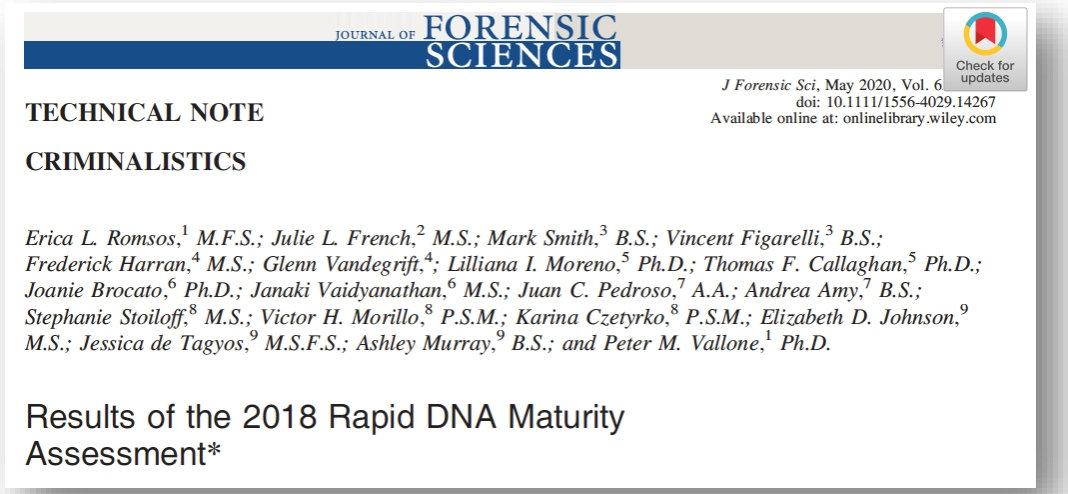


Email Katherine.Gettings@nist.gov
…for questions/ideas

Topics for today



Rapid DNA Maturity Assessment



JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, May 2020, Vol. 6
doi: 10.1111/1556-4029.14267
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

CRIMINALISTICS

Erica L. Romsos,[1] M.F.S.; Julie L. French,[2] M.S.; Mark Smith,[3] B.S.; Vincent Figarelli,[3] B.S.; Frederick Harran,[4] M.S.; Glenn Vandegrift,[4]; Lilliana I. Moreno,[5] Ph.D.; Thomas F. Callaghan,[5] Ph.D.; Joanie Brocato,[6] Ph.D.; Janaki Vaidyanathan,[6] M.S.; Juan C. Pedroso,[7] A.A.; Andrea Amy,[7] B.S.; Stephanie Stoiloff,[8] M.S.; Victor H. Morillo,[8] P.S.M.; Karina Czetyrko,[8] P.S.M.; Elizabeth D. Johnson,[9] M.S.; Jessica de Tagyos,[9] M.S.F.S.; Ashley Murray,[9] B.S.; and Peter M. Vallone,[1] Ph.D.

Results of the 2018 Rapid DNA Maturity Assessment*

TABLE 1—*Samples tested across nine participating agencies*

Instrument	Chemistry	Independent Instruments	Total Samples Tested	Analysis Method
ANDE 6C System	FlexPlex	5	100	Rapid DNA Analysis
RapidHIT 200	GlobalFiler Express	3	60	Modified Rapid DNA Analysis
RapidHIT ID	GlobalFiler Express	4	80	Modified Rapid DNA Analysis

What is "Rapid DNA"?

A fully automated instrument capable of generating a DNA profile from a swab ('swab in – profile out')

The maturity assessment was an interlaboratory study assessing DNA typing success and accuracy results on three Rapid DNA platforms

Twenty swabs were provided to each participant (crime labs, **police agencies**, **vendors**)

The results support the implementation of Rapid DNA instrumentation at the booking stations for single source samples

Stakeholders: FBI laboratory, DNA databasing labs, booking stations, Rapid DNA community, ANDE, Thermo Fisher

Assessment Scheme

Participant Runs RH-ID GFE

- Participants may choose one chemistry per ***20*** NIST provided swabs
- Additional packages may be requested

NIST provides **20** reference buccal swabs to each participant

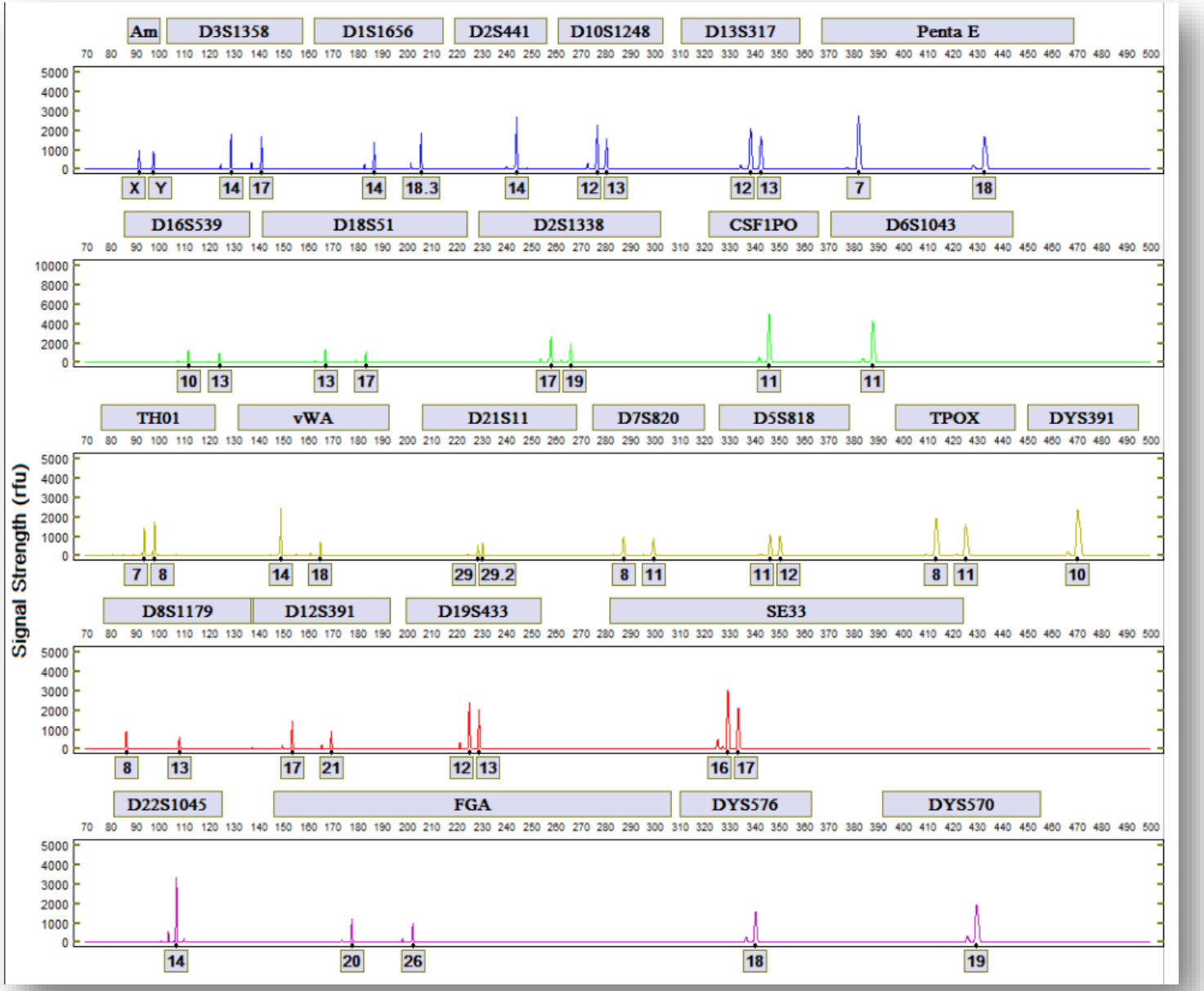
Participant Runs RH200 GFE

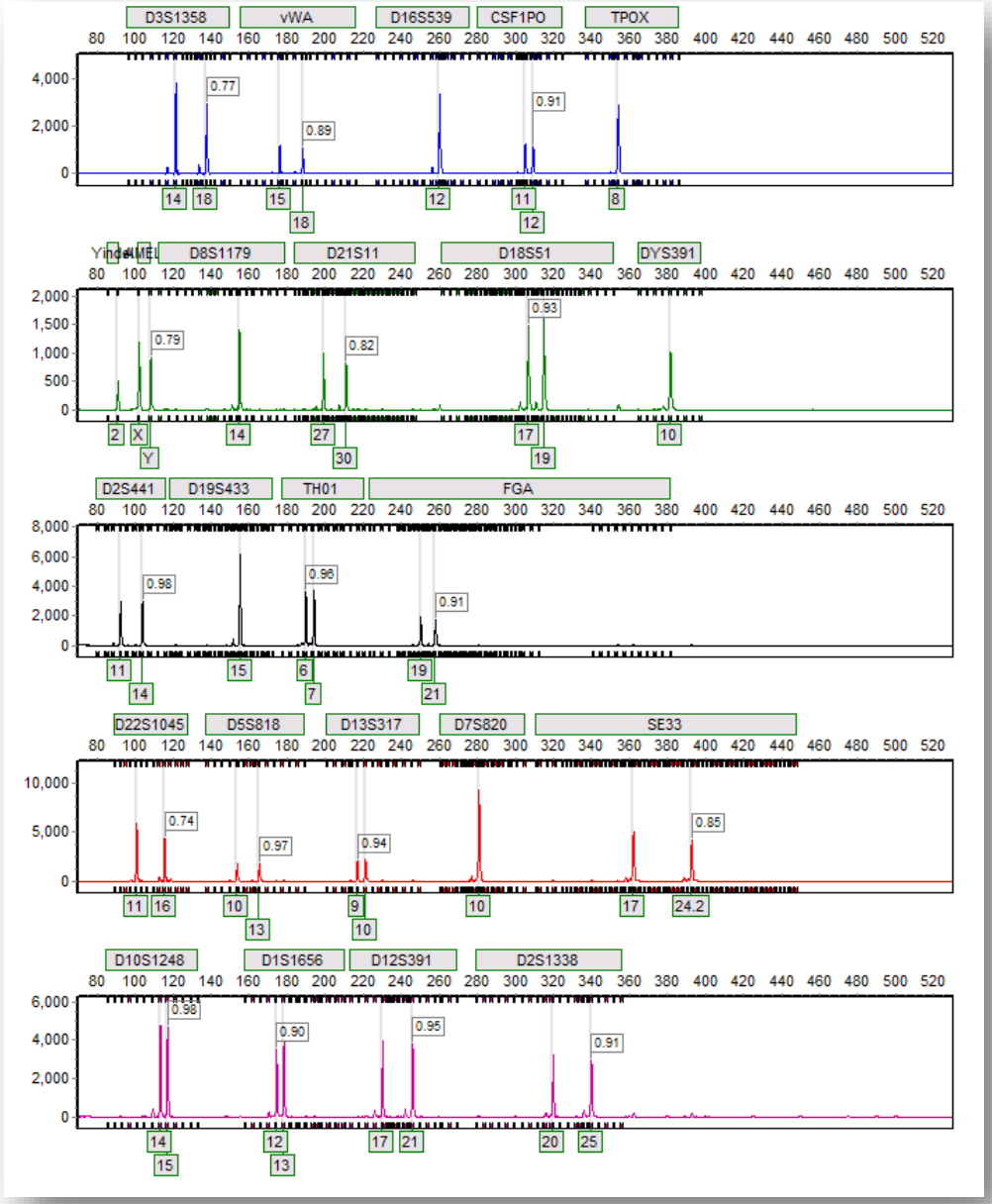
NIST reports CODIS 20 success rate for all data combined (% success)

Data transferred back to NIST via electronic format

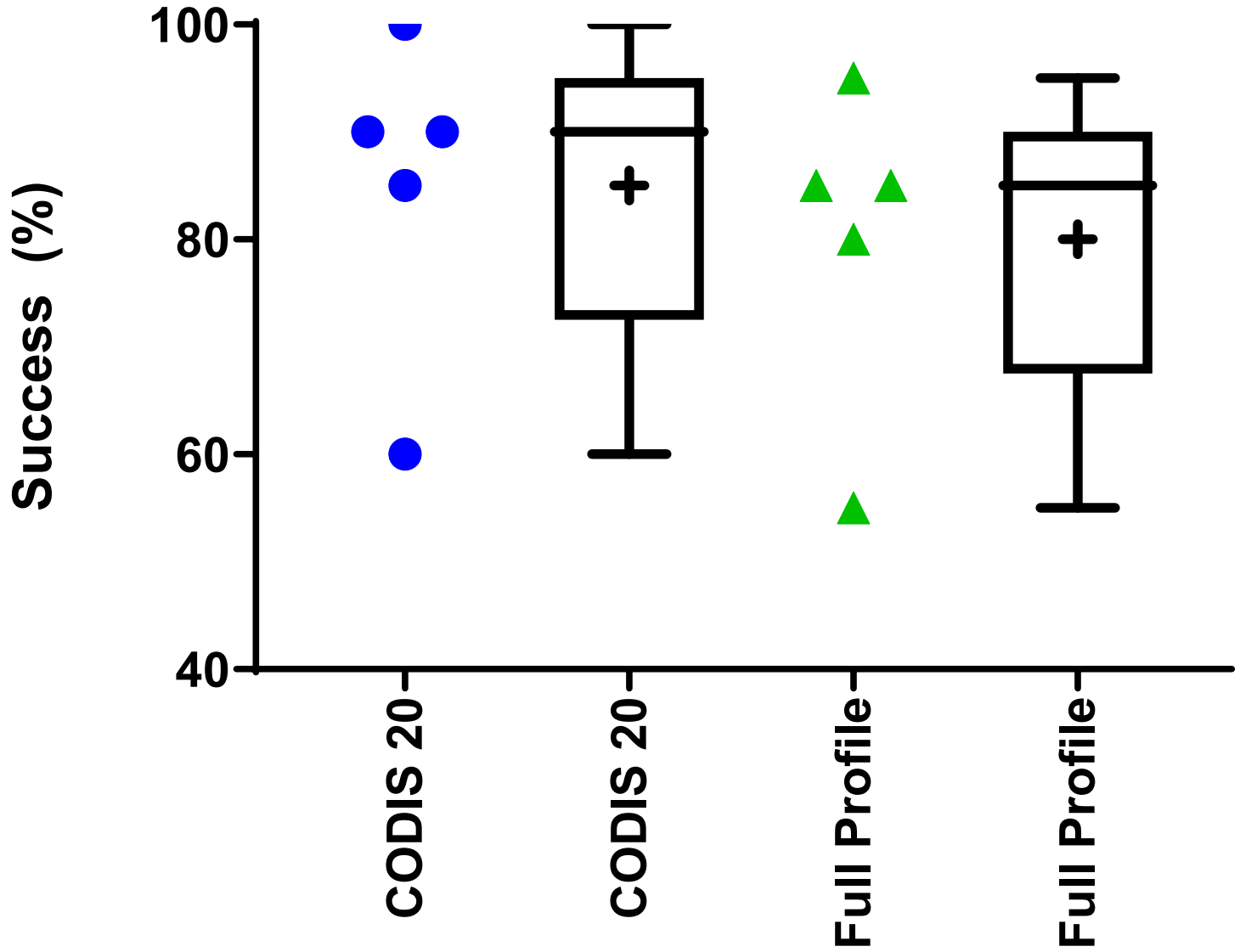
Participant Runs ANDE FlexPlex

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





Genotyping Success: Rapid DNA Analysis

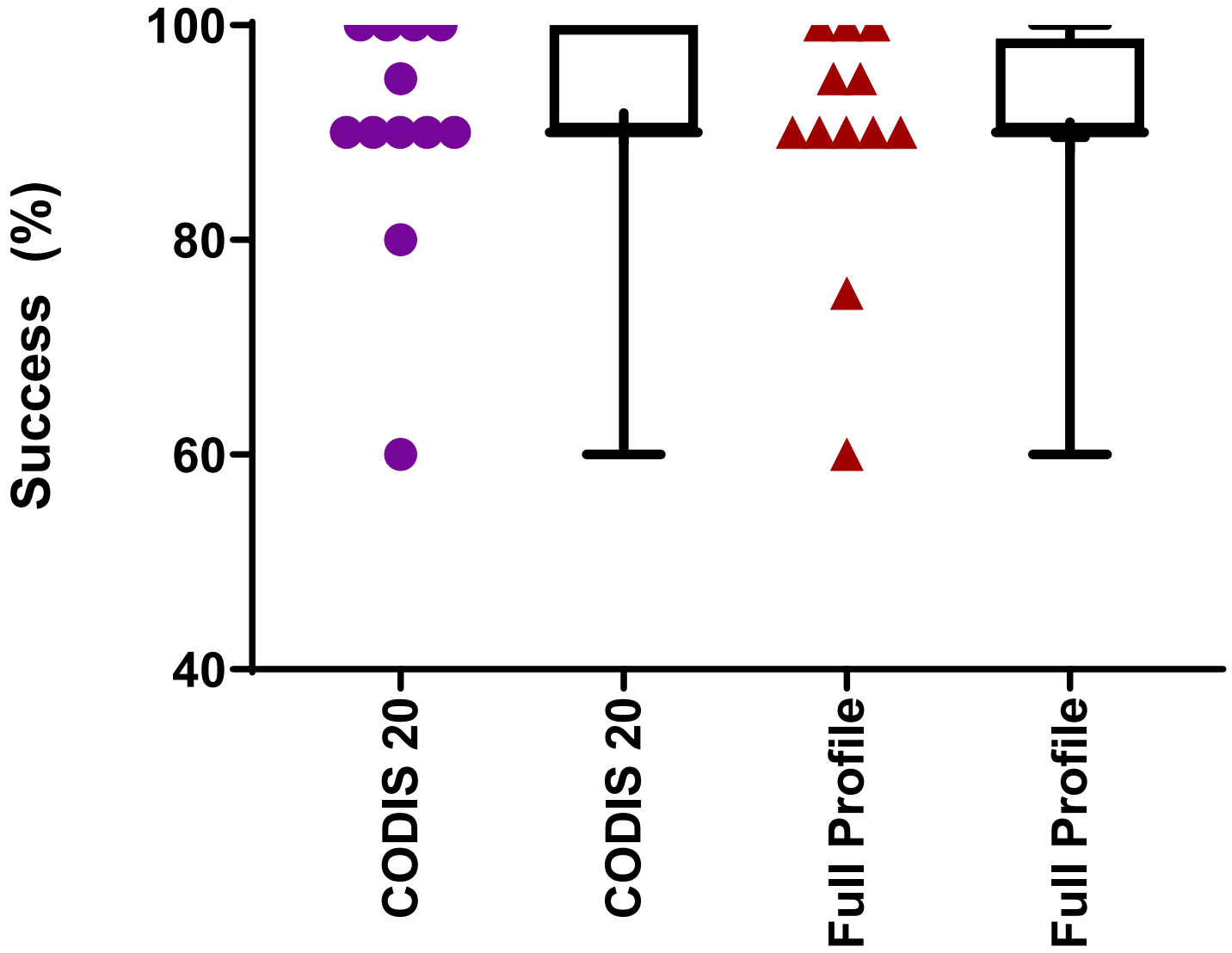


CODIS 20: 85%

Full Profile: 80%

Each data point is an independent instrument

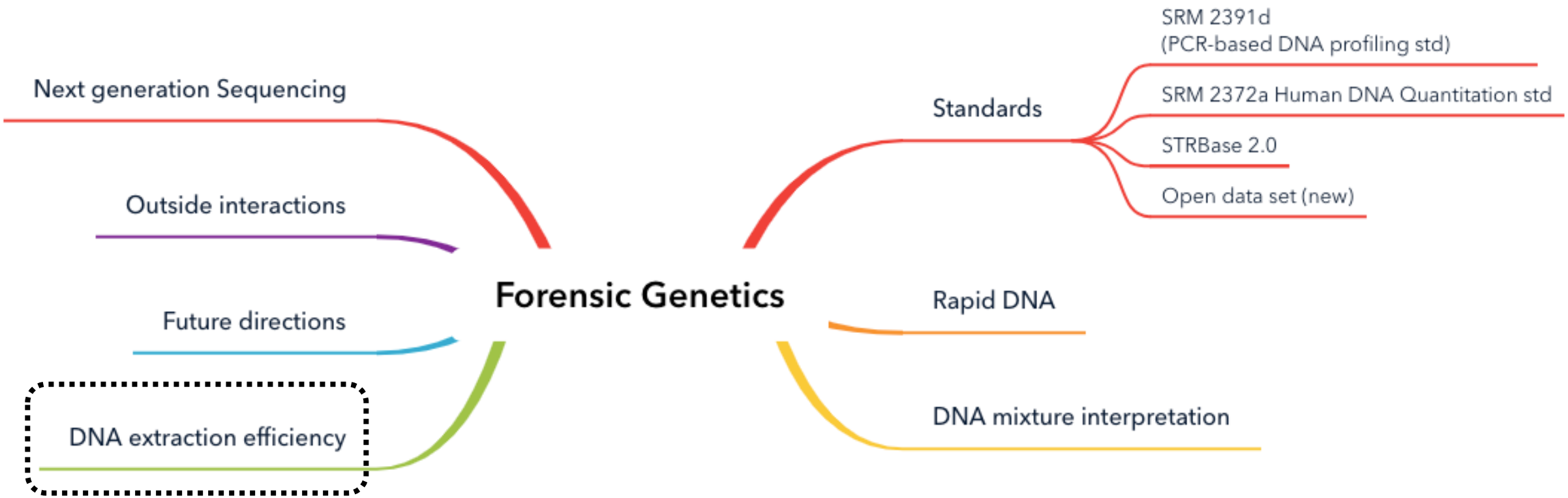
Genotyping Success: Modified Rapid DNA Analysis



CODIS 20: 90%
Full Profile: 90%

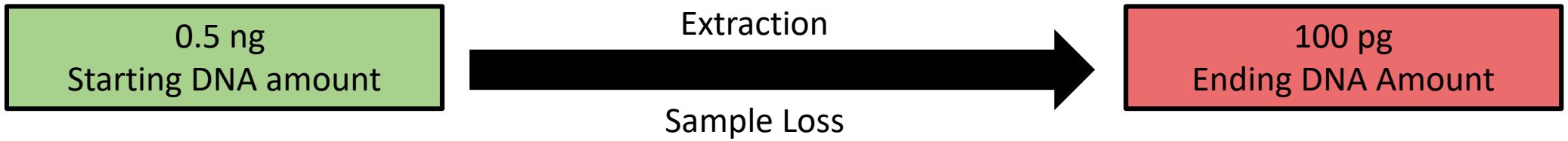
Each data point is an independent instrument

Topics for today



Examination of Front-End Methods in DNA Typing

- Problem: Assess the amount of sample loss during the extraction
  - Low extraction efficiency cold result in overall lower sample quantity
  - May fail to yield full STR profiles or minor components in mixtures



Methods for determining extraction efficiency and sample loss vary

Absolute Extraction Efficiency

$$\frac{\text{DNA Recovered}}{\text{Original Amount DNA}} = \text{Absolute Extraction Efficiency}$$

Measured by digital PCR

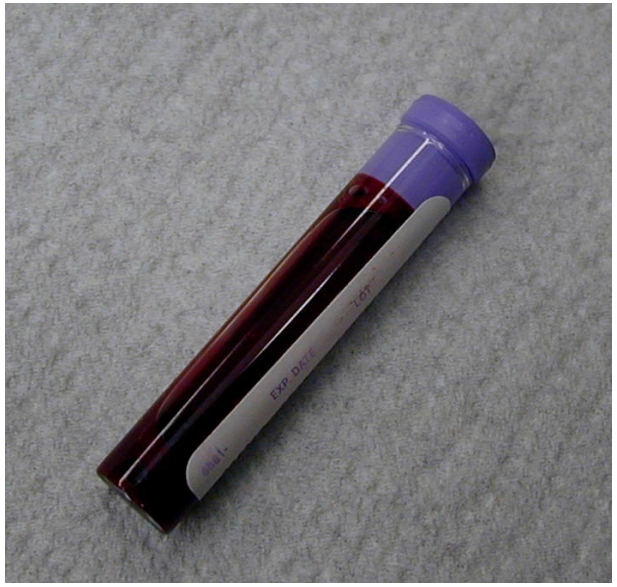
Offers the ability to evaluate individual extraction processes and their efficiency independent of another method

DNA Sources

Component A of SRM 2372a: Human DNA Quantitation Standard

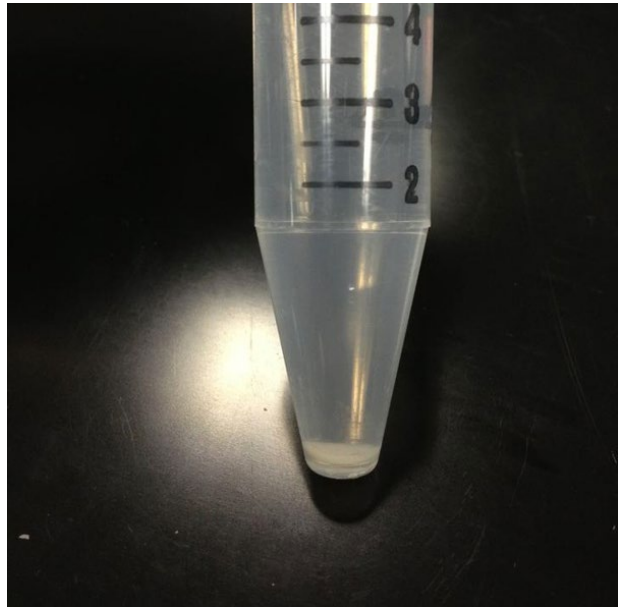
Known concentration of 49.8 ng/µL
Determined by digital PCR

Freshly collected whole blood



Known WBC of 4.6 x10³ per µL
WBC reported by blood bank

Washed cell suspension in dPBS

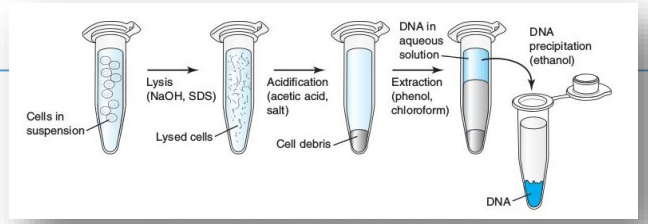


Known cell count of 1x10⁶ per mL
Determined by flow cytometry

DNA Extraction Methods

Phenol Chloroform (Organic)

- Often referred to as the "gold standard"
- Proteinase K digestion of the cells
- Equal volumes of Phenol Chloroform added
- Phase lock gel tubes used for promoting separation
- DNA was precipitated with Ethanol and resolubilized with TE^{-4} buffer



QIAamp Spin Columns

- Manual method commonly used in forensic DNA laboratories
- Silica columns for collection of DNA
- Elution in proprietary buffer
  - Similar to TE^{-4}

Qiagen EZ1 Advanced XL

- Robotic purification instrument
- Cell lysis takes place on the benchtop in a thermomixer
- Purification with paramagnetic bead collection
- Elution in TE^{-4}

Extraction methods available in our lab

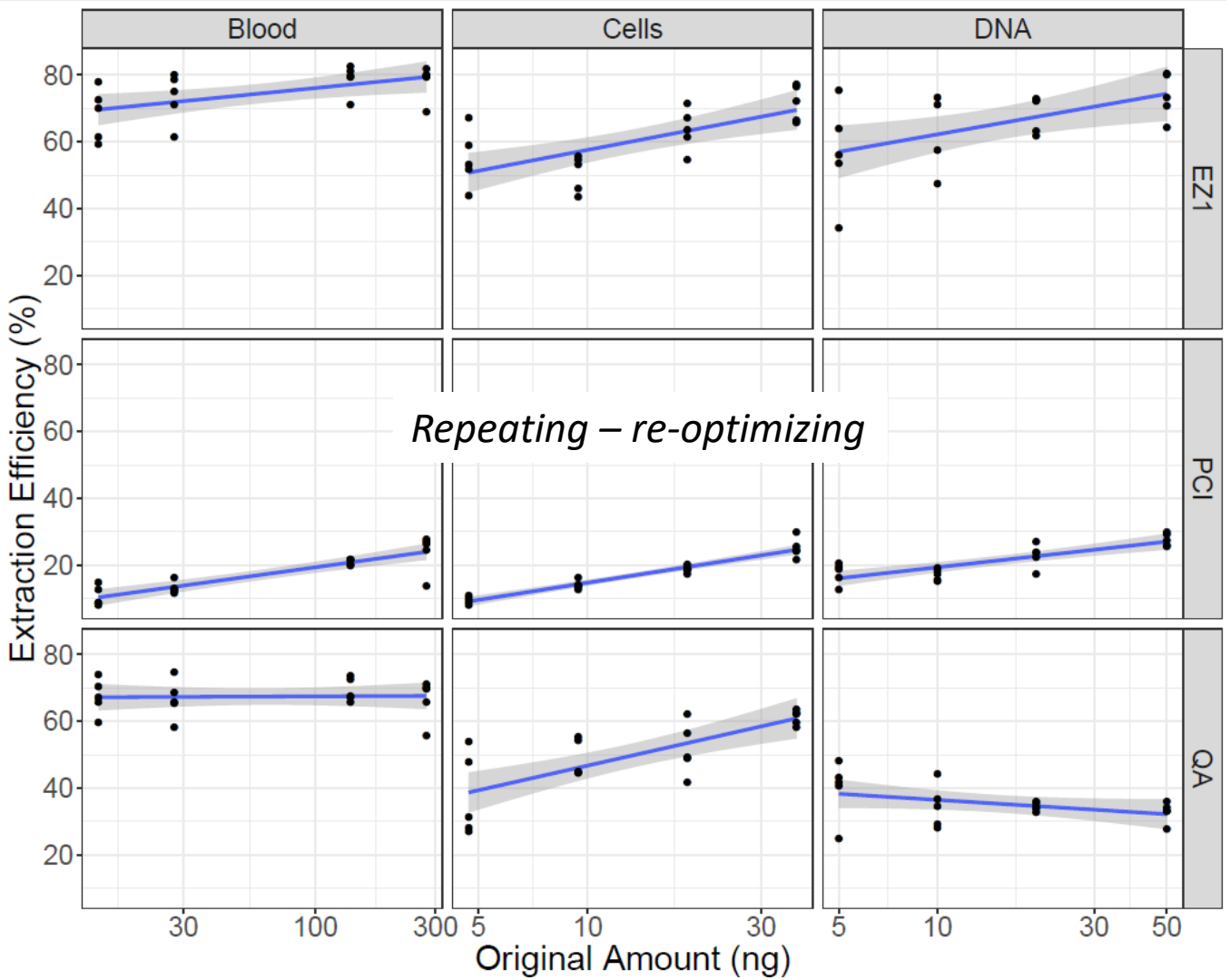
Four DNA input amounts were tested in replicates of five for each extraction method

60 Samples per DNA Source

	Amount (ng)	# of Cells	Uncertainty (± # Cells)	# of Replicates
Extracted DNA	50	8,333	833	5 per amount (20 per DNA Source)
	20	3,333	333	
	10	1,667	167	
	5	781	78	
Cells	38	6,250	313	5 per amount (20 per DNA Source)
	19	3,125	156	
	9	1,563	78	
	5	781	39	
Blood	276	46,000	2,300	5 per amount (20 per DNA Source)
	138	23,000	1,150	
	28	4,667	233	
	14	2,333	117	

60 Samples per Extraction Method

DNA Source



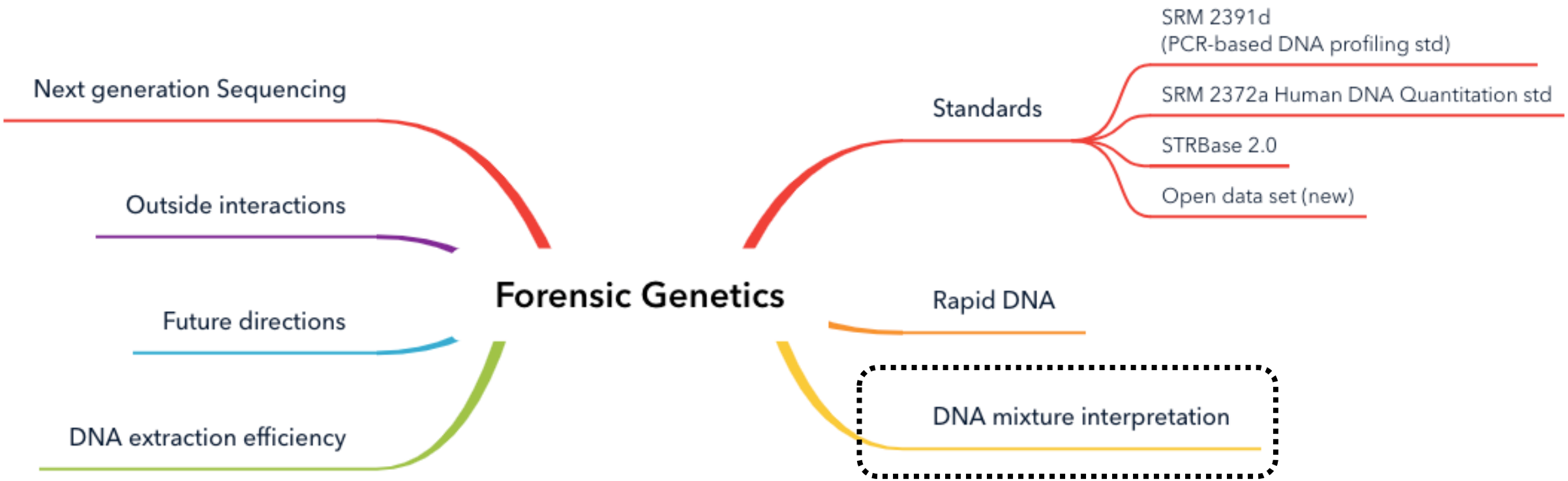
Extraction Method

Digital PCR was used to determine the concentration post-extraction

The efficiency of the extraction methods are shown between methods and DNA sources

Romsos, E.L & Vallone, P.M. (2019) Estimation of extraction efficiency by droplet digital PCR. Forensic Sci Int Genet. Sup. Ser. DOI: 10.1016/j.fsigss.2019.10.072

Topics for today



DNA Mixture Interpretation

LR system

DNA Extraction | DNA Quantitation | PCR Amplification | Genotyping & Sequencing | Interpretation

Statistics

measurement

Assign a strength of evidence
Likelihood ratio (LR)

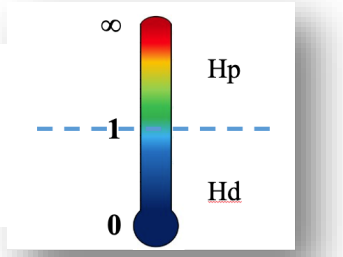
H_p: the DNA from the POI **IS** in the mixture

H_d: the DNA from the POI **IS NOT** in the mixture

I: background information

E: evidence

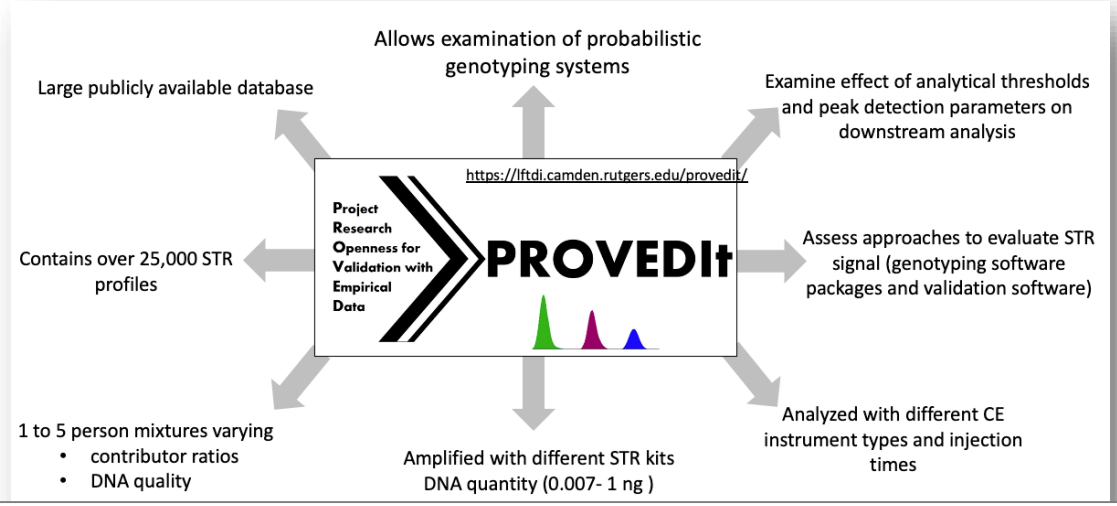
$$LR = \frac{Pr(E|Hp, I)}{Pr(E|Hd, I)}$$



DNA Mixture Interpretation

- Examine methods to assess the performance of LR systems using publicly available ground truth data (mixture profiles)

- Examine the similarities and differences between the LR systems

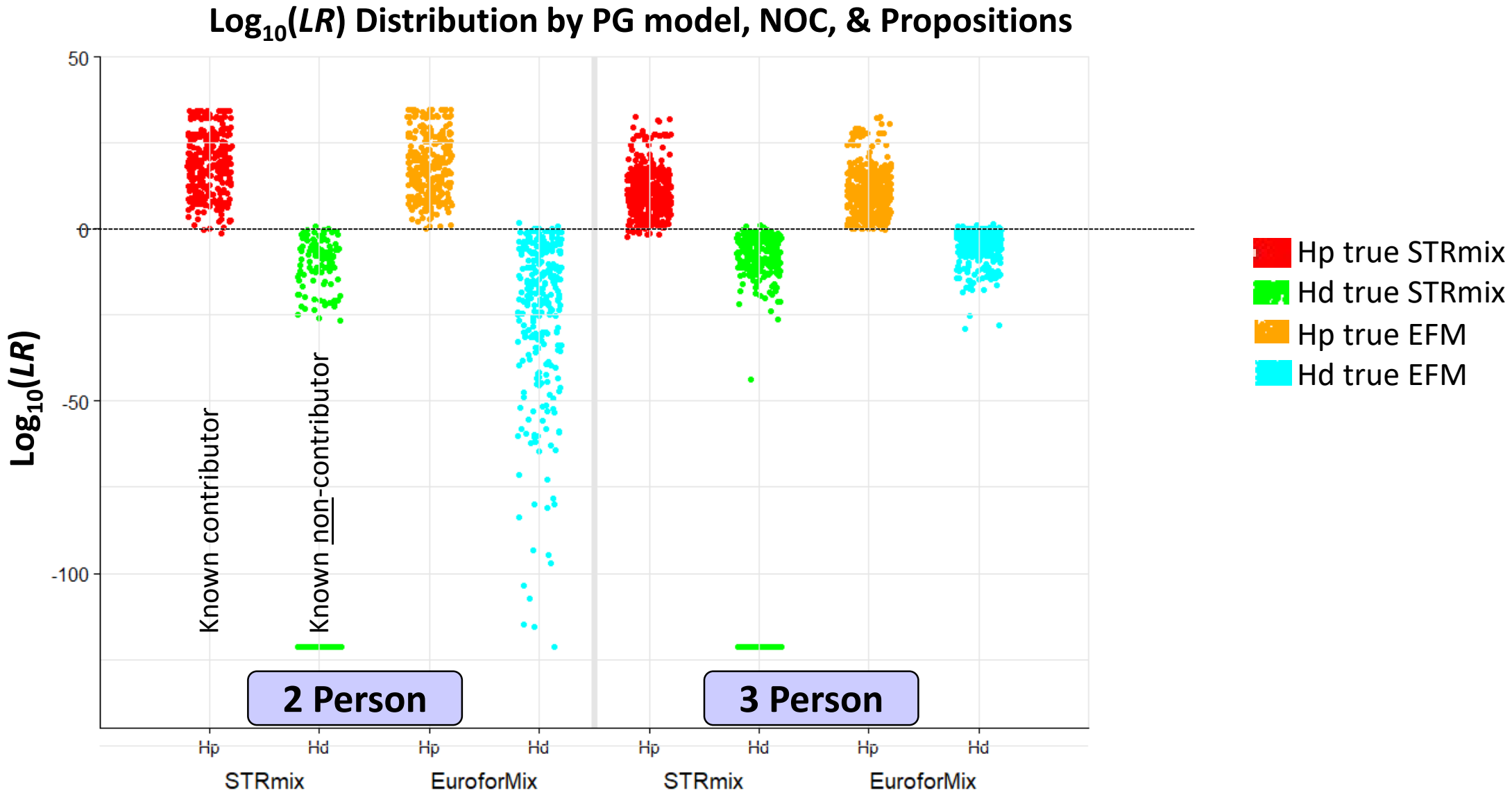


Alfonse, L.E., et al. A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. Forensic Sci. Int. Genetics 32, 62-70.

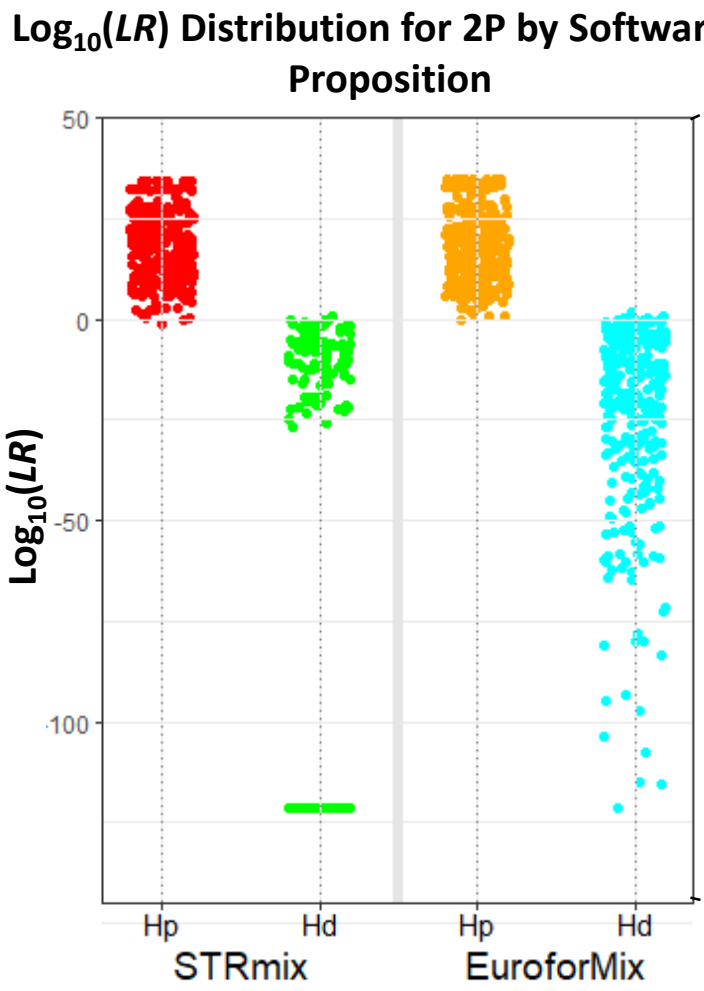
STRmix v2.6	EuroForMix v2.1.0
▪ N-1, N-2 and N+1 stutter peaks modeled ▪ Drop-in frequency = 0.0015 and maximum cap = 180 RFU ▪ Saturation threshold = 30,000 RFU ▪ MCMC settings: 8 chains of 100,000 burn-in accepts, 50,000 post burn-in accepts per chain ▪ > 300 single source profiles used for Model Maker ▪ Sub-source LR	▪ MLE (Maximum likelihood estimation) approach ▪ Degradation and stutter models on ▪ Default parameters except for a 35 RFU detection threshold, Pr(C) = 0.0015 and λ = 0.018. ▪ MLE based LR
https://www.strmix.com/	http://www.euroformix.com/

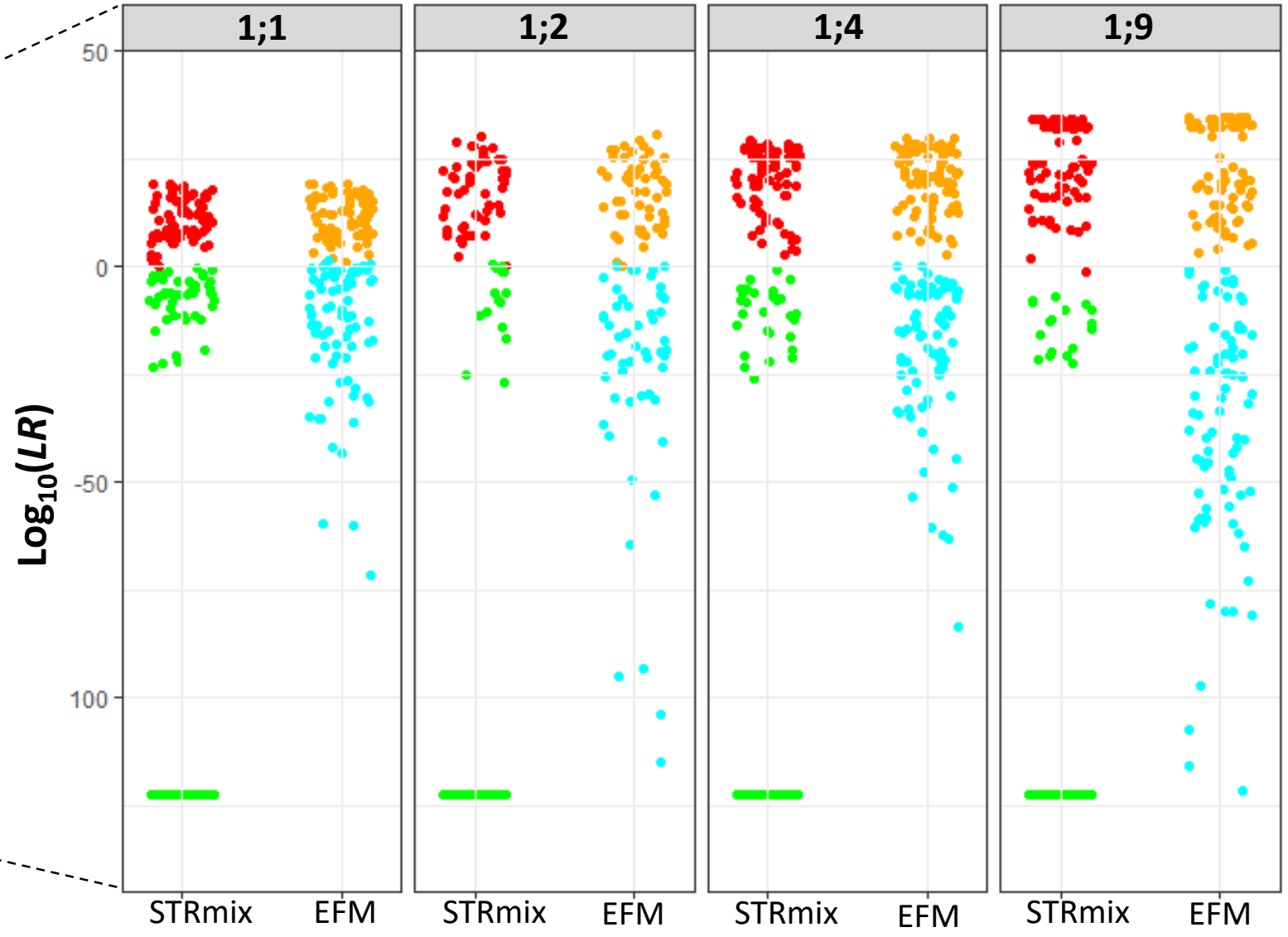
- Profiles were analyzed using the per dye ATs
- NIST 1036-Caucasian allele frequencies
- θ correction was applied using an F_{st} (θ) = 0.01
- True NOC and same propositions were used in both software

Discrimination power of LR Systems using Hp true & Hd true LR distribution

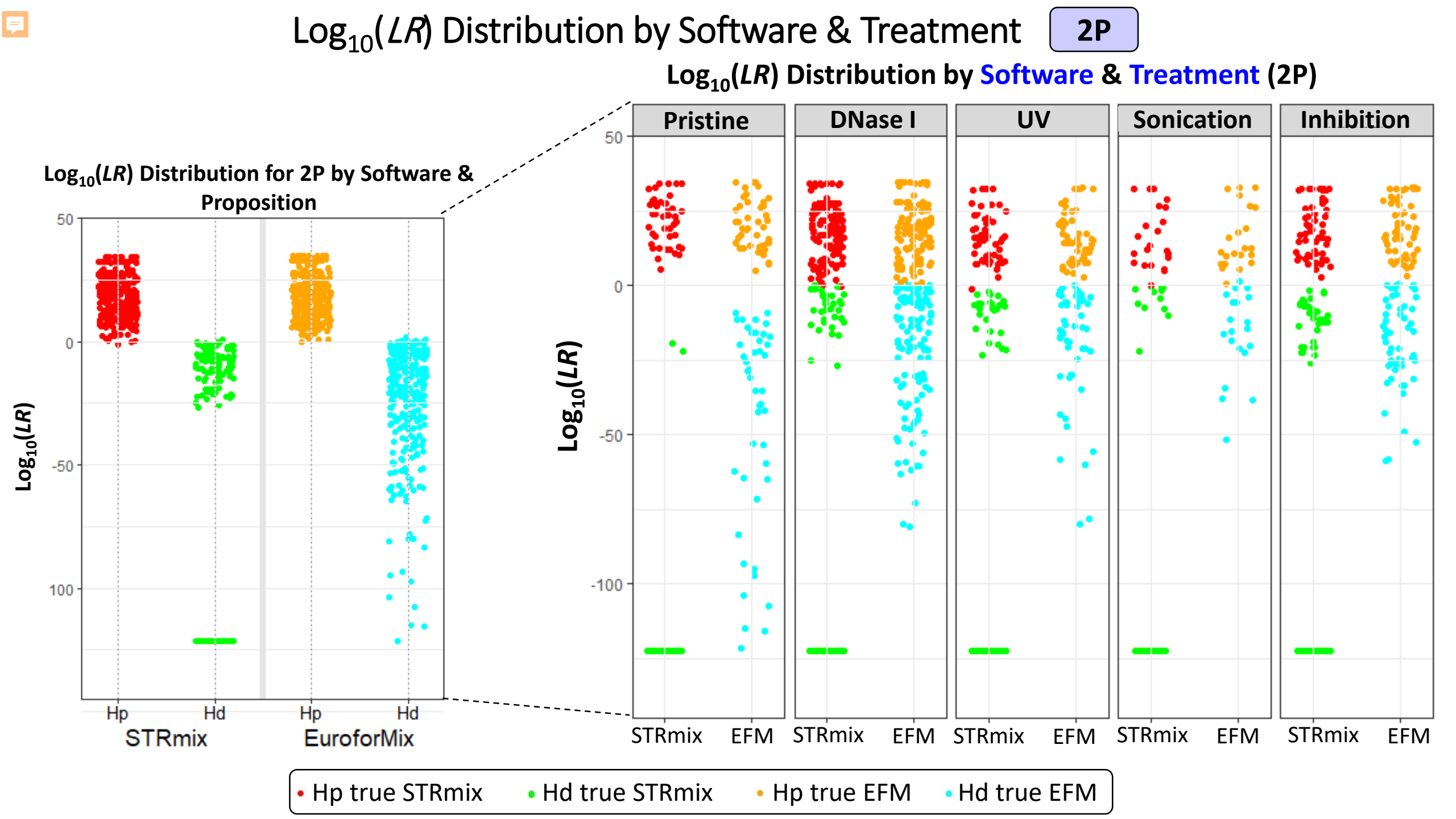


$Log_{10}(LR)$ Distribution by Software & Mixture Ratios **2P**

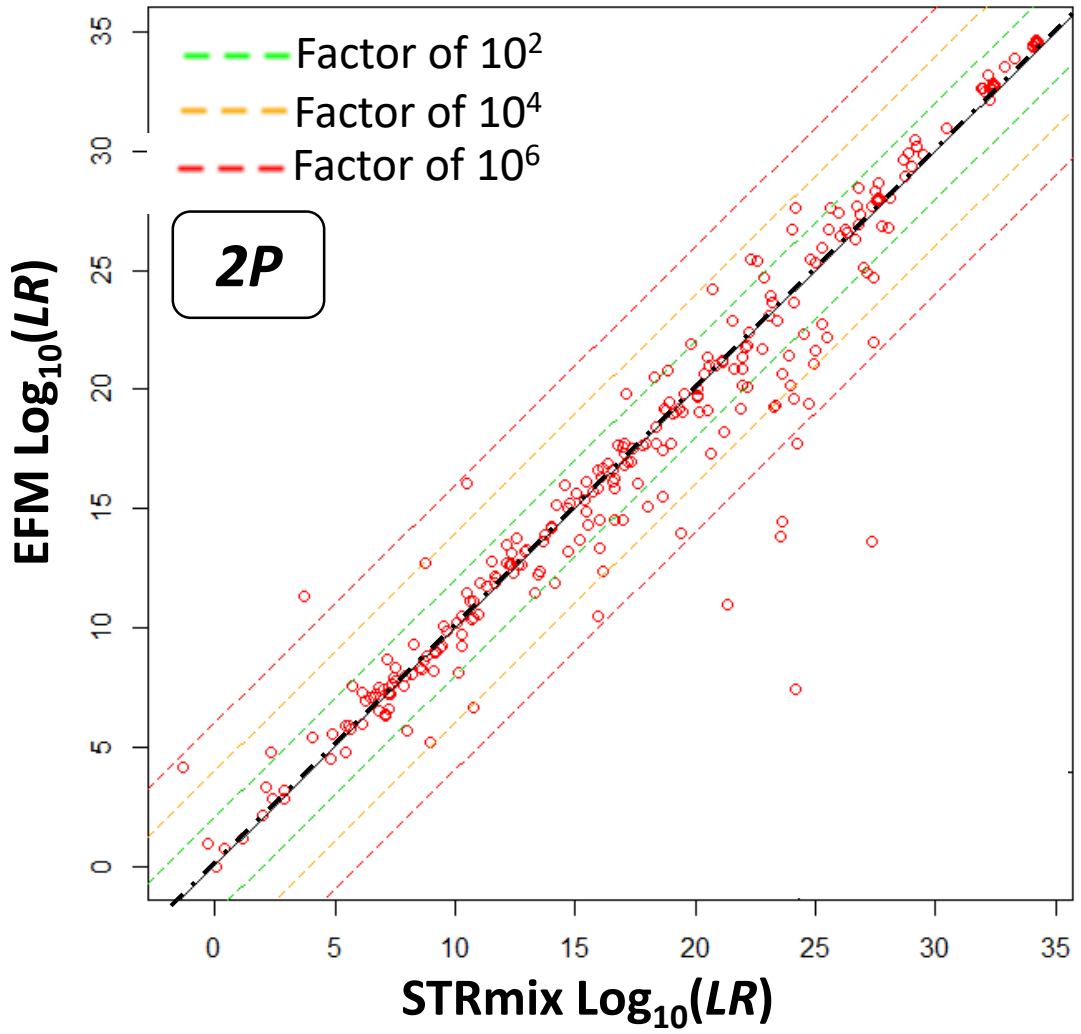


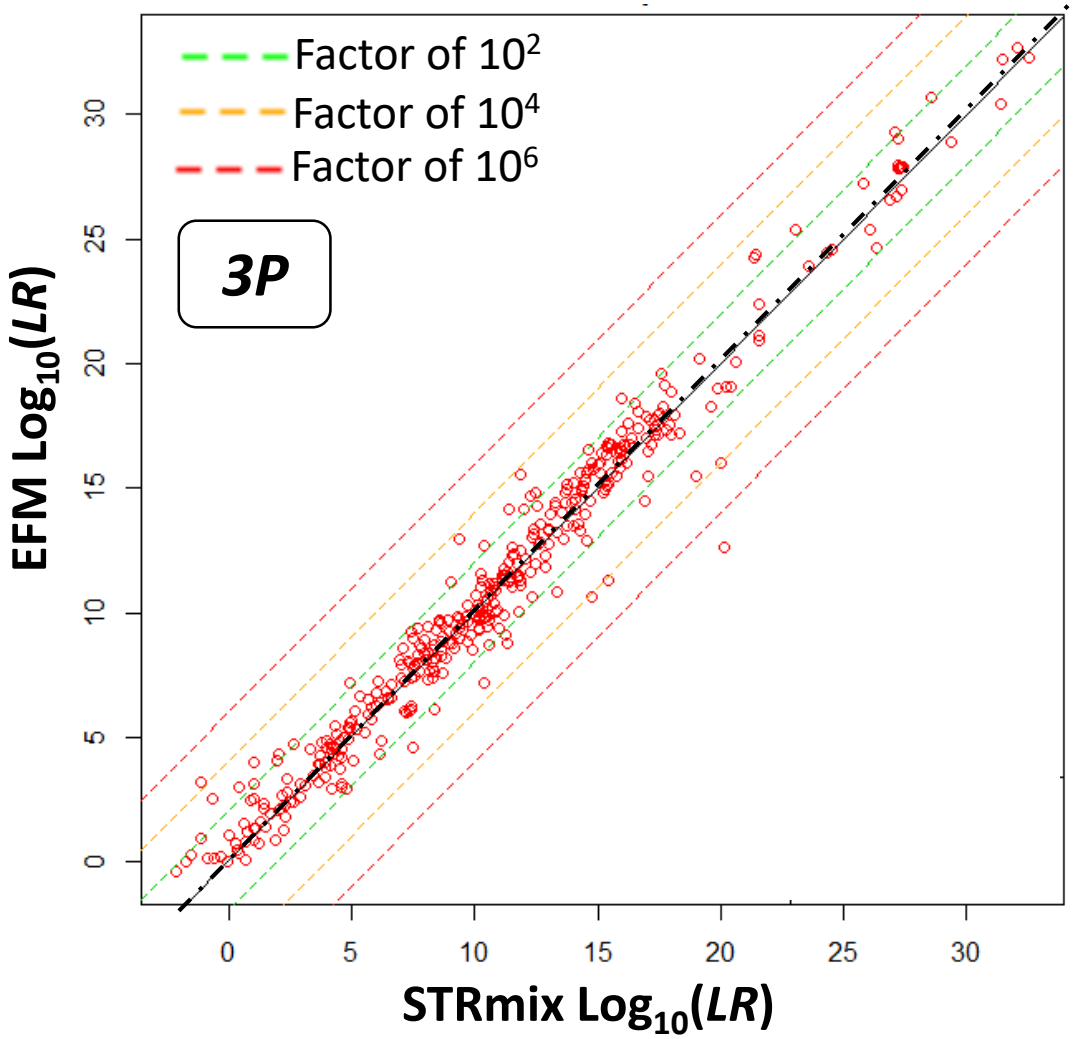


$Log_{10}(LR)$ Distribution by Software & Treatment 2P

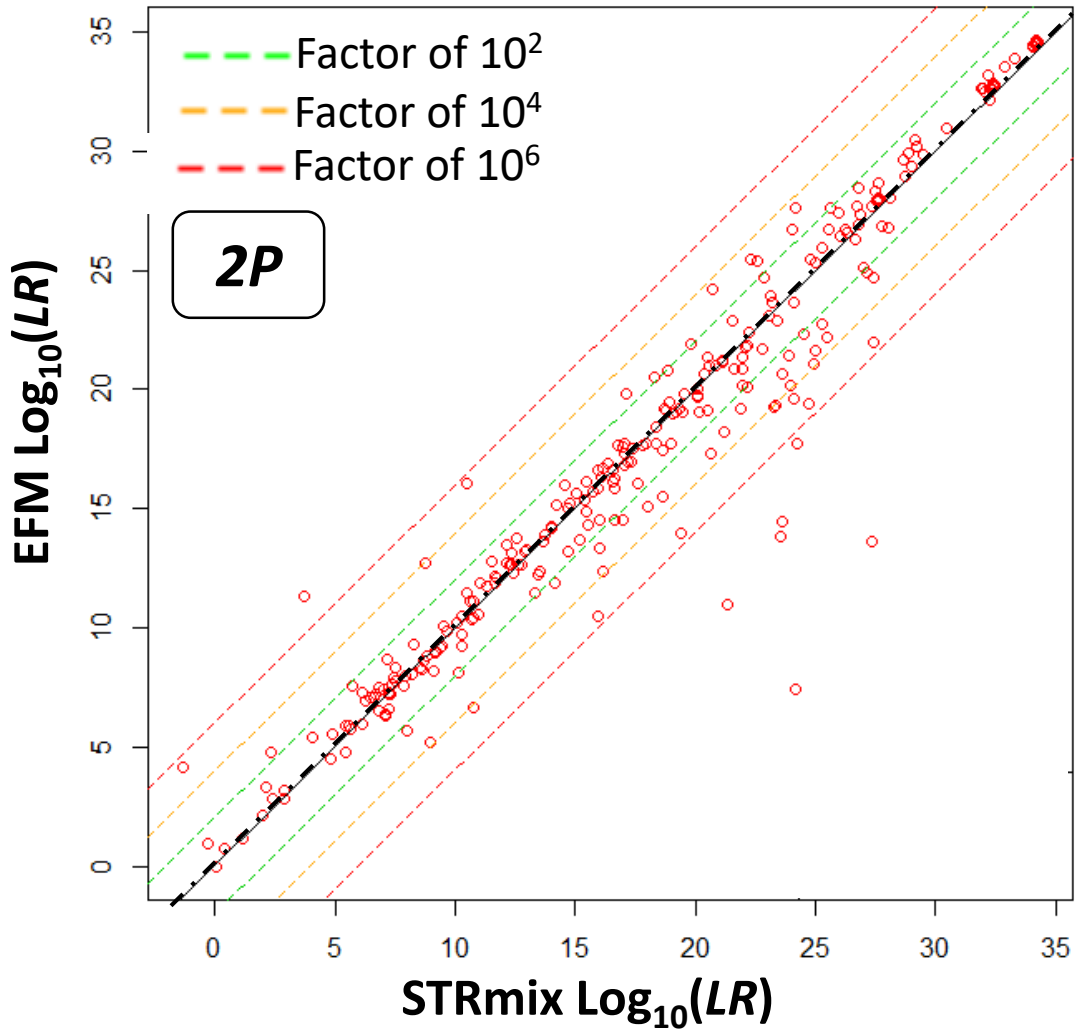


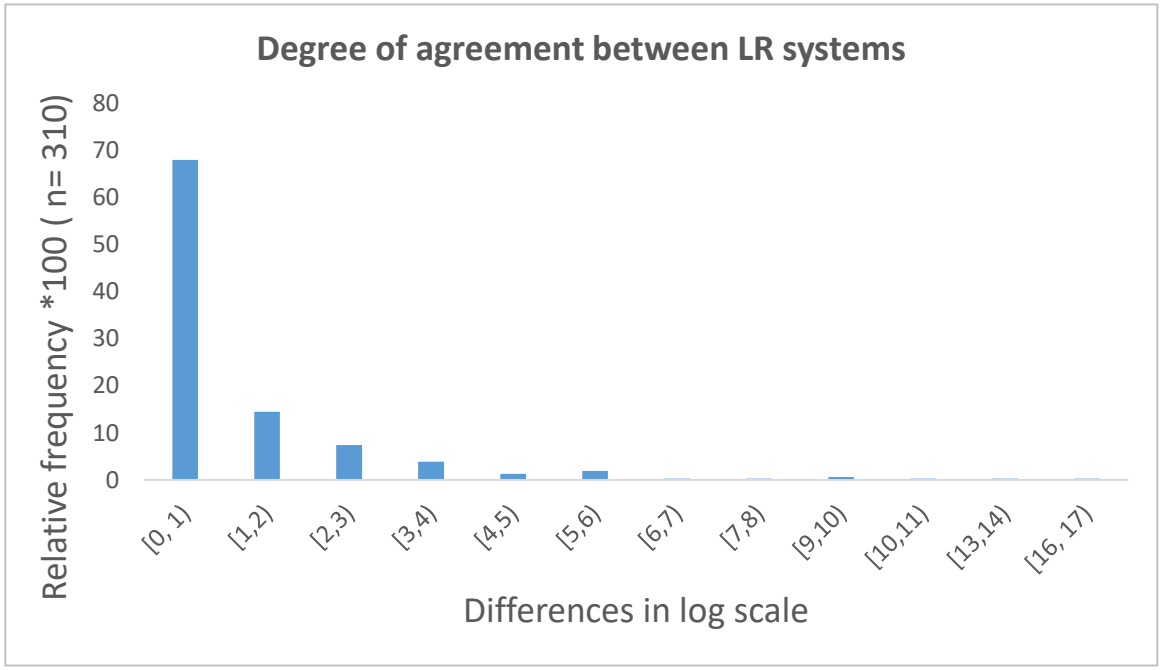
Global profile $Log_{10}(LR)$ from 2P and 3P for Hp true





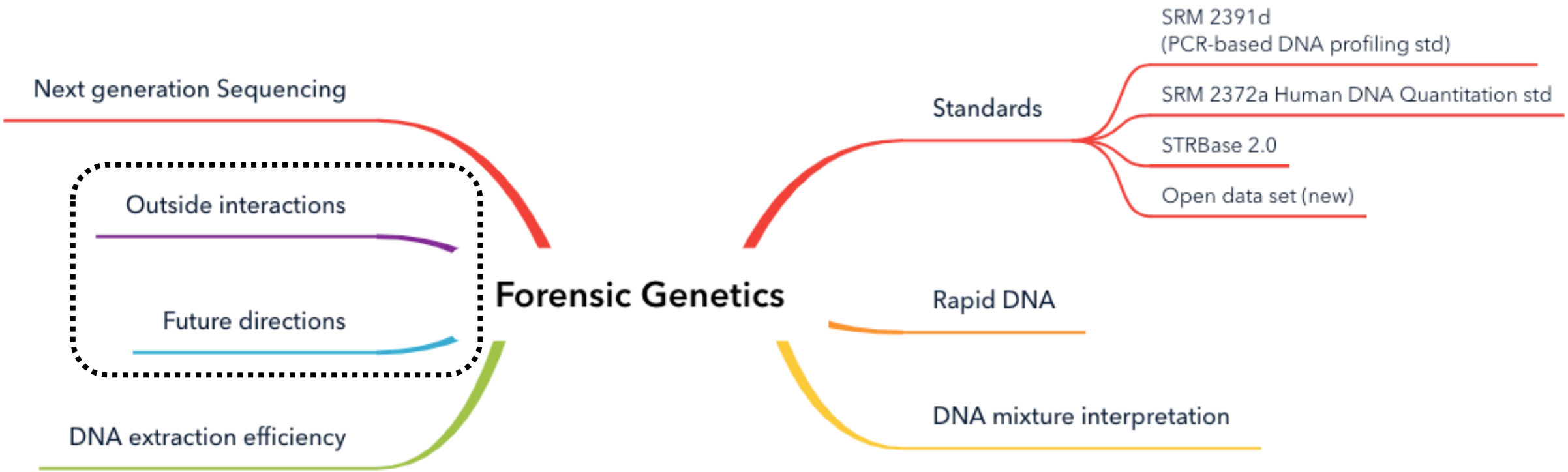
Global profile $Log_{10}(LR)$ from 2P for Hp true





- In the process of finalizing 4-person mixtures
- Investigating the sources of variation
- 155 two-person mixtures
- 147 three-person mixtures
- 132 four-person mixtures

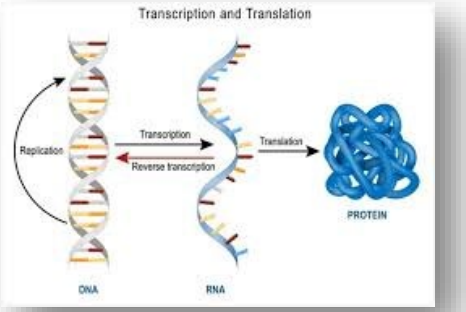
Topics for today



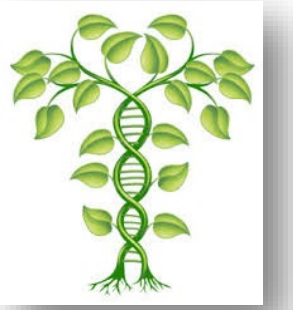
Working group and stakeholder engagement

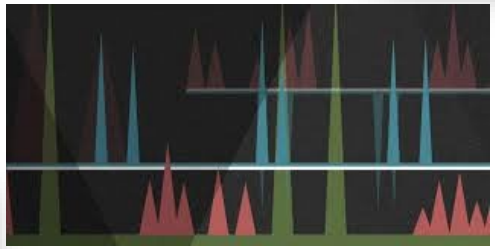
- NIST: Human Factors and Mixture Review
- NIJ FLNTWG – Sequencing white paper
- NIJ Technology Working Group
- FBI: SWGDAM
  - Laboratory Operations, Sequencing, Body Fluid Identification
- FBI: Rapid DNA task groups
- ISFG: STR Nomenclature (Recommendations)
- NIST/ANSI Type 18 DNA Standard Working Group
- OSAC: Sequencing subgroup
- Pre-release testing for Promega, Thermo Fisher, QIAGEN

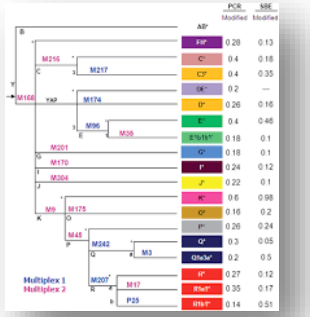
Looking forward into



- Skin protein work with IARPA (genetically variant peptides – GVP)
- Assessing the genetic genealogy landscape – needs for validation, further foundational research, and written standards
- Probabilistic Modeling for Forensic Interpretation of DNA Mixtures Using Next Generation Sequencing Data
- The application of AI/machine learning for DNA mixtures (NoC, deconvolution)
- Y SNP interlaboratory study (Thermo Fisher, 800+ Y SNPs)
- Hosting a continuing education day ***at NIST***
- Develop digital PCR assays for the Y chromosome







Thank you for your attention


Peter Vallone


Becky Steffen


Erica Romsos


Katherine Gettings


Kevin Kiesler


Margaret Kline


Lisa Borsuk


Sarah Riman


David Duewer


Hari Iyer

Statistical Support


Tunde Huszar PostDoc

peter.vallone@nist.gov

Forensic Genetics: Next Generation Sequencing

Katherine Butler Gettings, Ph.D.
Research Biologist, Applied Genetics Group
Forensics at NIST 2020
November 5, 2020

disclaimer

Points of view in this document are those of the author 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 Research Protections Office.

Forensic Genetics Team


Peter Vallone


Becky Steffen


Erica Romsos


Katherine Gettings


Kevin Kiesler


Margaret Kline


Lisa Borsuk


Sarah Riman


David Duewer

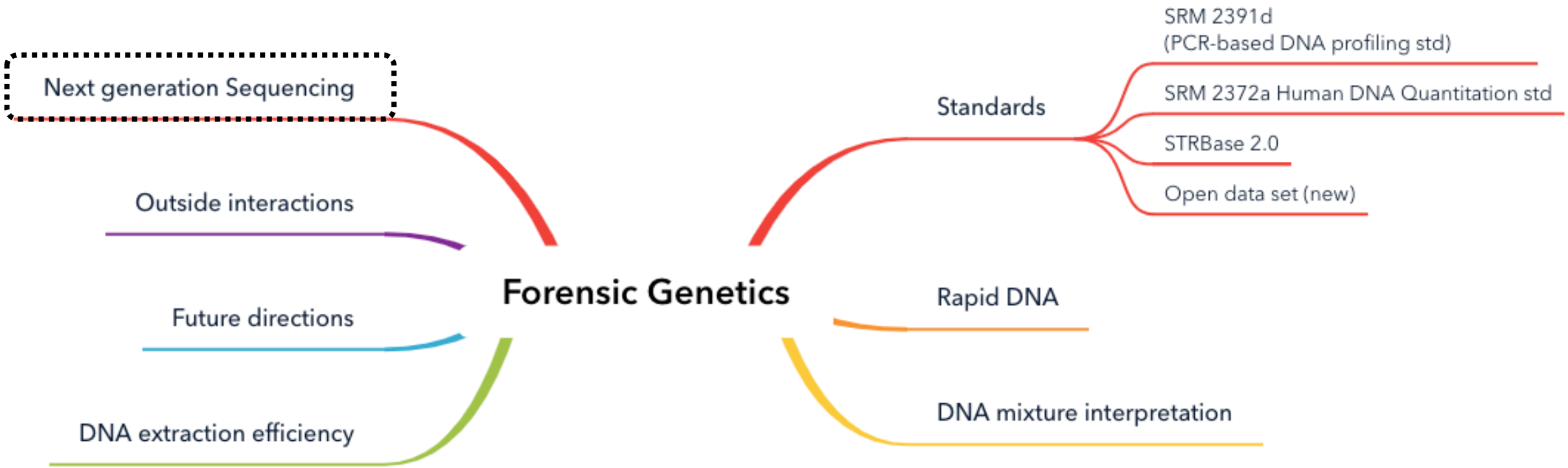

Hari Iyer

Statistical Support

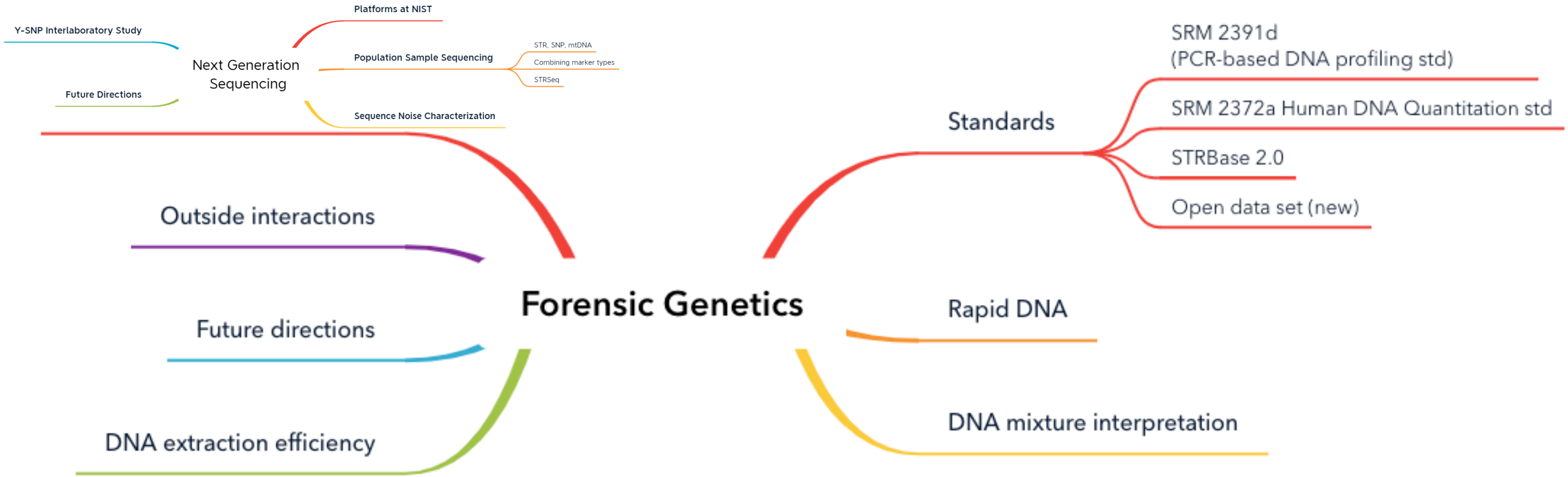

Tunde Huszar PostDoc

Margaret Kline will be retiring on November 20, 2020
Congratulations Margaret on a 35 year career at NIST
We'll miss you!

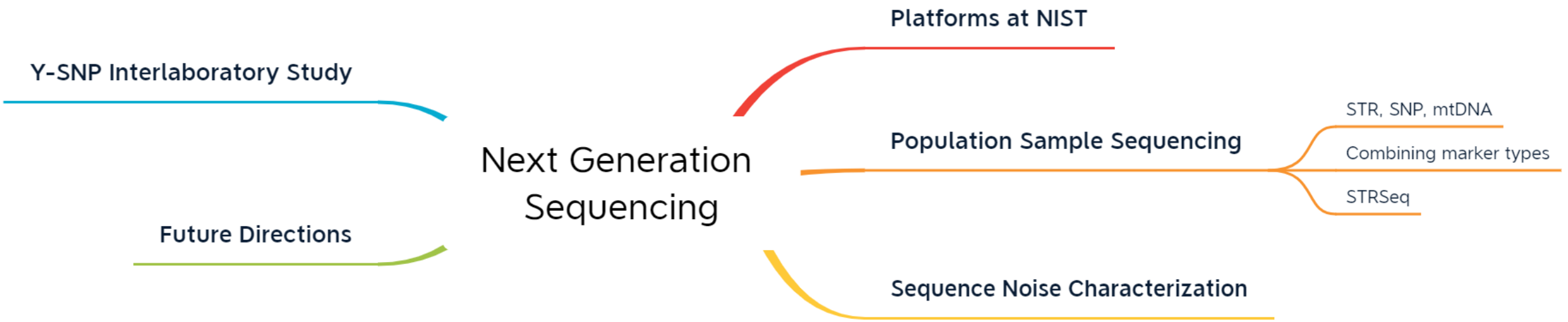
Topics for today



Topics for today



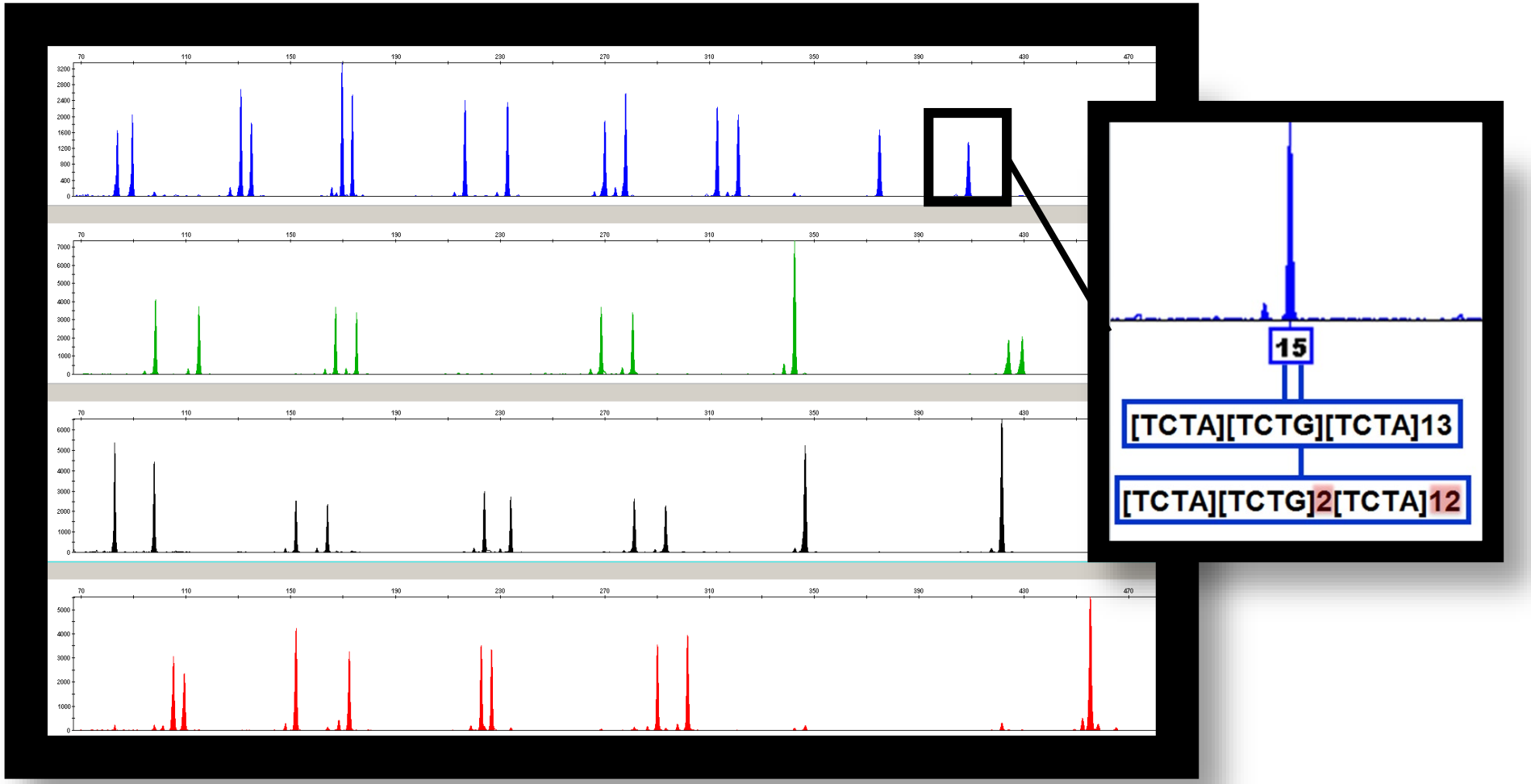
Topics for today



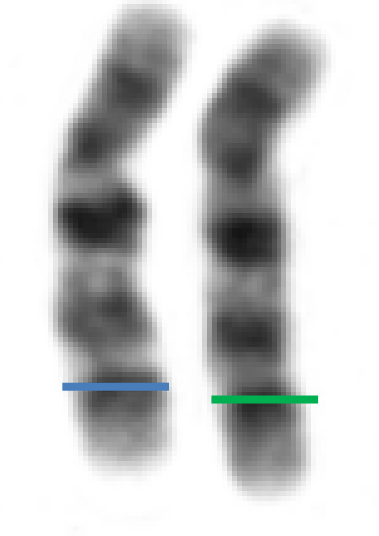
Short Tandem Repeats

Markers and peaks are separated by size (time)

Colors separated by fluorescent dye labels

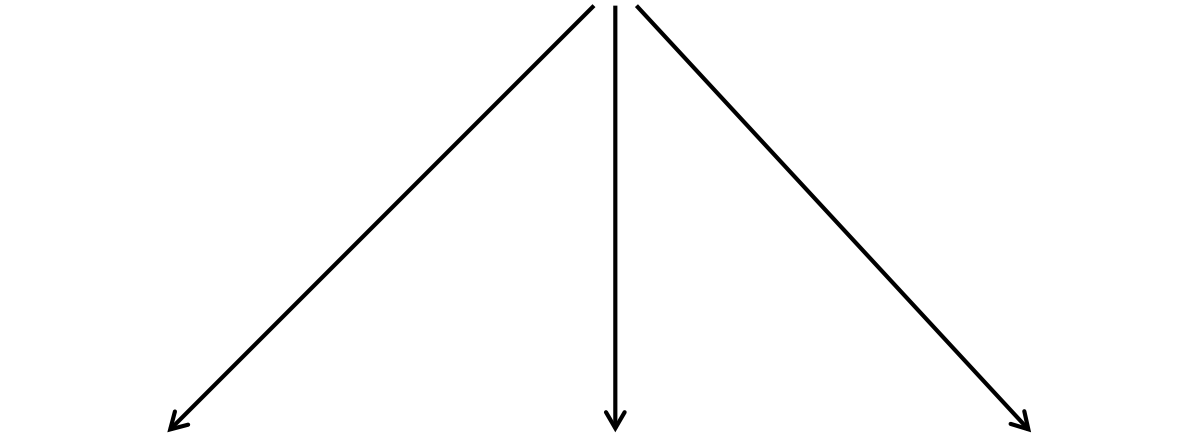


Single Nucleotide Polymorphism



Allele 1: TAGGATCGT**G**CCGATGACTG

Allele 2: TAGGATCGT**A**CCGATGACTG

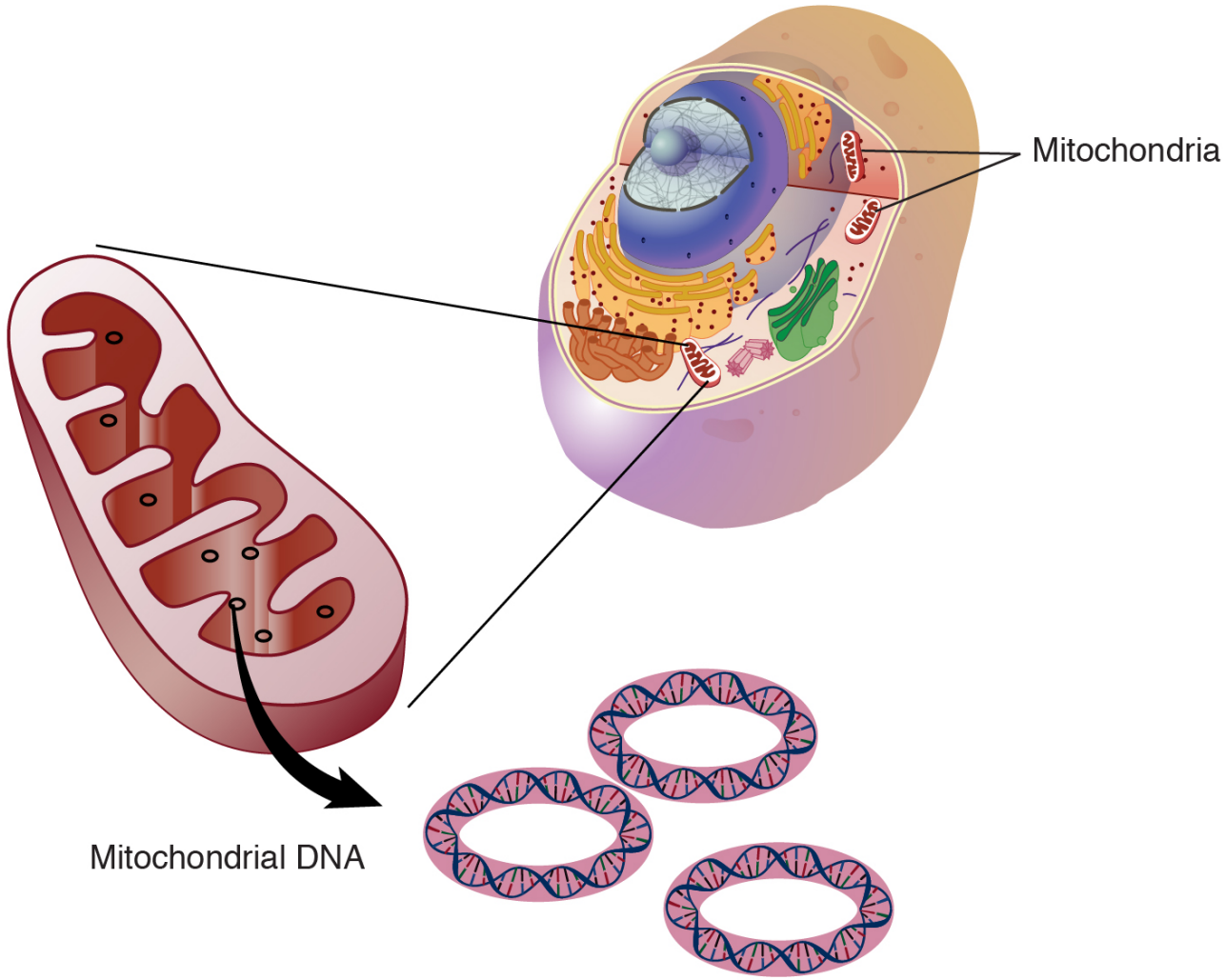


A/A Homozygous

A/G Heterozygous

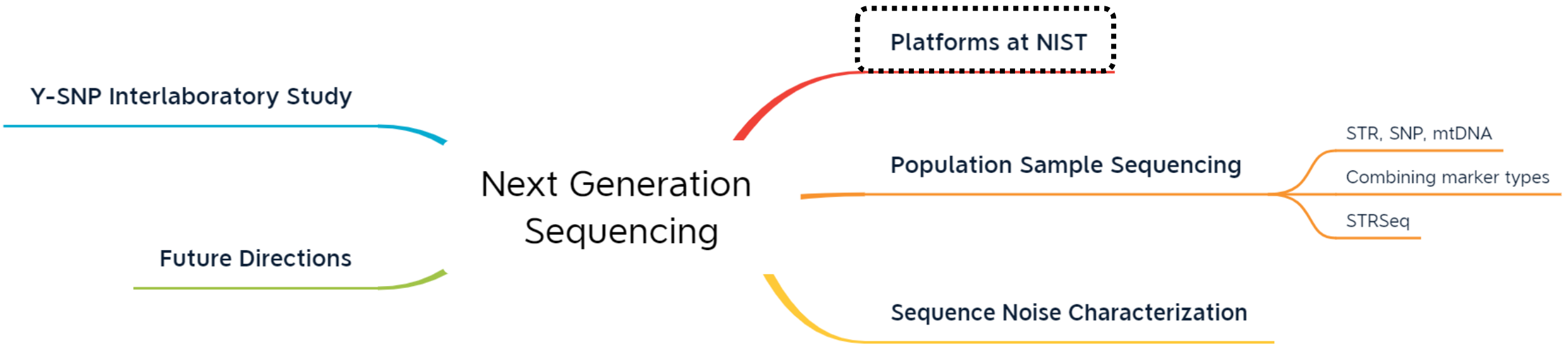
G/G Homozygous

Mitochondrial Genome




15
[TCTA][TCTG][TCTA]13
[TCTA][TCTG]2[TCTA]12
15 16 17 18 19
Forensic Science International: Genetics 37 (2018) 106–115
Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsigen
Research paper
Sequence-based U.S. population data for 27 autosomal STR loci
Katherine Butler Gettings, Lisa A. Borsuk, Carolyn R. Steffen, Kevin M. Kiesler, Peter M. Vallone
Forensic Science International: Genetics 44 (2020) 102192
Understanding the characteristics of sequence-based single-source DNA profiles
Sarah Riman, Hari Iyer, Lisa A. Borsuk, Peter M. Vallone
Research Article
Sequence-based US population data for the SE33 locus
Article
Platinum quality mitogeno
United States populations
MULTI-MARKER MATCH STATISTICS
Combining Results Across Sequence-Based STR and Identity SNP Markers
K.B. Gettings | A. Tillmar | P.M.Vallone
NCBI
SCIENTIFIC WORKING GROUP
SWGDAM
DNA ANALYSIS METHODS
Estimating number of contributors in massively parallel sequencing data of STR loci
Brian A Young, Katherine Butler Gettings, Bruce McCord, Peter M. Vallone

Topics for today



NGS Platforms @NIST

MiSeq FGx / RUO

- Verogen ForenSeq DNA Signature Prep
- Promega PowerSeq 46GY
- Promega PowerSeq CRM Nested System
- QIAseq Targeted Human Mitochondria Panel
- QIAgen GeneRead DNAseq Targeted V2 Panel (Identity SNP)

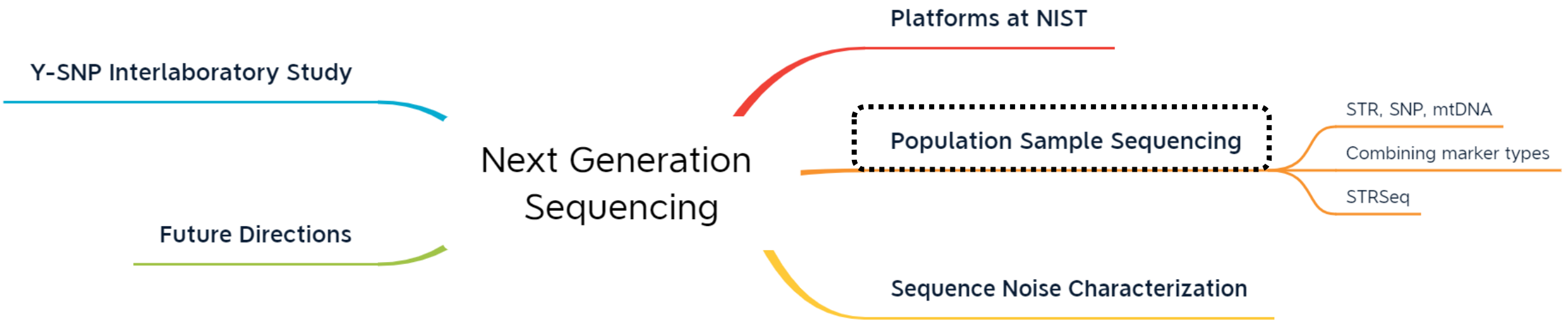
Ion Chef and Ion S5

- Precision ID GlobalFiler NGS STR Panel v2
- Precision ID Identity and Ancestry SNP Panels
- Ion AmpliSeq DNA Phenotyping Panel

MinION

- mtDNA whole genome and microbial genome

Topics for today



Population Sample Sequencing

When a match is made in a forensic case,

allele frequencies are used to calculate the rarity of the DNA profile

Length
8,9
2pq
2*0.144*0.375
0.108
1 in 9.3

D4S2408

Allele	N	Freq	Sequence Allele	N	Freq
7	1	0.6%	[ATCT]7	1	0.6%
8	23	14.4%	[ATCT]8	23	14.4%
9	60	37.5%	[ATCT]9	18	11.3%
			[ATCT] GTCT [ATCT]7	42	26.3%
10	53	33.1%	[ATCT]10	53	33.1%
11	21	13.1%	[ATCT]11	21	13.1%
12	2	1.3%	[ATCT]12	2	1.3%

Sequence
[ATCT]8, [ATCT]9
2pq
2*0.144*0.113
0.033
1 in 30.7

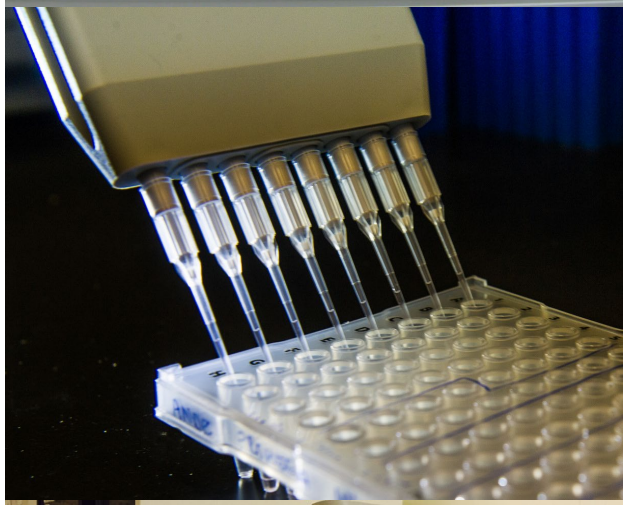
Population Sample Sequencing

STR, Y-STR, X-STR and SNP

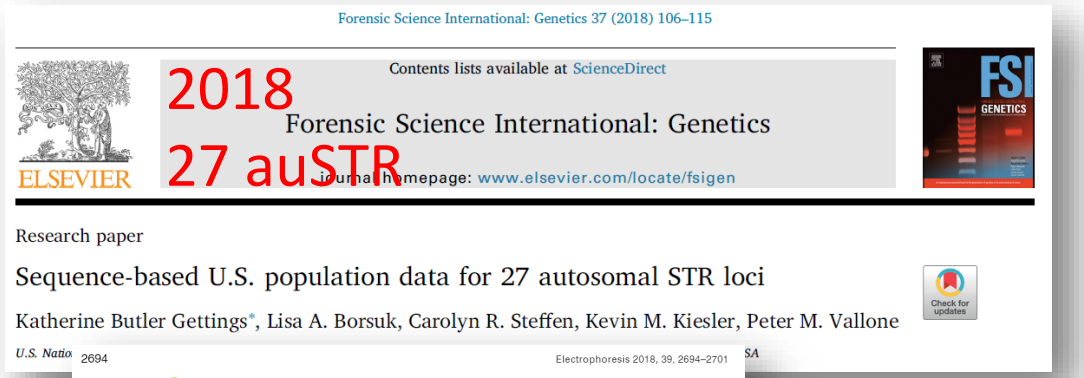
Illumina MiSeq FGx instrument, ForenSeq

- 27 autosomal STRs + 24 Y-STR + 7 X-STR + Amel
- 94 HID-SNPs + 56 ancestry SNPs + 22 phenotype SNPs

- 1036 Samples
- Sequenced in batches of 24 or 32
- 41 total sequencing runs in 2016



Population Sample Sequencing


Forensic Science International: Genetics 37 (2018) 106–115

Contents lists available at ScienceDirect

Forensic Science International: Genetics

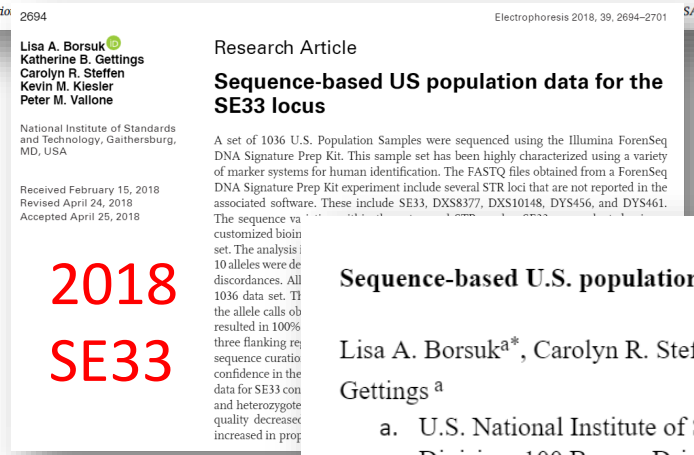
journal homepage: www.elsevier.com/locate/fsigen

2018
27 auSTR

Research paper

Sequence-based U.S. population data for 27 autosomal STR loci

Katherine Butler Gettings*, Lisa A. Borsuk, Carolyn R. Steffen, Kevin M. Kiesler, Peter M. Vallone


2694

Electrophoresis 2018, 39, 2694–2701

Lisa A. Borsuk
Katherine B. Gettings
Carolyn R. Steffen
Kevin M. Kiesler
Peter M. Vallone

National Institute of Standards and Technology, Gaithersburg, MD, USA

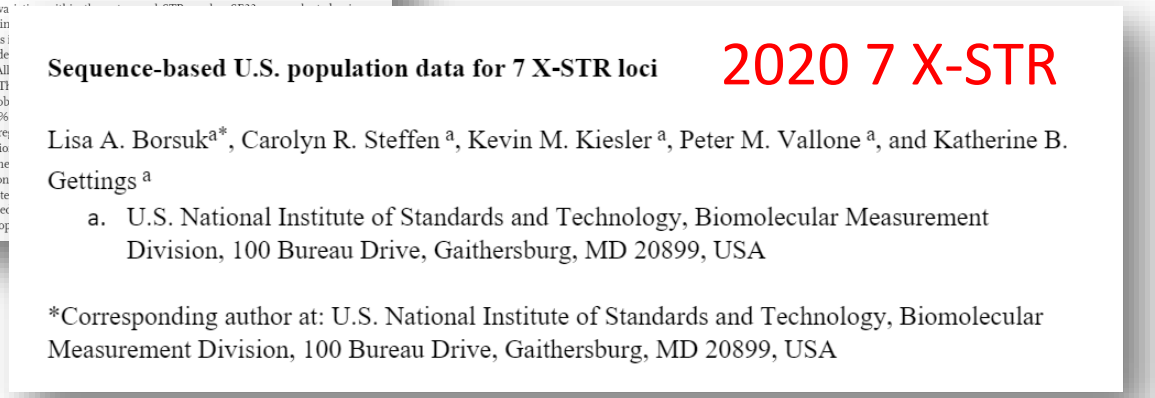
Received February 15, 2018
Revised April 24, 2018
Accepted April 25, 2018

Research Article

Sequence-based US population data for the SE33 locus

A set of 1036 U.S. Population Samples were sequenced using the Illumina ForenSeq DNA Signature Prep Kit. This sample set has been highly characterized using a variety of marker systems for human identification. The FASTQ files obtained from a ForenSeq DNA Signature Prep Kit experiment include several STR loci that are not reported in the associated software. These include SE33, DXS8377, DXS10148, DYS456, and DYS461.

2018
SE33


Sequence-based U.S. population data for 7 X-STR loci

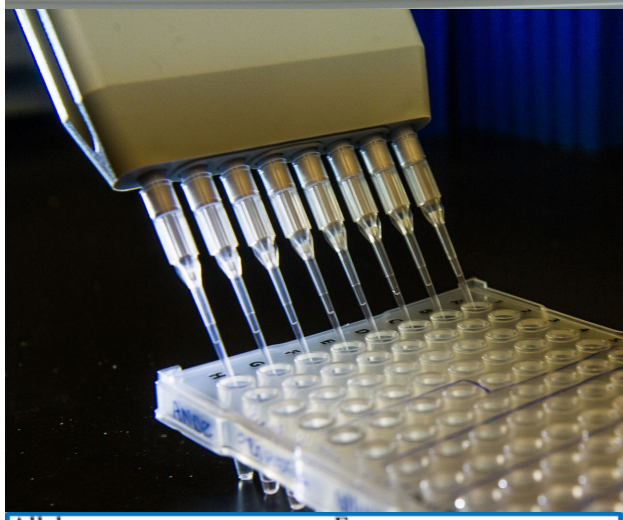
2020 7 X-STR

Lisa A. Borsuk[a*], Carolyn R. Steffen [a], Kevin M. Kiesler [a], Peter M. Vallone [a], and Katherine B. Gettings [a]

a. U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA

*Corresponding author at: U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA

24 Y-STR
and
94 HID-SNP
in progress



Allele	Frequency		
	D3S1358	D5S818	D8S1179
8			1.1%
9		3.1%	1.1%
10		10.4%	17.0%
11		33.3%	9.1%
12		37.5%	8.0%
13		12.5%	9.1%
14	7.1%	3.1%	29.5%
15	31.6%		17.0%
16	30.6%		8.0%

Population Sample Sequencing

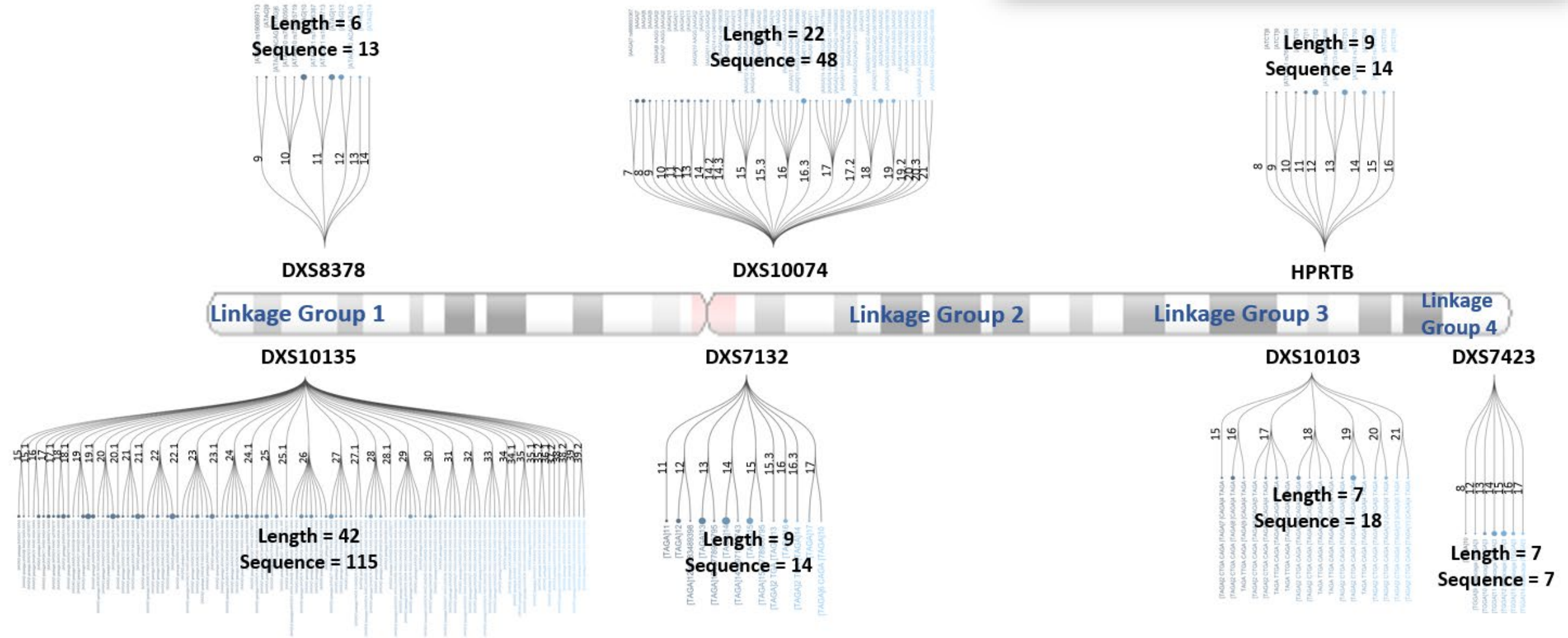
2020 7 X-STR

Sequence-based U.S. population data for 7 X-STR loci

Lisa A. Borsuk[a*], Carolyn R. Steffen [a], Kevin M. Kiesler [a], Peter M. Vallone [a], and Katherine B. Gettings [a]

a. U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA

*Corresponding author at: U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA



Population Data in NIST PDR

Data Publication

Sequence-based U.S. population data for 27 autosomal STR loci

Lisa A. Borsuk, Katherine B. Gettings, Kevin M. Kiesler, Carolyn R. Steffen, Peter M. Vallone

Contact: Katherine Gettings..

Identifier: doi:10.18434/t4/1500024
Version: **1.0.0+ (in edit)**... Last modified: **2018-06-14**

Description

This information and data are supplemental files associated with: K.B. Gettings, L.A. Borsuk, C.R. Steffen, K.M. Kiesler, P.M. Vallone, Sequence-based U.S. population data for 27 autosomal STR loci, Forensic Science International: Genetics 37 (2018) 106-115. The primary data consists of sequence-based allele frequencies for N=1036 anonymized U.S. population samples at 27 autosomal Short Tandem Repeat (auSTR) loci: D1S1656, TPOX, D2S441, D2S1338, D3S1358, D4S2408, FGA, D5S818, CSF1PO, D6S1043, D7S820, D8S1179, D9S1122, D10S1248, TH01, vWA, D12S391, D13S317, Penta E, D16S539, D17S1301, D18S51, D19S433, D20S482, D21S11, Penta D, and D22S1045. This information is expected to support the implementation of sequence-based STR analysis in forensic applications. /"NIST1036_auSTR_Seq_SuppTables.xls/" is an excel file containing the following worksheets: run metrics for the 42 sequencing runs performed to generate the allele frequency data (S1 - Run Metrics); coverage per locus per sample for all N=1036 at the 27auSTR, 7XSTR, and 24 YSTR loci reported by the manufacturer in this assay (S2 - Coverage); allele frequency data (S3 - Frequencies); GRCh38 reference coordinates for genomic regions reported in the 27 auSTRs (S4 - Ref Coordinates); summary of polymorphisms detected and reported in STR flanking regions (S5 - Flank Polymorph); number of alleles, expected and observed heterozygosity, and p-values associated with HWE testing by population for the 27 auSTR loci (S6 - Hexp_Hobs_pHW); p-values associated with testing for linkage disequilibrium (S7 - LD p-values); and pairwise Fst values by population for the 27 auSTR loci (Supp Table 8 - Pairwise Fst). Lastly, /"NIST1036_auSTR_Seq_SuppFile1.pdf/" contains information on optimization, sequencing, and quality control of the data.

Research Topics: Forensics: DNA and biological evidence

Subject Keywords: STR, forensic, sequence, population, allele frequency

Data Access

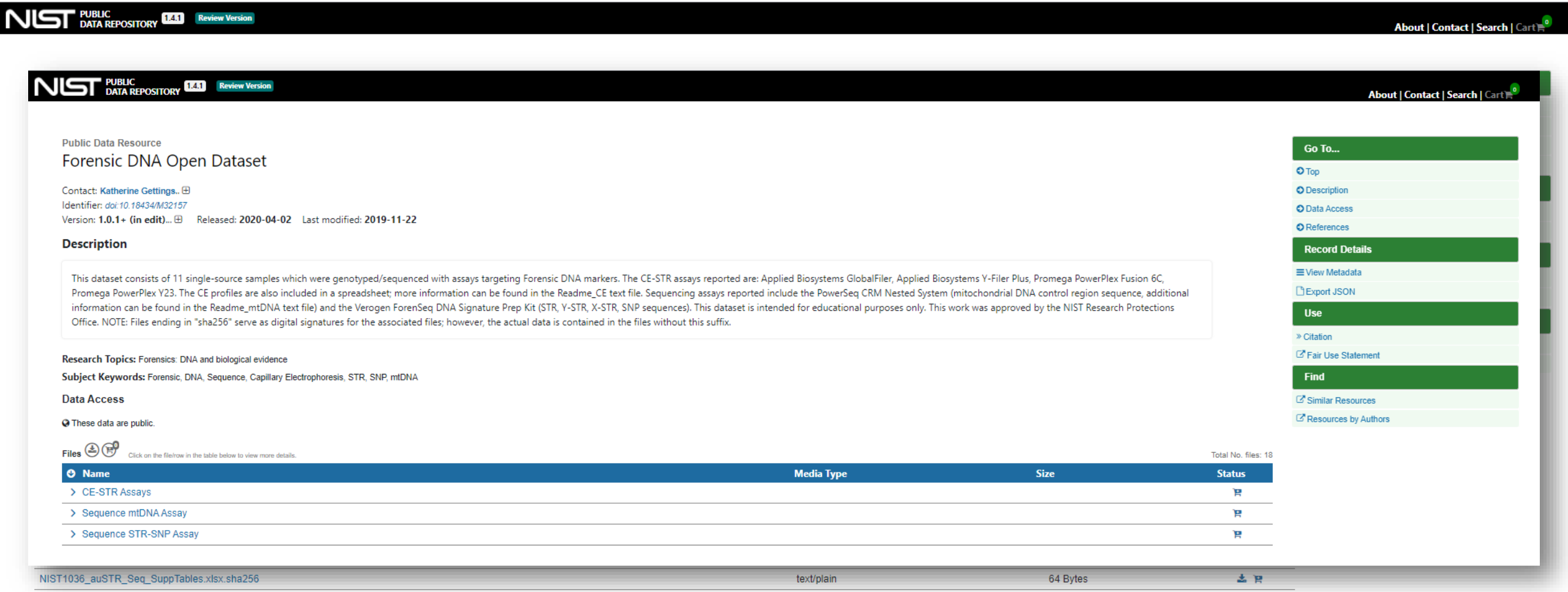
These data are public.

Files Click on the file/row in the table below to view more details. Total No. files: 4

Name	Media Type	Size	Status
NIST1036_auSTR_Seq_SuppFile1.pdf	application/pdf	817.9 kB	
NIST1036_auSTR_Seq_SuppFile1.pdf.sha256	text/plain	64 Bytes	
NIST1036_auSTR_Seq_SuppTables.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	648.5 kB	
NIST1036_auSTR_Seq_SuppTables.xlsx.sha256	text/plain	64 Bytes	

Go To...
Top
Description
Data Access
References
Record Details
View Metadata
Export JSON
Use
Citation
Fair Use Statement
Find
Similar Resources
Resources by Authors

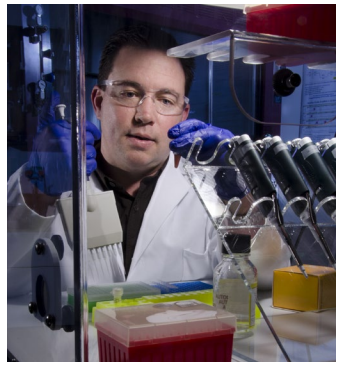
Population Data in NIST PDR



Mitochondrial Genome
Population Sample Sequencing

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

659 samples attempted
- African American
- U.S. Caucasian
- U.S. Hispanic



ARMED FORCES MEDICAL EXAMINER SYSTEM

704 samples attempted
- African American
- U.S. Caucasian
- U.S. Hispanic
- Native American
- Asian

Cassandra Taylor
Kimberly Sturk-Andreaggi
Charla Marshall

genes
MDPI

Article

Platinum-Quality Mitogenome Haplotypes from United States Populations

Cassandra R. Taylor [1,2], Kevin M. Kiesler [3], Kimberly Sturk-Andreaggi [1,2], Joseph D. Ring [1,2], Walther Parson [4,5], Moses Schanfield [6], Peter M. Vallone [3] and Charla Marshall [1,2,5,*]

1 Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Dover Air Force Base, DE 19002, USA; cassandra.r.taylor7.ctr@mail.mil (C.R.T.); kimberly.s.andreaggi.ctr@mail.mil (K.S.-A.); joseph.d.ring2.ctr@mail.mil (J.D.R.)
2 SNA International, LLC; Alexandria, VA 22314, USA
3 National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA; kevin.kiesler@nist.gov (K.M.K.); peter.vallone@nist.gov (P.M.V.)
4 Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck 6020, Austria; walther.parson@i-med.ac.at
5 Forensic Science Program, The Pennsylvania State University, State College, PA 16801, USA
6 Department of Forensic Sciences, The George Washington University, Washington, DC 20007, USA; mschanfi@gwu.edu
* Correspondence: charla.k.marshall.ctr@mail.mil; Tel.: +1-302-346-8519

Received: 2 October 2020; Accepted: 27 October 2020; Published: 29 October 2020

check for updates

Abstract: A total of 1327 platinum-quality mitochondrial DNA haplotypes from United States (U.S.) populations were generated using a robust, semi-automated next-generation sequencing (NGS) workflow with rigorous quality control (QC). The laboratory workflow involved long-range PCR to minimize the co-amplification of nuclear mitochondrial DNA segments (NUMTs), PCR-free library preparation to reduce amplification bias, and high-coverage Illumina MiSeq sequencing to produce an average per-sample read depth of 1000 × for low-frequency (5%) variant detection. Point heteroplasmies below 10% frequency were confirmed through replicate amplification, and length heteroplasmy was quantitatively assessed using a custom read count analysis tool. Data analysis involved a redundant, dual-analyst review to minimize errors in haplotype reporting with additional QC checks performed by EMPOP. Applying these methods, eight sample sets were processed from five U.S. metapopulations (African American, Caucasian, Hispanic, Asian American, and Native American) corresponding to self-reported identity at the time of sample collection. Population analyses (e.g., haplotype frequencies, random match probabilities, and genetic distance estimates) were performed to evaluate the eight datasets, with over 95% of haplotypes unique per dataset. The platinum-quality mitogenome haplotypes presented in this study will enable forensic statistical calculations and thereby support the usage of mitogenome sequencing in forensic laboratories.

Keywords: mtDNA; mitogenome; next-generation sequencing; haplotype; haplogroup; population statistics

Project Aims

Generate high-quality data

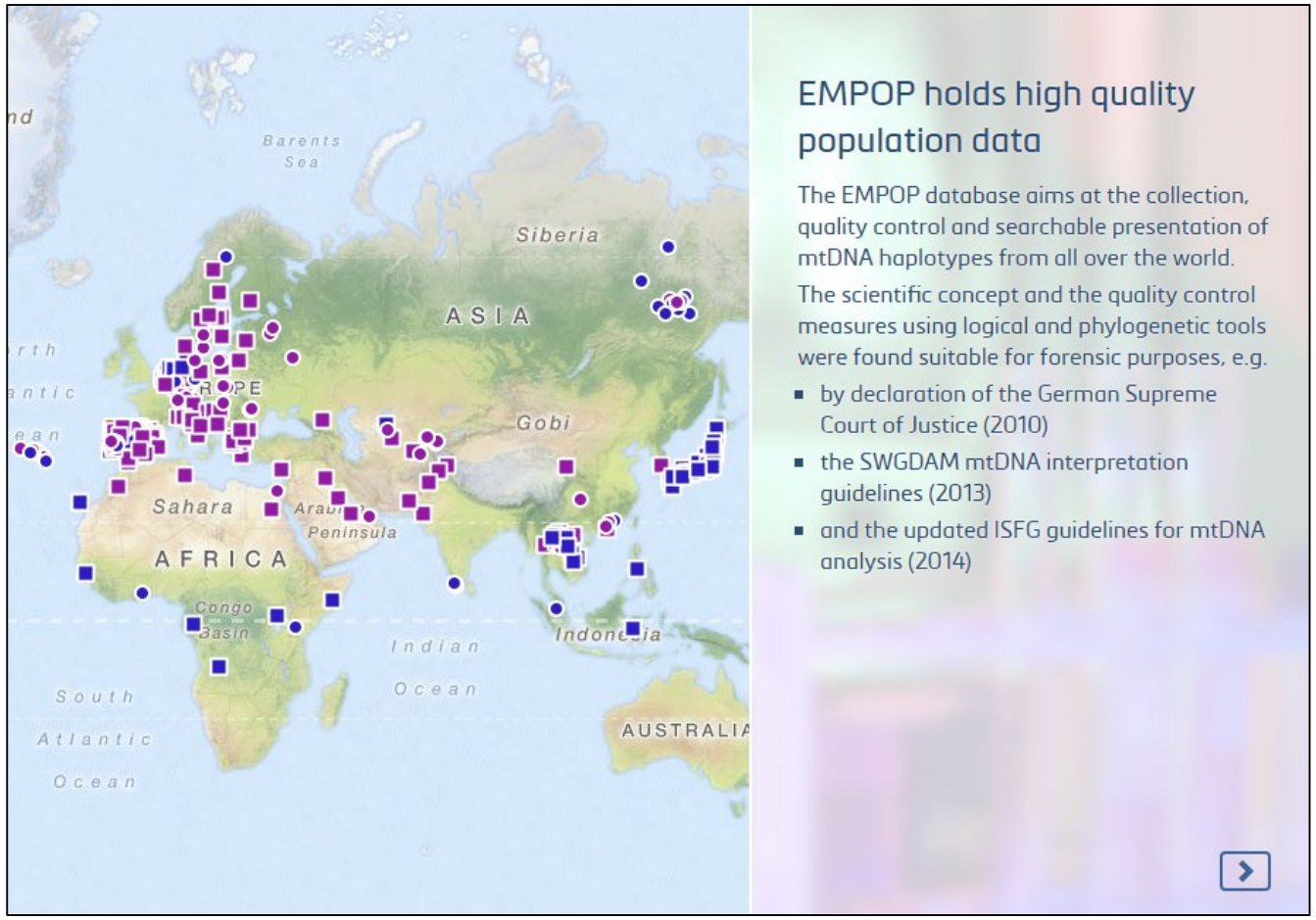
- CLC Bio Genomics Workbench Plugin
  - “AQME” developed by AFDIL
- Double review of variant calls
- Confirmation of heteroplasmy < 10 %

Phylogenetic QC by EMPOP

- International mtDNA database
- Identify unlikely variants

Large dataset of mtGenomes

- 1,327 total passed QC
- Searchable in EMPOP
- Enable match statistics (mtGenome)



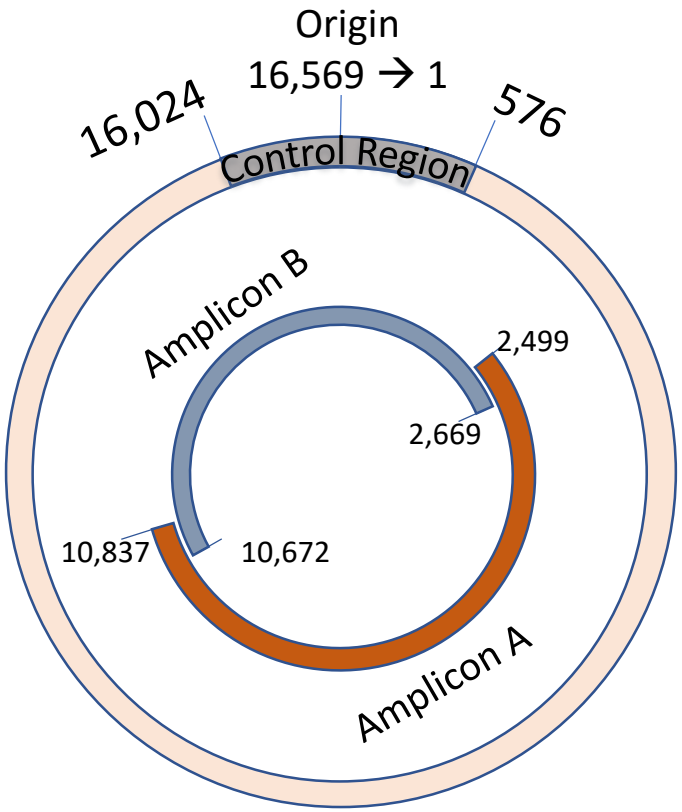
Long-PCR Workflow

PCR
- Two long amplicons ~ 8.5 kb each
- QC +/- on AATI Fragment Analyzer

Library Preparation
- Fragmentation (enzymatic) of PCR products
- Ligate adaptors/barcodes

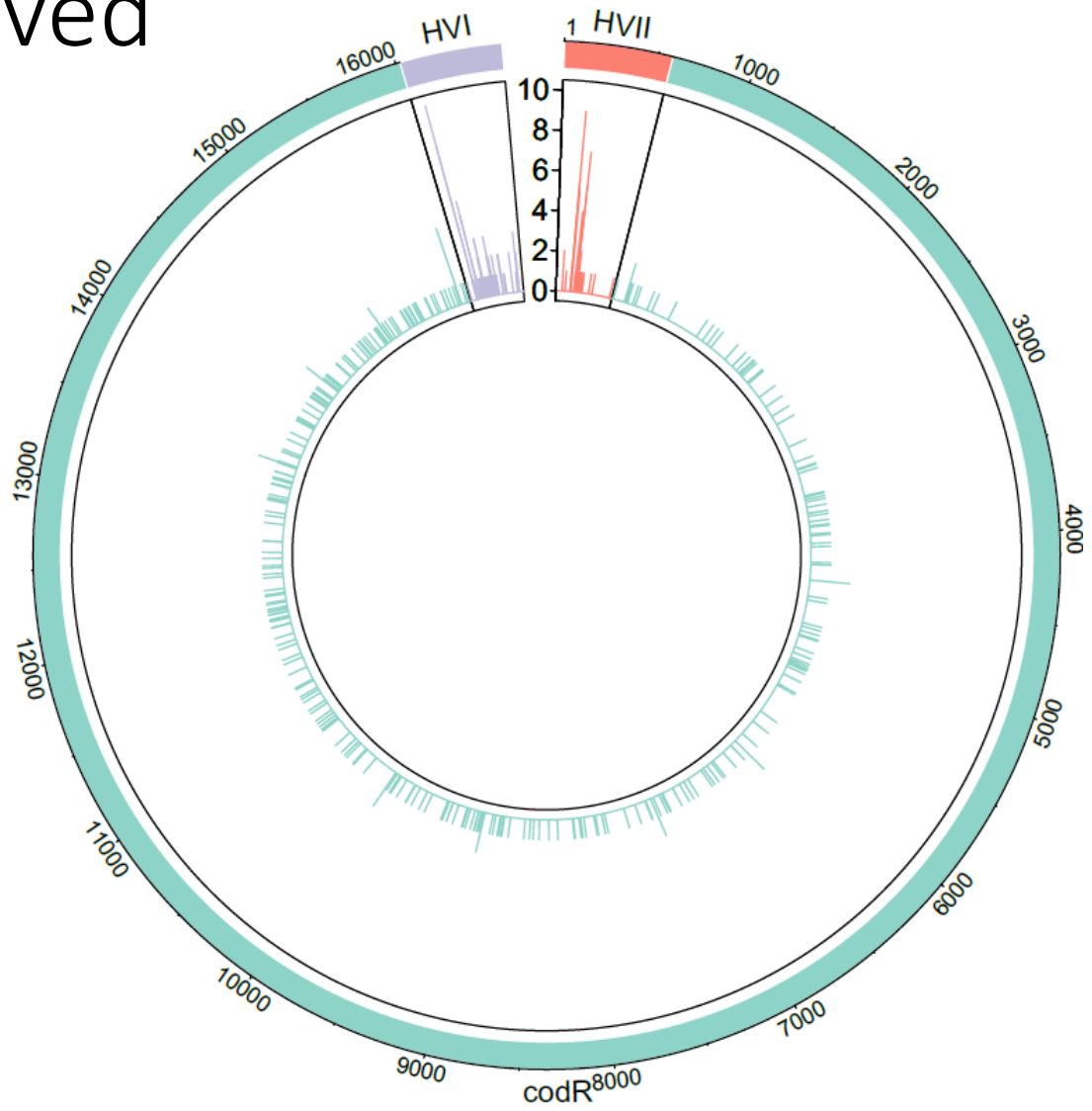
Sequencing
- QC & Quantitate libraries on AATI F/A
- Run on MiSeq FGx

Data Analysis
- Align to reference genome (rCRS)
- Dual Review QC of data



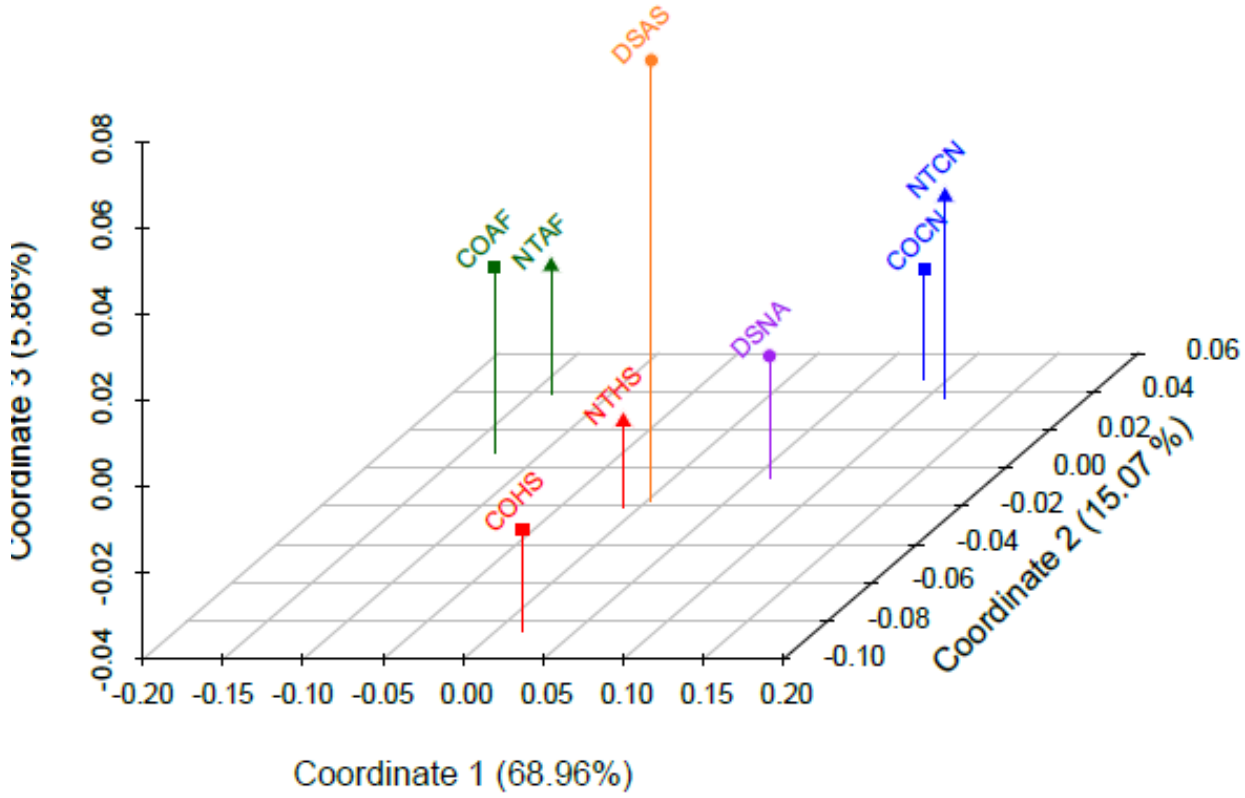
Heteroplasmy Observed

Dataset	Total Individuals	Total PHPs	Individuals with PHPs	Individuals with 1 PHP	Individuals with 2 PHPs	Individuals with 3 PHPs
COAF	112	37	31 (28 %)	26	4	1
COCN	112	41	30 (27 %)	20	9	1
COHS	109	36	27 (25 %)	20	5	2
NTAF	256	77	60 (23 %)	43	17	0
NTCN	260	92	77 (30 %)	65	10	2
NTHS	138	53	43 (31 %)	34	8	1
DSAS	169	62	54 (32 %)	49	2	3
DSNA	171	48	43 (25 %)	38	5	0
All	1327	446	365 (28 %)	295	60	10

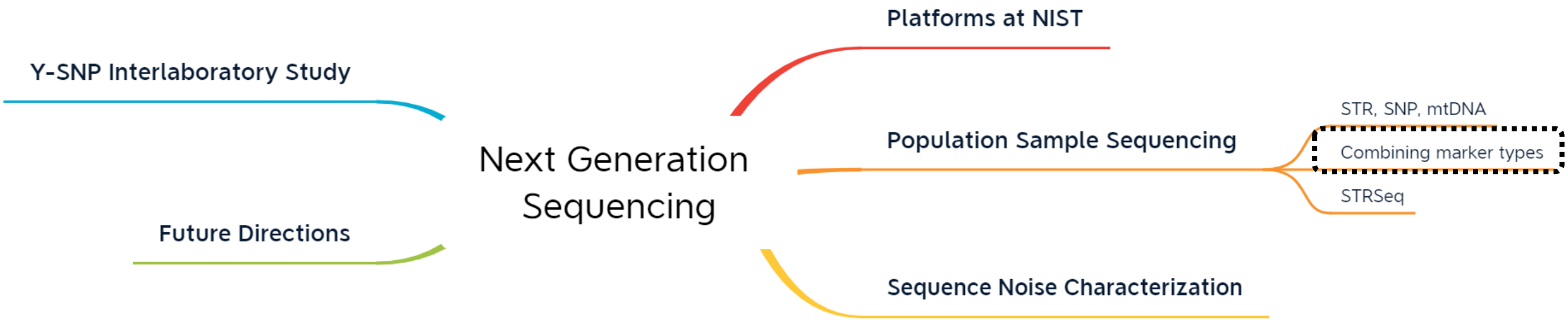


Population Pairwise F_{st} Comparisons

- Similar population ‘samplings’
  - African American and U.S. Caucasian
  - Homogeneous across geographic samples
- Differences in U.S. Hispanics
  - Could be due to site of collection
  - Western U.S. vs Eastern



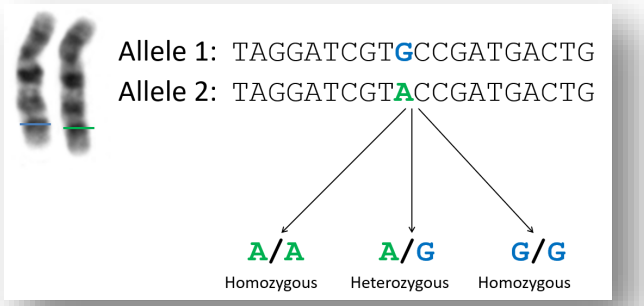
Topics for today



Combining Marker Types

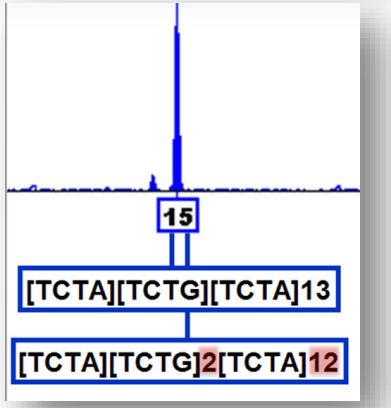
We calculate Match Statistics by multiplying allele frequencies across markers

- Requires markers to be independent



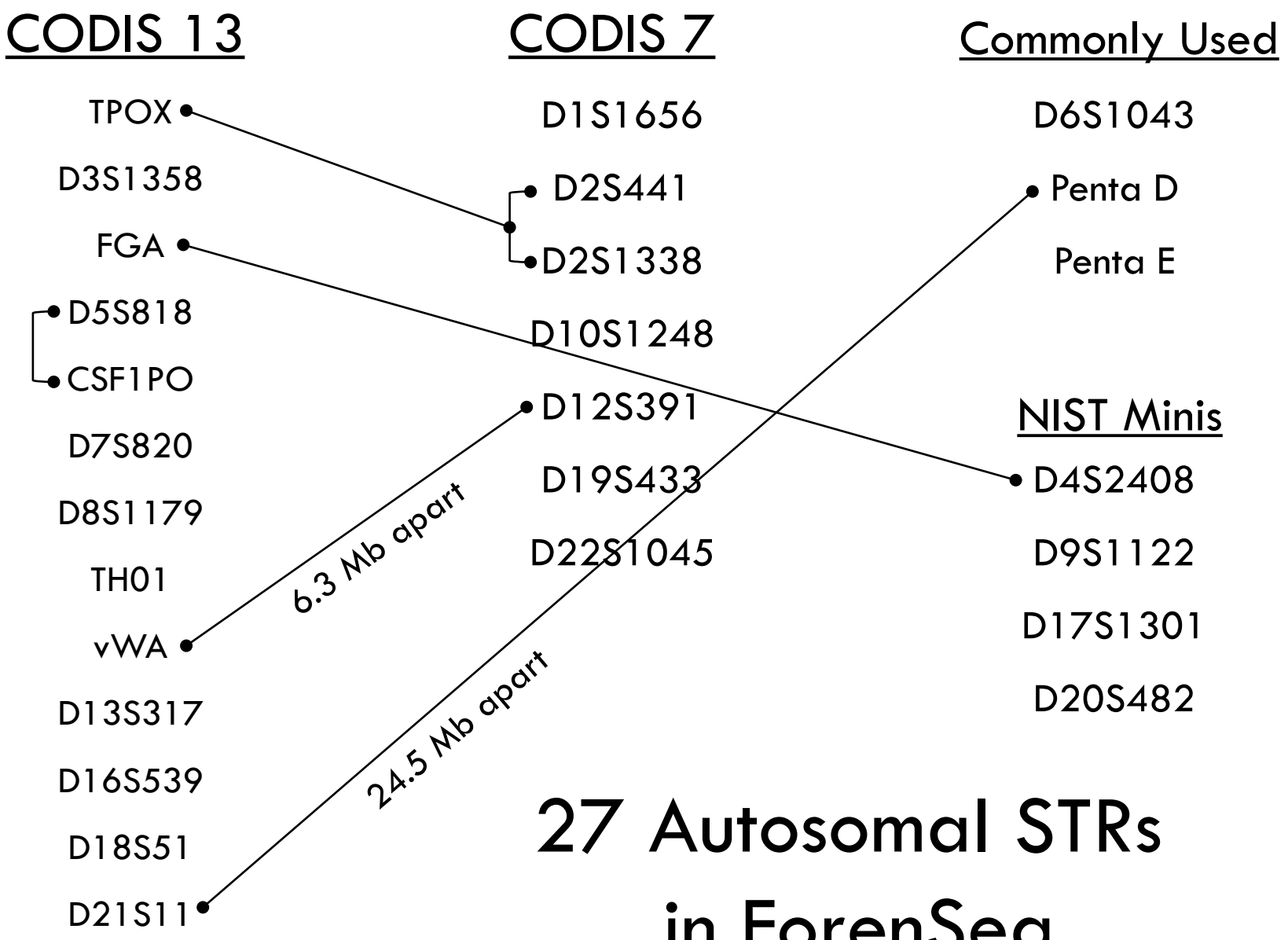
Sequencing allows typing more markers and different marker types

Laboratories need guidance on calculating the appropriate match statistics



Evaluating Linkage Disequilibrium in auSTR and IISNP loci in NIST 1036

- Collaboration with Andreas Tillmar, National Board of Forensic Medicine



27 Autosomal STRs in ForenSeq

94 IISNPs in ForenSeq

Electrophoresis 2006, *27*, 1713–1724

Juan J. Sanchez[1]
Chris Phillips[2]
Claus Børsting[1]
Kinga Balogh[3]
Magdalena Bogus[3]
Manuel Fondevila[2]
Cheryl D. Harrison[4]
Esther Musgrave-Brown[4]
Antonio Salas[2]
Denise Syndercombe-Court[4]
Peter M. Schneider[3]
Angel Carracedo[2]
Niels Morling[1]

[1]Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark
[2]Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
[3]Institute of Legal Medicine, Johannes Gutenberg University, Mainz, Germany
[4]Centre for Haematology, ICMS, Barts and The London, Queen Mary's School of Medicine and Dentistry, London, UK

Received September 10, 2005
Revised October 15, 2005
Accepted October 16, 2005

Research Article

A multiplex assay with 52 single nucleotide polymorphisms for human identification

A total of 52 SNPs reported to be polymorphic in European, Asian and African populations were selected. Of these, 42 were from the distal regions of each autosome (except chromosome 19). Nearly all selected SNPs were located at least 100 kb distant from known genes and commonly used STRs. We established a highly sensitive and reproducible SNP-typing method with amplification of all 52 DNA fragments in one PCR reaction followed by detection of the SNPs with two single base extension reactions analysed using CE. The amplicons ranged from 59 to 115 bp in length. Complete SNP profiles were obtained from 500 pg DNA. The 52 loci were efficiently amplified from degraded samples where previously only partial STR profiles had been obtained. A total of 700 individuals from Denmark, Greenland, Somalia, Turkey, China, Germany, Taiwan, Thailand and Japan were typed, and the allele frequencies estimated. All 52 SNPs were polymorphic in the three major population groups. The mean match probability was at least 5.0×10^{-19} in the populations studied. Typical paternity indices ranged from 336 000 in Asians to 549 000 in Europeans. Details of the 52 SNP loci and population data generated in this work are freely available at http://www.snpforid.org.

Keywords: Autosomes / Human identification / Multiplex PCR / Single base extension / Single nucleotide polymorphism
DOI 10.1002/elps.200500671

1 Introduction

SNPs have a number of characteristics that make them ideal markers for human identification. First, they have lower mutation rates than the STR and VNTR (variable number tandem repeat) loci typically used for relationship analysis in paternity and immigration testing. Second, SNPs can be analysed after PCR amplification of very short DNA-regions surrounding the substitution site, making SNPs preferable for anthropological and crime case investigations where the DNA is often degraded. Third, SNPs can be genotyped with a growing range of high-throughput technologies; an important factor in the implementation of large criminal DNA databases [1, 2]. Finally, SNPs, as binary polymorphisms, are comparatively easy to validate, because precise allele frequency estimates, required for the accurate interpretation of forensic genotyping data, can be obtained by analysing fewer samples compared to those needed for allele frequencies estimates of STRs and VNTRs. Seeking to match the discriminatory power of the 10–15 multiple allele STRs routinely used in forensic investigations, a set of about 50 polymorphic SNP markers are predicted to be required [3, 4]. Furthermore, it has been suggested that 50 unlinked SNP loci with high overall heterozygosity should be sufficient to adjust for population stratification in population-based associations studies [5]. SNPs that are polymorphic in one population may be almost or completely monomorphic in another population [6, 7], while others are known to be polymorphic in all major population groups. Thus, it should be possible to select SNPs that are useful for human identification purposes in the majority of populations, and to supplement these with SNPs

Correspondence: Dr. Juan J. Sanchez, Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, 11 Frederik V's Vej, DK-2100 Copenhagen, Denmark
E-mail: juan.sanchez@forensic.ku.dk
Fax: +45-35-32-61-20

Abbreviations: **RFU**, relative fluorescence unit; **SBE**, single base extension

50 from SNPforID

44 from Kidd

LD tests for pairs of SNPs on same chromosome demonstrated no significant deviation from expectations

Hum Genet (2010) 127:315–324
DOI 10.1007/s00439-009-0771-1

ORIGINAL INVESTIGATION

SNPs for a universal individual identification panel

Andrew J. Pakstis · William C. Speed · Rixun Fang · Fiona C. L. Hyland · Manohar R. Furtado · Judith R. Kidd · Kenneth K. Kidd

Received: 9 September 2009 / Accepted: 13 November 2009 / Published online: 24 November 2009

Abstract An efficient method to uniquely identify every individual would have value in quality control and sample tracking of large collections of cell lines or DNA as is now often the case with whole genome association studies. Such a method would also be useful in forensics. SNPs represent the best markers for such purposes. We have developed a globally applicable resource of 92 SNPs for individual identification (IISNPs) with extremely low probabilities of any two unrelated individuals from anywhere in the world having identical genotypes. The SNPs were identified by screening over 500 likely/candidate SNPs on samples of 44 populations representing the major regions of the world. All 92 IISNPs have an average heterozygosity >0.4 and the F_{st} values are all <0.06 on our 44 populations making these a universally applicable panel irrespective of ethnicity or ancestry. No significant linkage disequilibrium (LD) occurs for all unique pairings of 86 of the 92 IISNPs (median LD = 0.011) in all of the 44 populations. The remaining 6 IISNPs show strong LD in most of the 44 populations for a small subset (7) of the unique pairings in which they occur due to close linkage. 45 of the 86 SNPs are spread across the 22 human autosomes and show very loose or no genetic linkage with each other. These 45 IISNPs constitute an excellent panel for individual identification including paternity testing with associated probabilities of individual genotypes less than 10^{-15}, smaller than achieved with the current panels of forensic markers. This panel also improves on an interim panel of 40 IISNPs previously identified using 40 population samples. The unlinked status of the subset of 45 SNPs we have identified also makes them useful for situations involving close biological relationships. Comparisons with random sets of SNPs illustrate the greater discriminating power, efficiency, and more universal applicability of this IISNP panel to populations around the world. The full set of 86 IISNPs that do not show LD can be used to provide even smaller genotype match probabilities in the range of 10^{-31}–10^{-35} based on the 44 population samples studied.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-009-0771-1) contains supplementary material, which is available to authorized users.

A. J. Pakstis · W. C. Speed · J. R. Kidd · K. K. Kidd (✉)
Department of Genetics, Yale University School of Medicine, 333 Cedar Street, 208005, New Haven, CT 06520, USA
e-mail: kenneth.kidd@yale.edu

R. Fang · F. C. L. Hyland · M. R. Furtado
Applied Markets, Applied Biosystems/Life Technologies, Foster City, CA 94404, USA

Introduction

In previous papers (Kidd et al. 2006; Pakstis et al. 2007), we described the rationale and our strategy for developing a panel of SNPs for individual identification (IISNPs) and presented some potentially useful IISNPs. Such a panel would have use in sample tracking in large collections of human DNA samples and in forensics and paternity testing. Others have also addressed the value of such panels in forensics (Inagaki et al. 2004; Lee et al. 2005; Sanchez et al. 2006; Butler et al. 2008; Pakstis et al. 2008). One panel of 52 SNPs has been accepted for forensic use in several European countries (Sanchez et al. 2006; Phillips et al. 2009). An IISNP panel would provide a complementary tool for forensic applications in situations, such as highly degraded DNA (e.g., Fang et al. 2009), in which the standard STR markers of the widely used *CO*mbined *D*NA *I*ndex *S*ystem (CODIS) panel do not perform well. SNPs also offer a potentially cheaper,

45 SNPs are spread across the 22 human autosomes and show very loose or no genetic linkage with each other

LD in casework

When LD is detected, ideally:

- Rule out technical issues by testing on different platforms/assays
- Confirm with multiple sample sets from same population, and multiple test methods

Designing panel/assay: Evaluate LD, eliminate loci as needed based on informativeness

Implementing established panel/assay:

Best – Determine haplotype frequency for pair or block

- for polymorphic loci the sample size would be unfeasible

Alternative – Exclude one of the two markers during validation

- Keep the more informative, similar to assay design

Problematic – Exclude one of the two markers case-by-case

- RMP vs Kinship

STRSeq Catalog of Sequences

Forensic Science International: Genetics 31 (2017) 111–117

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

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

Katherine Butler Gettings[a,*], Lisa A. Borsuk[a], David Ballard[b], Martin Bodner[c], Bruce Budowle[d,e], Laurence Devesse[b], Jonathan King[d], Walther Parson[c,f], Christopher Phillips[g], Peter M. Vallone[a]

[a] U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[b] King's Forensics, King's College London, Franklin-Wilkins Building, 150 Stamford Street London, UK
[c] Institute of Legal Medicine, Medical University of Innsbruck, Austria
[d] Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[e] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University Jeddah, Saudi Arabia
[f] Forensic Science Program, The Pennsylvania State University, USA
[g] Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Spain

The STR Sequencing Project (human) Accession: PRJNA380127 ID: 380127

The purpose of STRSeq is to facilitate the description of sequence-based alleles at the Short Tandem Repeat (STR) loci targeted in human identification assays. This collaborative effort of the international forensic DNA community, which has been endorsed by the executive board of the ISFG (International Society of Forensic Genetics), provides a framework for communication among laboratories. Each record contains: (a) observed sequence of an STR region, (b) annotation of the repeat region ("bracketing") and flanking region polymorphisms, (c) information regarding the sequencing assay and data quality, and (d) backward compatible length-based allelic designation. Data within the umbrella project is organized into locus sub-projects, and can be accessed by browsing, BLAST searching, or ftp download at NCBI. For comments or questions, please contact strseq@nist.gov.

Accession	PRJNA380127
Type	Umbrella project
Publications (total 5)	1. Borsuk LA *et al.*, "Sequence-based US population data for the SE33 locus.", *Electrophoresis*, 2018 Nov;39(21):2694-2701 More...
Submission	Registration date: 22-Mar-2017 National Institute of Standards and Technology
Related Resources	• STRSeq • STRidER
Relevance	Human Identification

Project Data:

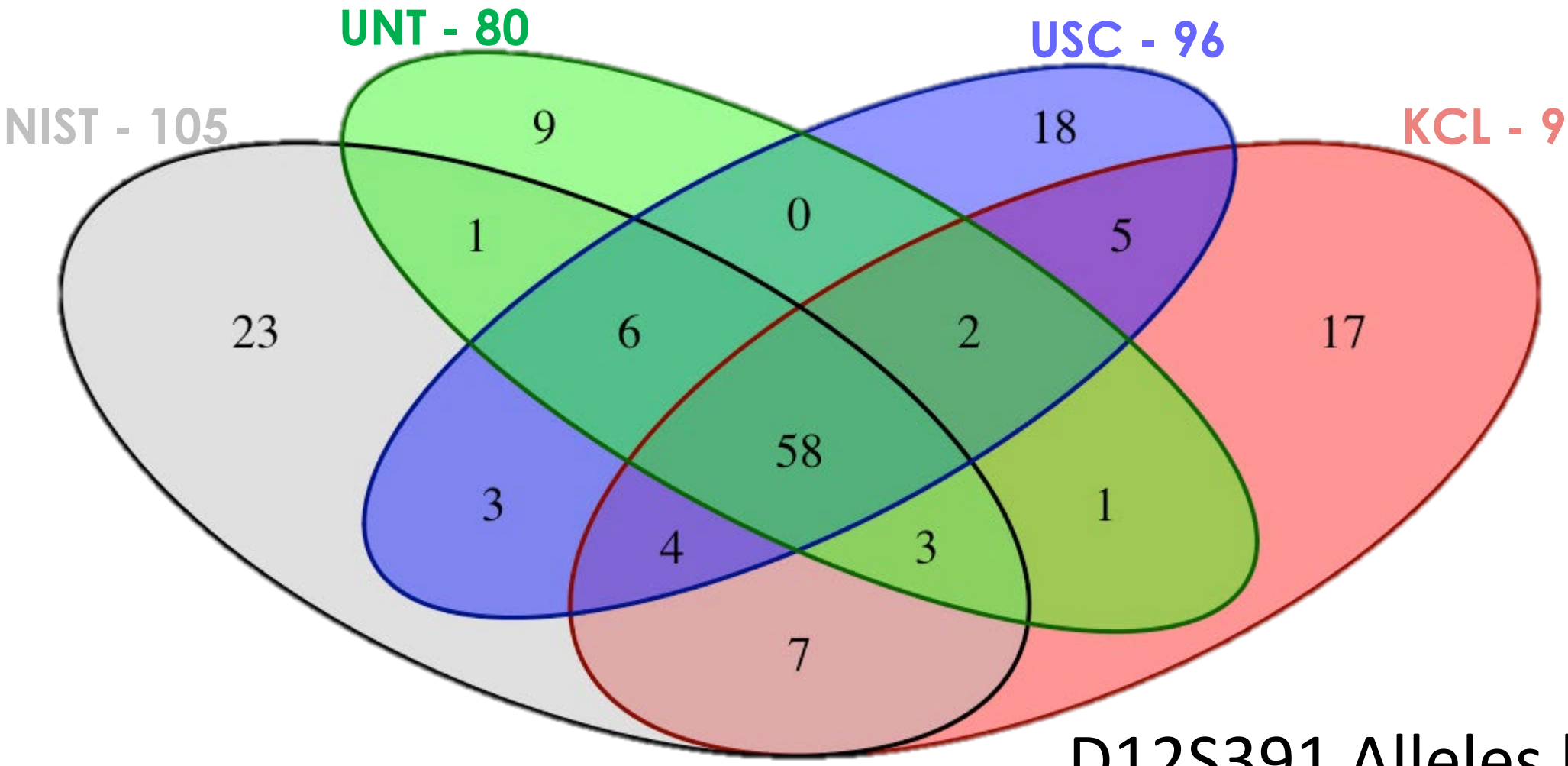
Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	2047
PUBLICATIONS	
PubMed	5
PMC	3

The STR Sequencing Project (human) encompasses the following 4 sub-projects:

Project Type	Number of Projects
Umbrella project	4

BioProject accession	Name	Title
PRJNA380345	Homo sapiens	STRSeq Commonly Used Autosomal STR Loci (National Institute of Standards...)
PRJNA380346	Homo sapiens	STRSeq Alternate Autosomal STR Loci (National Institute of Standards...)
PRJNA380347	Homo sapiens	STRSeq Y-Chromosomal STR Loci (National Institute of Standards...)
PRJNA380348	Homo sapiens	STRSeq X-Chromosomal STR Loci (National Institute of Standards...)

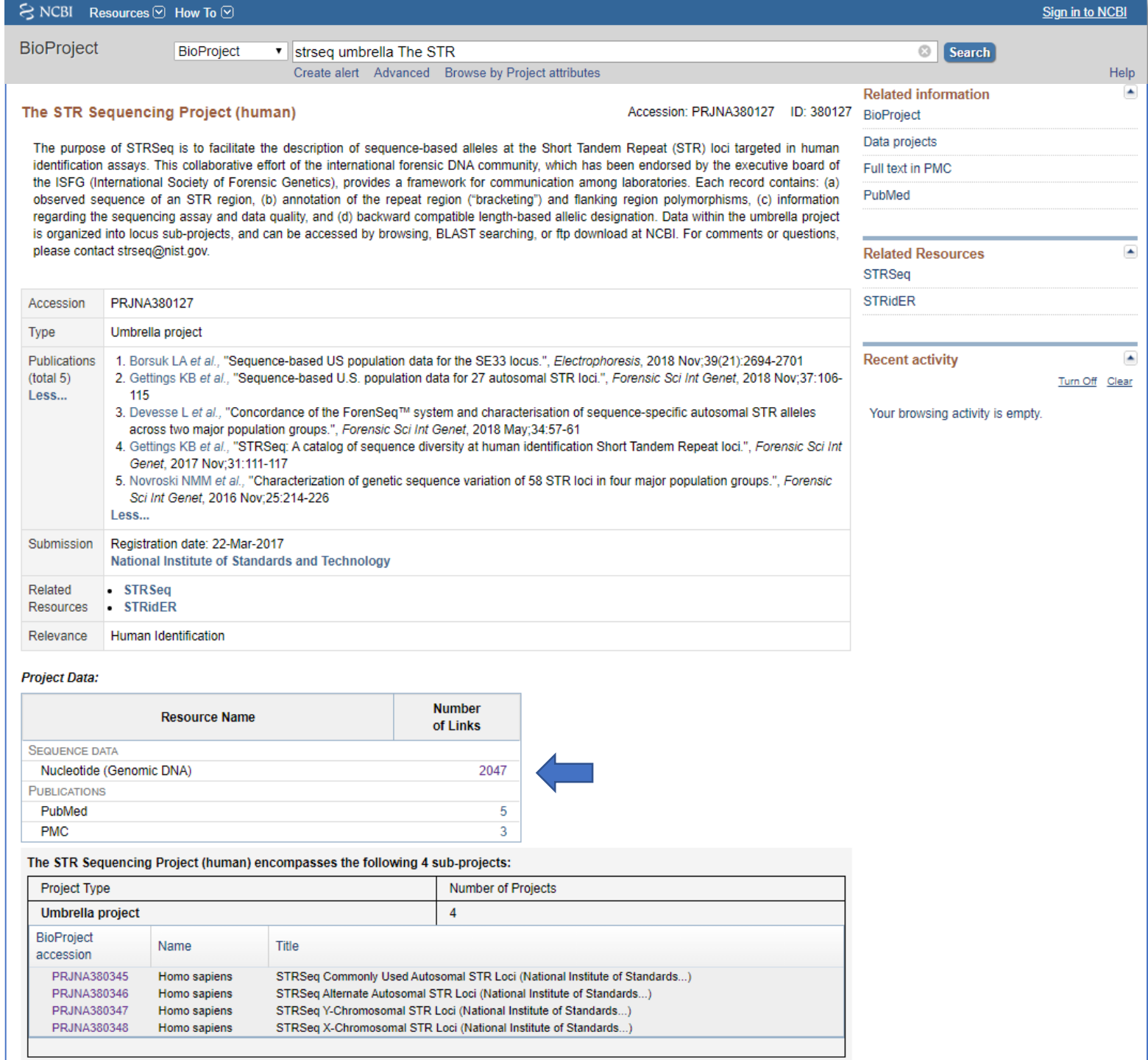
STRSeq Catalog of Sequences



D12S391 Alleles by Lab

STRSeq
Catalog of Sequences

https://www.ncbi.nlm.nih.gov/bioproject/380127



NCBI Resources How To Sign in to NCBI

BioProject BioProject strseq umbrella The STR Search

Create alert Advanced Browse by Project attributes Help

The STR Sequencing Project (human)

Accession: PRJNA380127 ID: 380127

The purpose of STRSeq is to facilitate the description of sequence-based alleles at the Short Tandem Repeat (STR) loci targeted in human identification assays. This collaborative effort of the international forensic DNA community, which has been endorsed by the executive board of the ISFG (International Society of Forensic Genetics), provides a framework for communication among laboratories. Each record contains: (a) observed sequence of an STR region, (b) annotation of the repeat region ("bracketing") and flanking region polymorphisms, (c) information regarding the sequencing assay and data quality, and (d) backward compatible length-based allelic designation. Data within the umbrella project is organized into locus sub-projects, and can be accessed by browsing, BLAST searching, or ftp download at NCBI. For comments or questions, please contact strseq@nist.gov.

Accession	PRJNA380127
Type	Umbrella project
Publications (total 5) Less...	1. Borsuk LA *et al.*, "Sequence-based US population data for the SE33 locus.", *Electrophoresis*, 2018 Nov;39(21):2694-2701 2. Gettings KB *et al.*, "Sequence-based U.S. population data for 27 autosomal STR loci.", *Forensic Sci Int Genet*, 2018 Nov;37:106-115 3. Devesse L *et al.*, "Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups.", *Forensic Sci Int Genet*, 2018 May;34:57-61 4. Gettings KB *et al.*, "STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci.", *Forensic Sci Int Genet*, 2017 Nov;31:111-117 5. Novroski NMM *et al.*, "Characterization of genetic sequence variation of 58 STR loci in four major population groups.", *Forensic Sci Int Genet*, 2016 Nov;25:214-226 Less...
Submission	Registration date: 22-Mar-2017 National Institute of Standards and Technology
Related Resources	• STRSeq • STRidER
Relevance	Human Identification

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	2047
PUBLICATIONS	
PubMed	5
PMC	3

The STR Sequencing Project (human) encompasses the following 4 sub-projects:

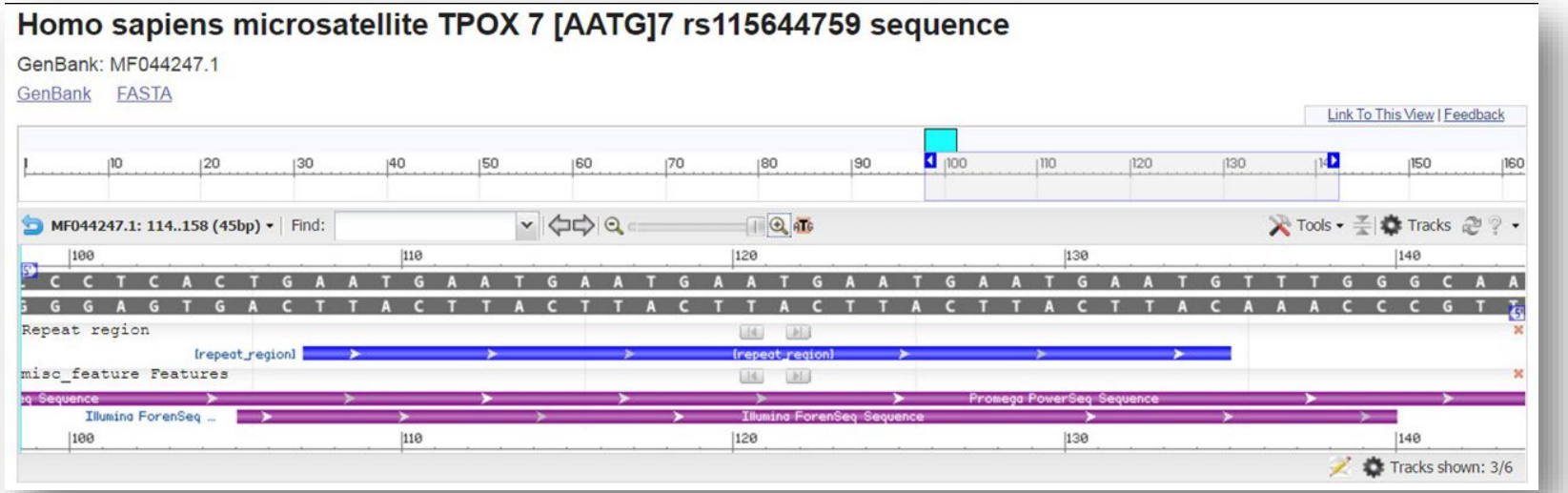
Project Type		Number of Projects
Umbrella project		4
BioProject accession	Name	Title
PRJNA380345	Homo sapiens	STRSeq Commonly Used Autosomal STR Loci (National Institute of Standards...)
PRJNA380346	Homo sapiens	STRSeq Alternate Autosomal STR Loci (National Institute of Standards...)
PRJNA380347	Homo sapiens	STRSeq Y-Chromosomal STR Loci (National Institute of Standards...)
PRJNA380348	Homo sapiens	STRSeq X-Chromosomal STR Loci (National Institute of Standards...)

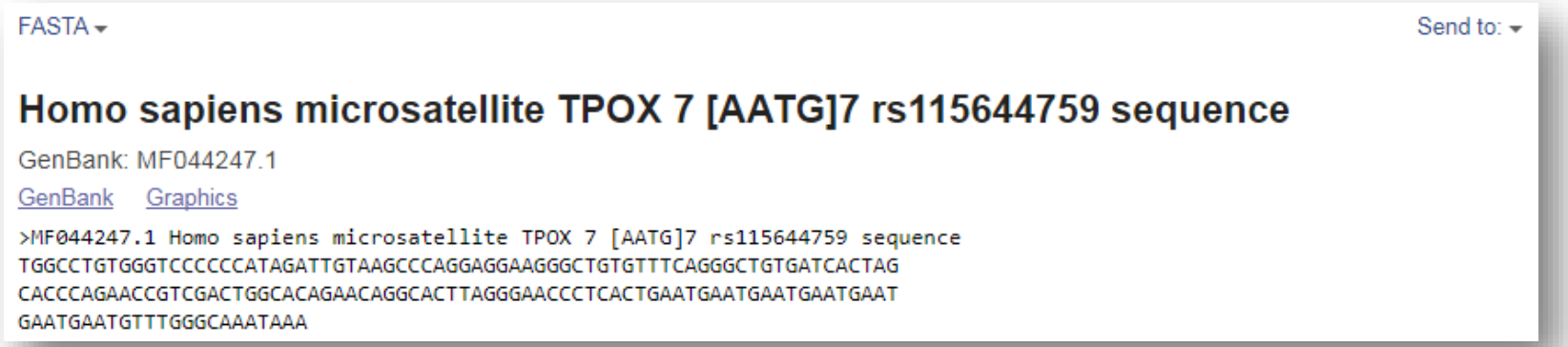
Related information
BioProject
Data projects
Full text in PMC
PubMed

Related Resources
STRSeq
STRidER

Recent activity
Turn Off Clear
Your browsing activity is empty.

STRSeq Catalog of Sequences





STRSeq Catalog of Sequences

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

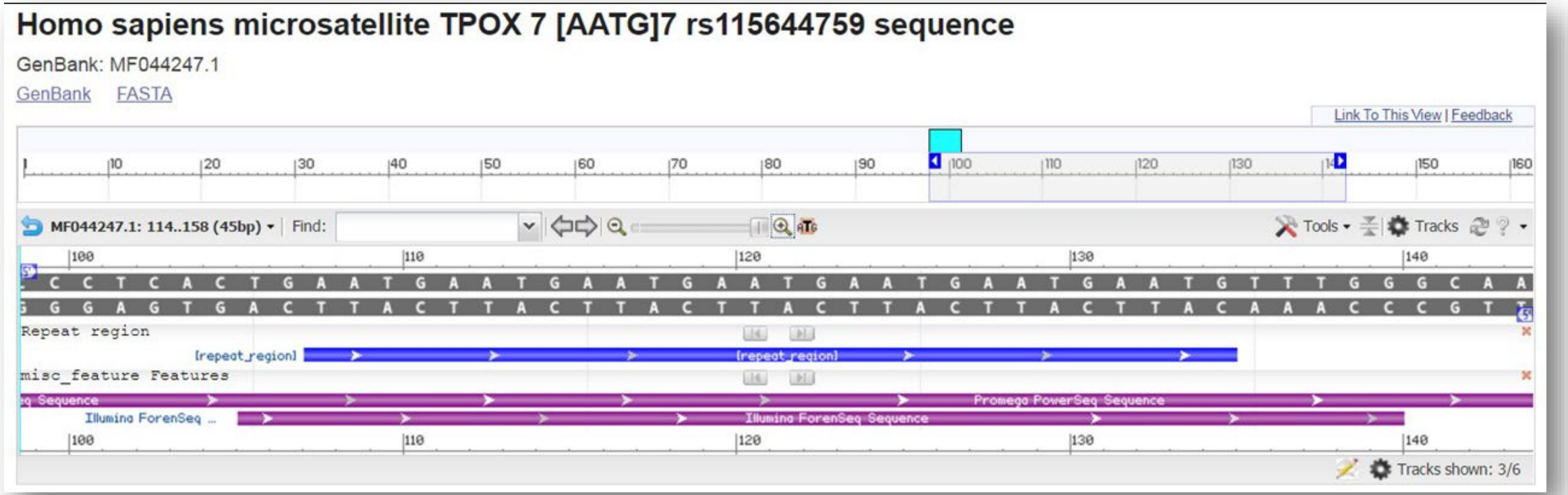
GenBank: MF044247.1

FASTA Graphics

Go to:

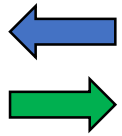
```
LOCUS       MF044247               163 bp    DNA     linear   PRI 30-MAY-2017
DEFINITION  Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence.
ACCESSION   MF044247
VERSION     MF044247.1
DBLINK      BioProject: PRJNA380554
KEYWORDS    STRSeq, STR, TPOX.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Gettings,K.B., Borsuk,L.A. and Vallone,P.M.
  TITLE     The STR Sequencing Project [manuscript in preparation]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163)
  AUTHORS   NIST,A.G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2017) Applied Genetics Group, National Institute
            of Standards and Technology, 100 Bureau Drive, MS-8314,
            Gaithersburg, MD 20899, USA
COMMENT     Annotation ('bracketing') of the repeat region is consistent with
            the guidance of the ISFG (International Society of Forensic
            Genetics), PMID: 26844919.  Lower case letters in the 'Bracketed
            repeat' region below denote uncounted bases.  The given
            length-based allele value was determined using the designated
            length-based technology.  Variation in the length-based allele
            between individuals or assays can result from indels in flanking
            regions.  The length of reported sequence is dependent on the assay
            (see 'Sequencing technology') and the quality of the flanking
            sequence. This information is provided as part of the STR
            Sequencing Project (STRseq), a collaborative effort of the
            international forensic DNA community.  The purpose of this project
            is to facilitate the description of sequence-based STR alleles.
            Additional resources can be found at strseq.nist.gov.  For
            questions or feedback, please contact strseq@nist.gov.  Allele
            frequency data can be accessed in the strider.online database.
            ##HumanSTR-START##
            STR locus name        :: TPOX
            Length-based allele   :: 7
            Bracketed repeat      :: [AATG]7
            Sequencing technology :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
            Coverage              :: >30X
            Length-based tech.    :: PowerPlex Fusion, ABI3500xl
            Assembly              :: GRCh38 (GCF_000001405)
            Chromosome            :: 2
            RefSeq Accession      :: NC_000002.12
            Chrom. Location       :: 1489532..1489698
            Repeat Location       :: 1489653..1489684
            Cytogenetic Location  :: 2p25.3
            ##HumanSTR-END##
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

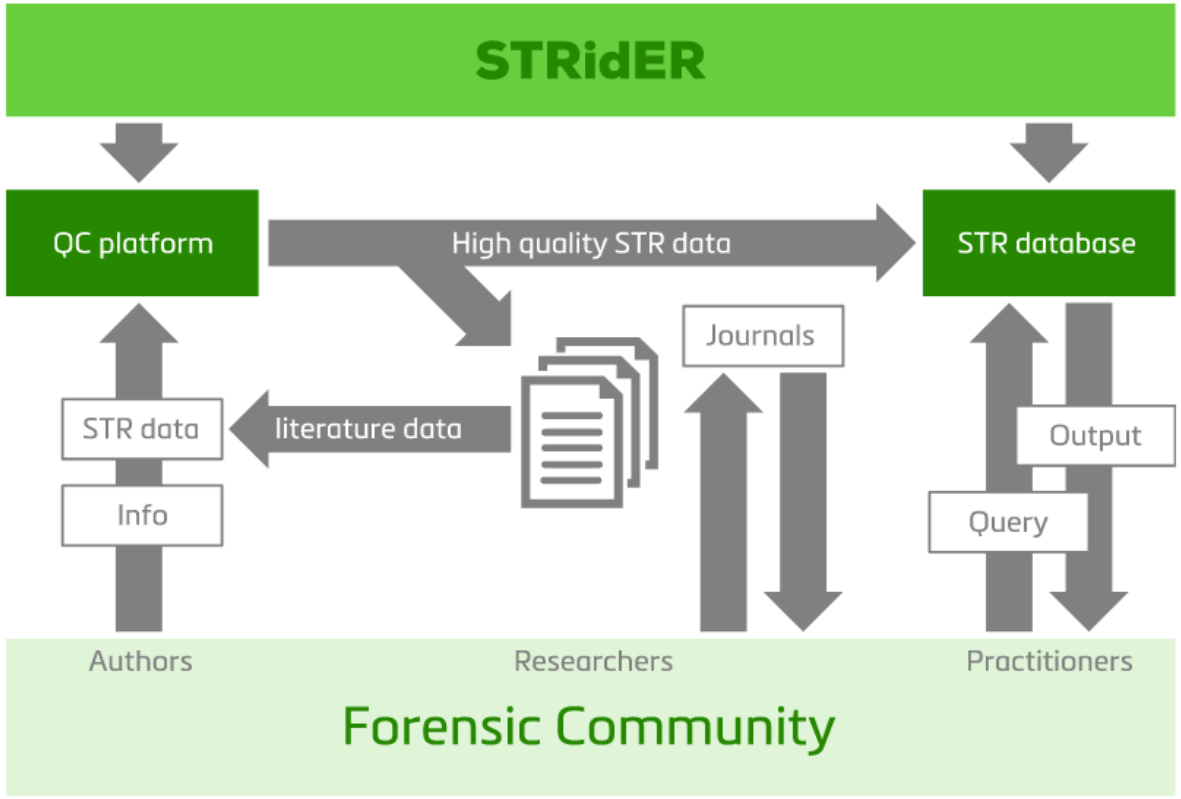
STRSeq Catalog of Sequences



STRSeq in Population Data

```
            ##HumanSTR-START##
            STR locus name        :: TPOX
            Length-based allele   :: 7
            Bracketed repeat      :: [AATG]7
            Sequencing technology :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
            Coverage              :: >30X
            Length-based tech.    :: PowerPlex Fusion, ABI3500xl
            Assembly              :: GRCh38 (GCF_000001405)
            Chromosome            :: 2
            RefSeq Accession      :: NC_000002.12
            Chrom. Location       :: 1489532..1489698
            Repeat Location       :: 1489653..1489684
            Cytogenetic Location  :: 2p25.3
            ##HumanSTR-END##
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```



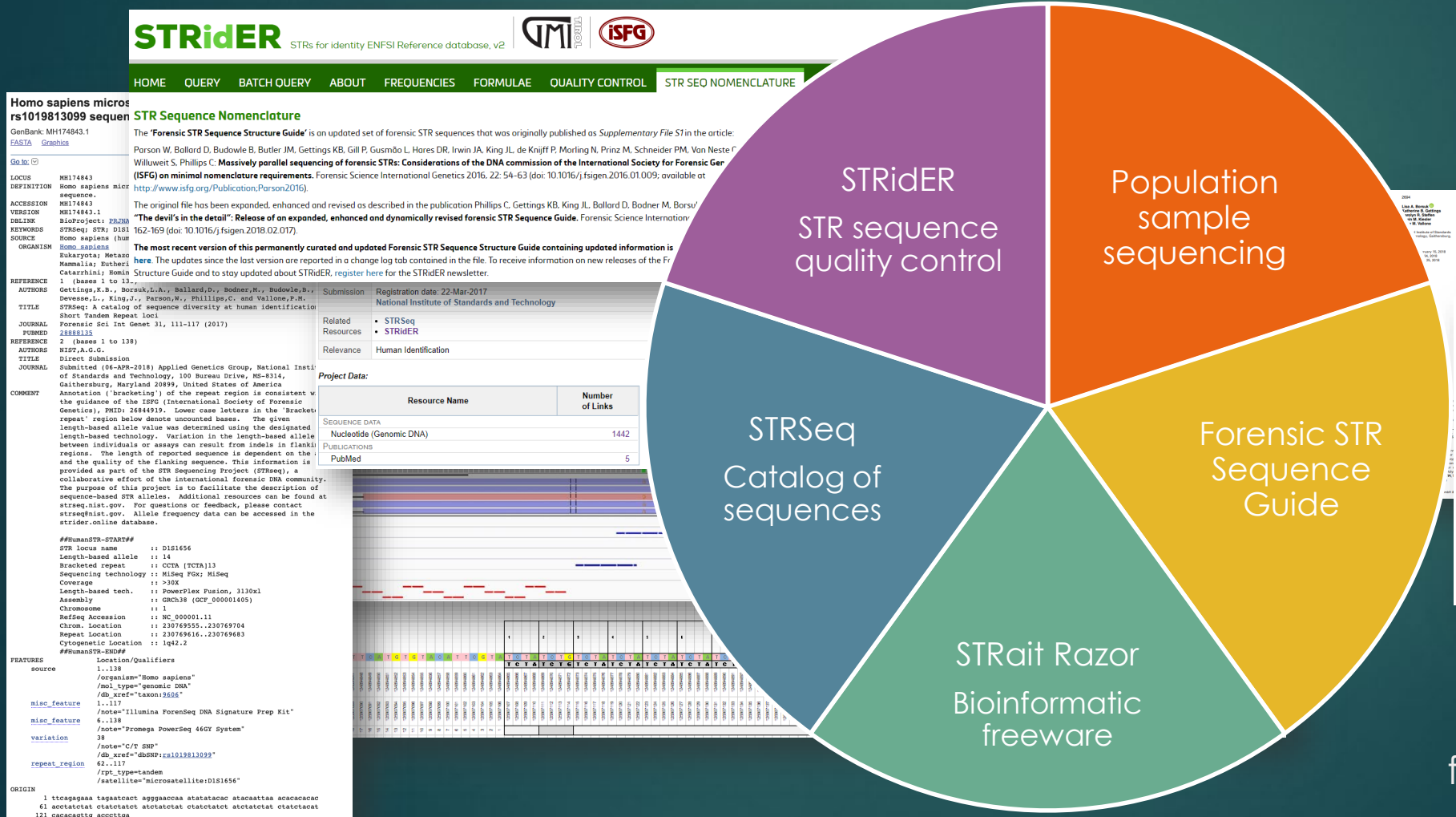


STRidER in the field of forensic STR typing (from Bodner et al. 2016)

STRAND *working group*

align | name | define

Our mission is to harmonize related efforts across member laboratories:



and to characterize additional STR loci present in the genome which may be useful for forensic purposes in the future.

STR Nomenclature Meeting

April 2019

Spring 2020:
ISFG EB approved
STRAND WG proposal for
DNA Commission on STR
Sequence Nomenclature
Recommendations

5' to 3':
Walther Parson, Lisa Borsuk, Peter Schneider, Brian Young, Rebecca Just, Jodi Irwin, David Ballard, Sascha Willuweit, Cydne Holt, Chris Phillips, Jonathan King, Tunde Huszar, Peter Gill, Christian Sell, Kris Van der Gaag, Laurence Devesse, Claus Borsting, Doug Hares, Katherine Gettings, Rob Lagace, Jerry Hoogenboom, Martin Bodner, Peter deKnijff, Sebastian Ganschow, Pedro Barrio, Teresa Gross

STRAND *working group*
align | name | define

D.Ballard:KCL,UK | M.Bodner&W.Parson:GMI,Austria | L.Borsuk&K.Gettings:NIST,US | J.King:UNT,US | C.Phillips:USC,Spain

Table 1
Publications containing STR sequence population data.

Citation	Year	First Author	Total Number of Samples	Populations	Sequenced STR Loci	Additional Data	Bioinformatic Method(s)
[6]	2016	Novroski	777	Caucasian Hispanic African American East Asian	27 Autosomal STR 24 Y-STR 7 X-STR	CE-STR	ForenSeq UAS STRait Razor v2.0
[21]	2016	van der Gaag	297	Netherlands Nepal Bhutan Central African Pygmy	17 Autosomal STR	CE-STR	TSSV (FDSTools)
[22,23]	2016, 2017	Wendt	62	Yavapai	27 Autosomal STR 24 Y-STR 7 X-STR	94 iiSNP 56 aiSNP 22 piSNP	STRait Razor v2s
[24]	2017	Casals	231	Spanish Roma Catalans	27 Autosomal STR 24 Y-STR 7 X-STR	94 iiSNP	ForenSeq UAS
[25]	2017	Silva	59	South Brazilian	22 Autosomal STR 23 Y-STR	CE-STR	Altius Cloud System
[26]	2018	Borsuk	1036	Caucasian African American Hispanic Asian	1 Autosomal STR (SE33)	CE-STR	STRait Razor v2.0
[7]	2018	Devesse	400	White British British Chinese	27 Autosomal STR	CE-STR	ForenSeq UAS
[9]	2018	Gettings	1036	Caucasian African American Hispanic Asian	27 Autosomal STR	CE-STR	ForenSeq UAS STRait Razor v2.0
[27]	2018	Huszar	100	African European Australian Asian Near and Middle Eastern American	23 Y-STR	CE-STR	FDSTools v1.1.1
[28]	2018	Kim	209	Korean	27 Autosomal STR 24 Y-STR 7 X-STR	CE-STR	ForenSeq UAS
[8]	2018	Phillips	944	CEPH (51 populations)	27 Autosomal STR 24 Y-STR 7 X-STR	CE-STR	ForenSeq UAS
[29]	2018	Salvador	143	Filipino	7 X-STR	CE-STR	ForenSeq UAS STRait Razor v2s

Table 2
STR Sequence Analysis Software.

Name	Availability
Agnostic, freeware	
FDSTools [34]	Python Package Index; www.fdstools.nl
Seqmapper [35]	http://forensic.mc.ntu.edu.tw:9000/SEQMapperWeb/Default.aspx
STRait Razor v2s [3] STRait Razor 3.0 [36]	https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/strait-razor/
STRinNGS [37]	Upon request from the University of Copenhagen
ToaSTR [38]	https://www.toastr.de/
Agnostic, for purchase	
ExactID	https://www.battelle.org/government-offerings/homeland-security-public-safety/security-law-enforcement/forensic-genomics/exactid
GeneMarkerHTS	https://softgenetics.com/GeneMarkerHTS.php
Armed Expert Mixture Ace	https://nichevision.com/mixtureace/
Assay specific, for purchase	
Converge	https://www.thermofisher.com/order/catalog/product/A35131
Universal Analysis Software	https://verogen.com/products/

Forensic Science International: Genetics 43 (2019) 102165

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Short communication

Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting

Katherine Butler Gettings[a,*], David Ballard[b], Martin Bodner[c], Lisa A. Borsuk[a], Jonathan L. King[d], Walther Parson[c,e], Christopher Phillips[f]

[a] U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD, 20899, USA
[b] King's Forensics, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London, UK
[c] Institute of Legal Medicine, Medical University of Innsbruck, Austria
[d] Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX, 76107, USA
[e] Forensic Science Program, The Pennsylvania State University, USA
[f] Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Spain

ARTICLE INFO

Keywords:
STR
Sequence
Nomenclature
Bioinformatics

ABSTRACT

This report summarizes topics discussed at the STR sequence nomenclature meeting hosted by the STRAND Working Group in April 2019. Invited attendees for this meeting included researchers known-to-us to be developing STR sequence-based nomenclature schemata, scientific representatives from vendors developing STR sequence bioinformatic methods, DNA intelligence database curators, and academic experts in STR genomics. The goal of this meeting was to provide a forum for individuals developing nomenclature schemata to present and discuss their ideas, encouraging mutual awareness, identification of differences in approaches, opposing aspects, and opportunities for parallelization while some approaches are still under development.

1. Introduction

Since 2016, the *ad hoc* formed STR Sequence Working Group (the authorship of this publication) has been collaborating to harmonize related efforts across our respective laboratories, consisting of: STRidER STR sequence quality control [1], STRSeq catalog of sequences [2], STRait Razor bioinformatic freeware [3], the Forensic STR Sequence Structure Guide [4,5], and large-scale population sample sequencing efforts [6–9] (see [10] for a comprehensive review).

To address the more broadly reaching issue of STR sequence nomenclature, we formalized our group in 2018 as the STRAND Working Group (Short Tandem Repeat: Align, Name, Define). Subsequently, we received the endorsement of the ISFG Executive Board to organize an STR sequence nomenclature meeting, which was held in London on April 11th and 12th, 2019. Invited attendees for this meeting included researchers known-to-us to be developing STR sequence-based nomenclature schemata, scientific representatives from vendors developing STR sequence bioinformatic methods, DNA intelligence database curators, and academic experts in STR genomics. Attendees and affiliations were as follows:

Attendee Name	Affiliation
David Ballard	King's College London, UK
Pedro A. Barrio	National Institute of Toxicology and Forensic Science, Spain
Martin Bodner	Medical University of Innsbruck, Austria
Claus Børsting	University of Copenhagen, Denmark
Lisa Borsuk	National Institute of Standards and Technology, US
Laurence Devesse	King's College London, UK
Kristiaan van der Gaag	Netherlands Forensic Institute, Netherlands
Sebastian Ganschow	LABCON-OWL, Germany
Katherine Gettings	National Institute of Standards and Technology, US
Peter Gill	Norwegian Institute of Public Health, Norway
Theresa Gross	University of Cologne, Germany
Douglas Hares	Federal Bureau of Investigation, US
Cydne Holt	Verogen, US
Jerry Hoogenboom	Netherlands Forensic Institute, Netherlands
Tunde Huszar	University of Leicester, UK
Jodi Irwin	Federal Bureau of Investigation, US
Rebecca Just	Federal Bureau of Investigation, US
Jonathan King	University of North Texas Health Science Center, US
Peter de Knijff	Leiden University, Netherlands
Robert Lagacé	Thermo Fisher, US
Walther Parson	Medical University of Innsbruck, Austria
Christopher Phillips	University of Santiago de Compostela, Spain
Peter Schneider	University of Cologne, Germany
Christian Sell	BKA Wiesbaden, Germany
Sascha Willuweit	Charité University of Medicine Berlin, Germany
Brian Young	NicheVision, US

* Corresponding author at: National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD, 20899-8314, USA.
E-mail address: katherine.gettings@nist.gov (K.B. Gettings).

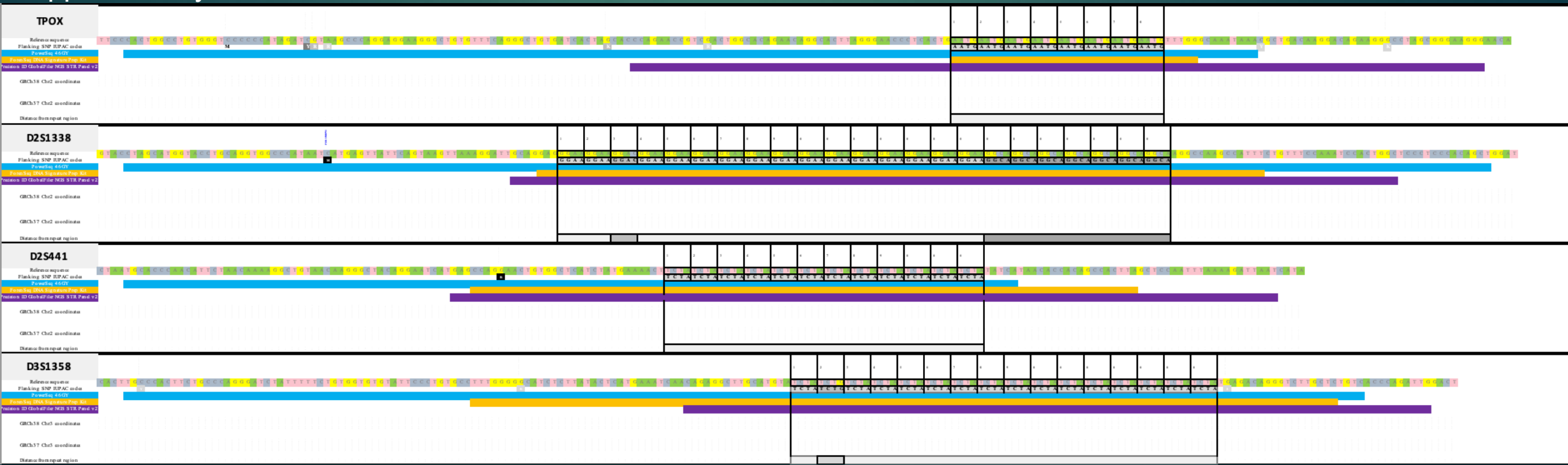
https://doi.org/10.1016/j.fsigen.2019.102165
Received 12 July 2019; Received in revised form 19 September 2019; Accepted 20 September 2019
Available online 21 September 2019

STR Nomenclature Meeting Report

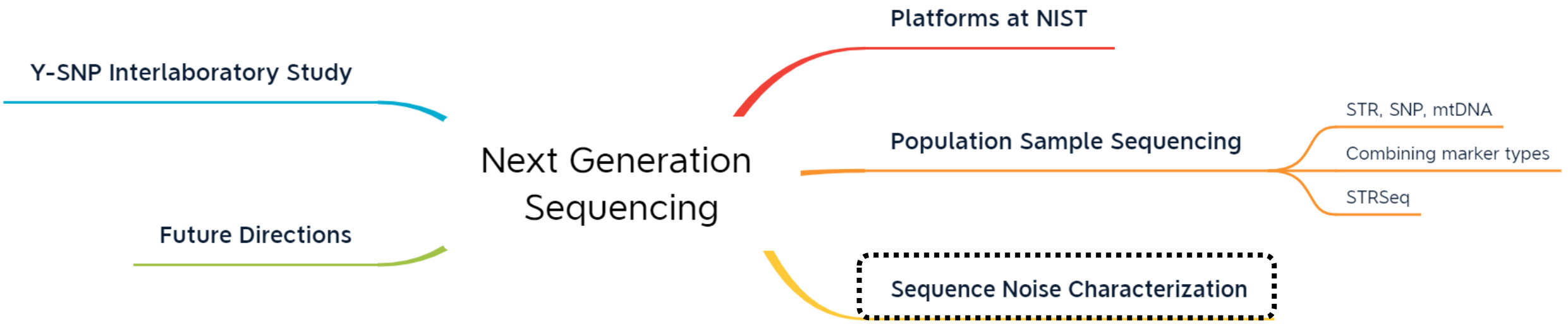
Defined Coordinates

PowerSeq 46GY	GeneMarker NGS Range
ForenSeq DNA Signature Prep Kit	UAS Flanking Region Report Range
Precision ID GlobalFiler NGS v2	Converge .bed file range

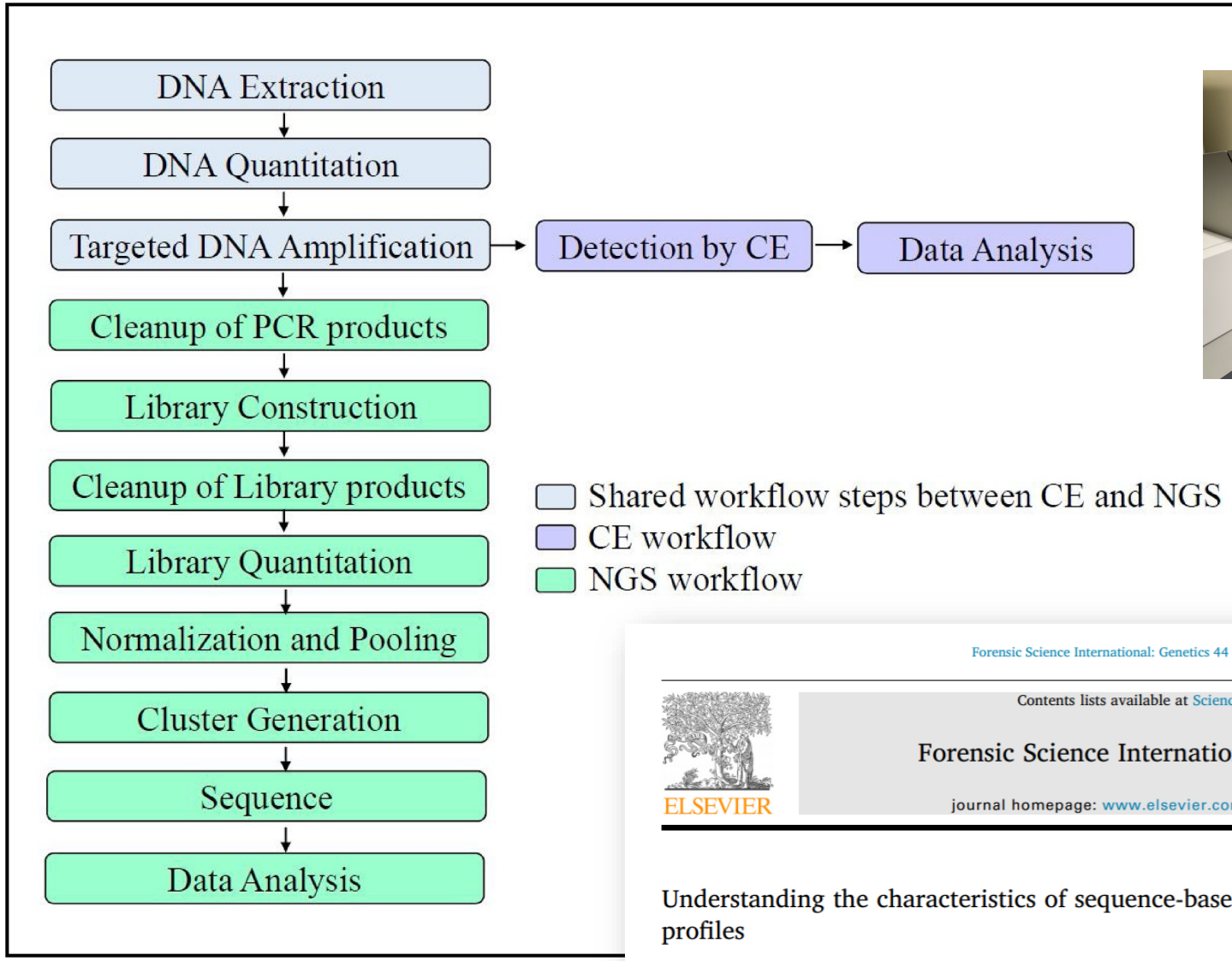
Supplementary File - 24 auSTRs



Topics for today



Comparison of conventional CE versus NGS-STR genotyping workflows



Riman et al. 2020

Forensic Science International: Genetics 44 (2020) 102192

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Understanding the characteristics of sequence-based single-source DNA profiles

Sarah Riman[a,*], Hari Iyer[b], Lisa A. Borsuk[a], Peter M. Vallone[a]

[a] U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA
[b] U.S. National Institute of Standards and Technology, Statistical Engineering Division, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA

Data Analysis by NGS

A = True known allele sequences

S1 = Primary back stutter (LUS of basic repeat motif)

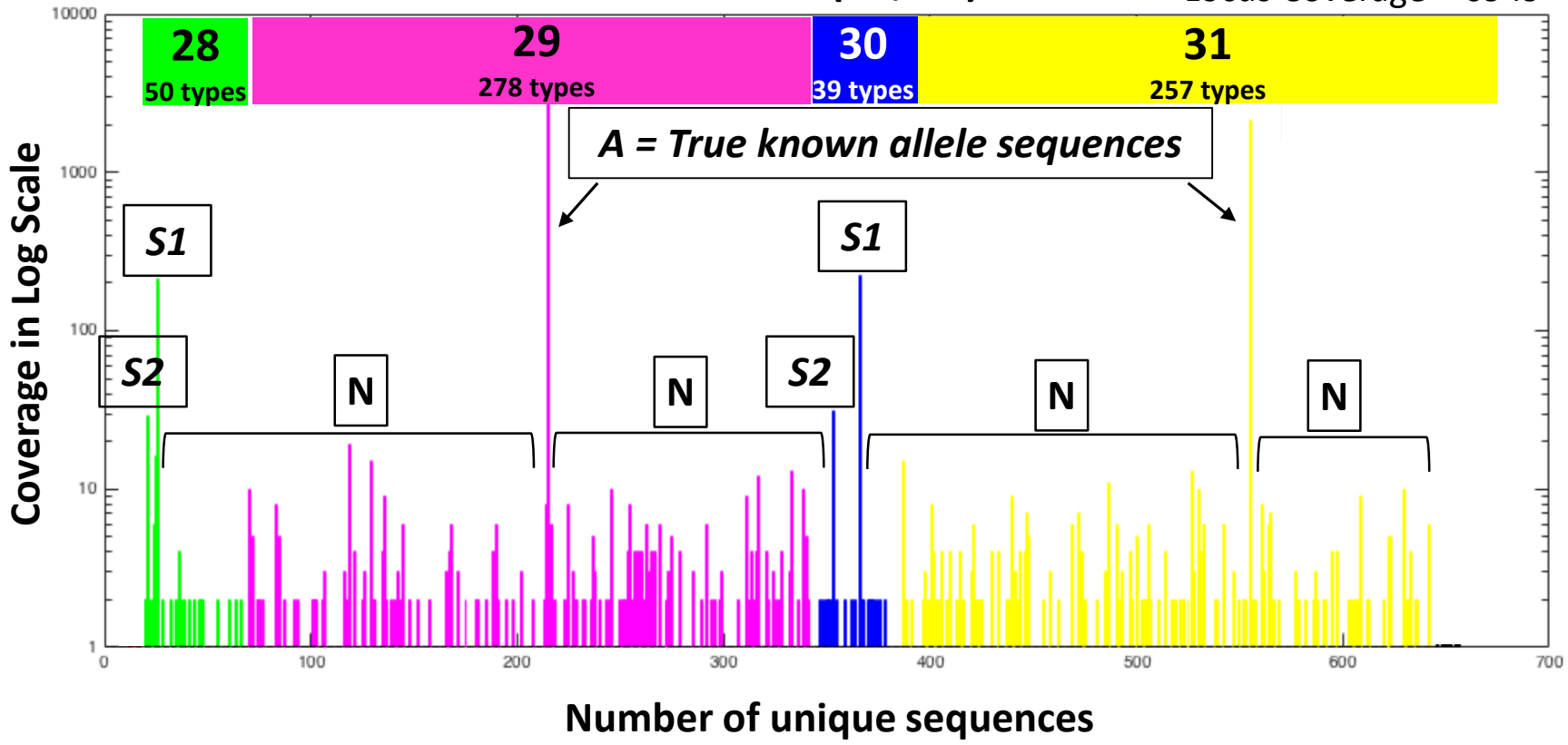
S2 = Back stutter sequences not attributed to S1

N = Noise sequences

Total Type of Sequences = 646

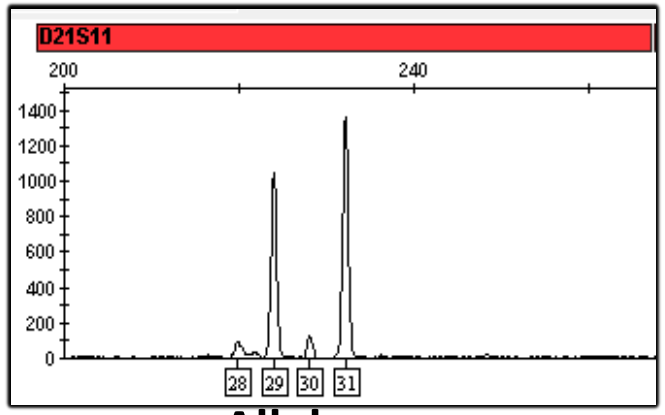
Locus Coverage = 6949

D21S11 (29, 31)

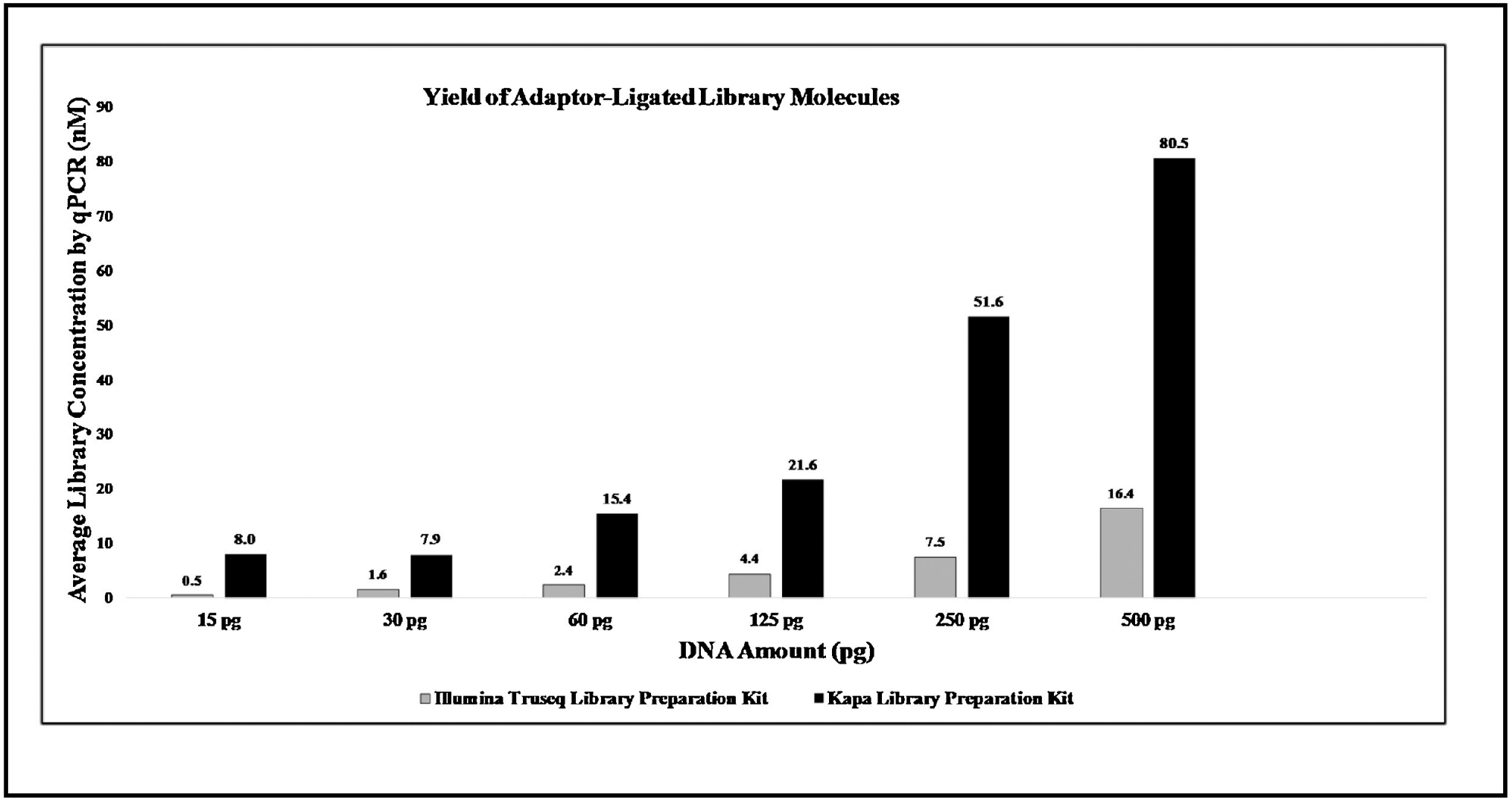


Data Analysis by CE

D21S11 (29, 31)

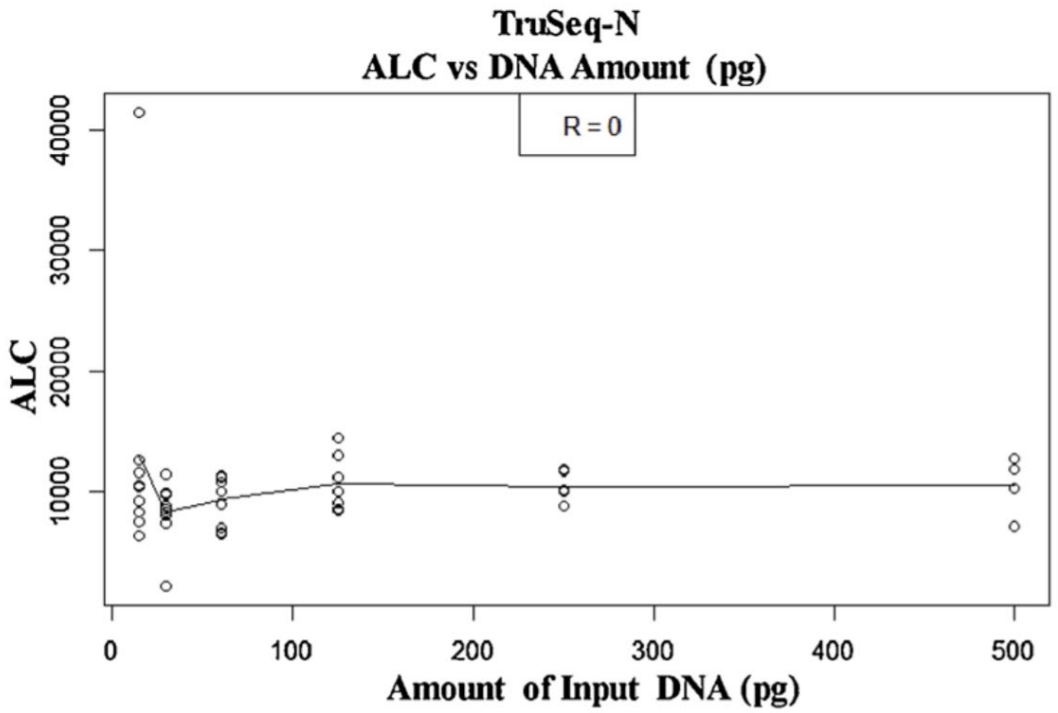


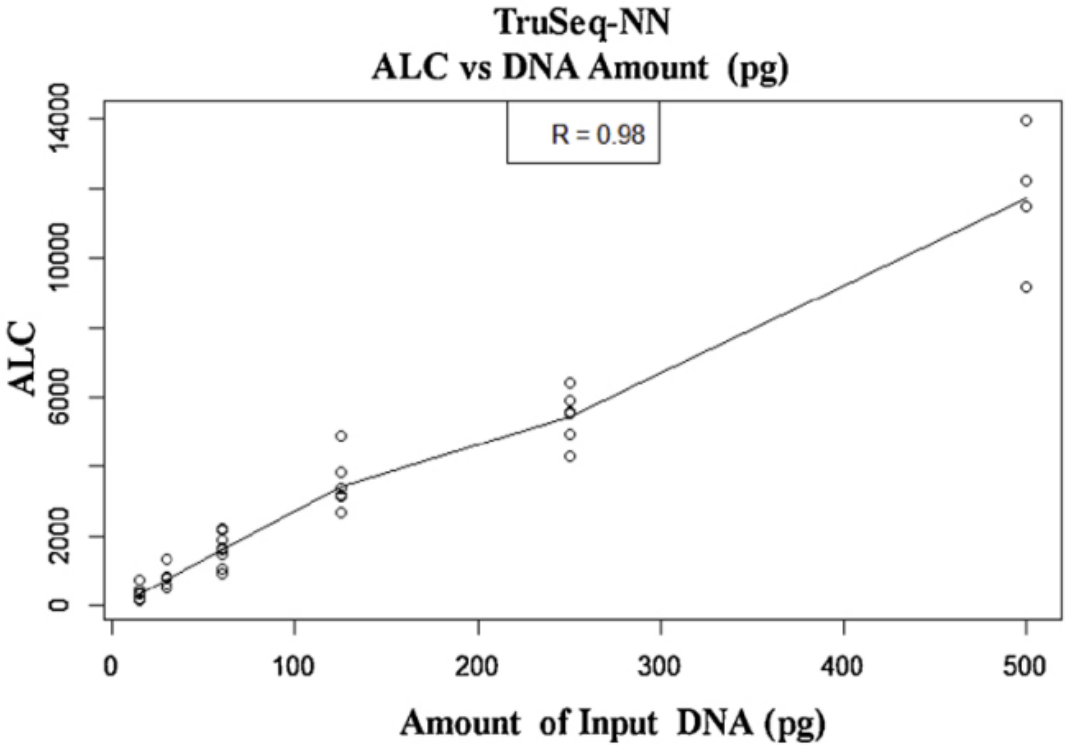
Average library concentration yield versus starting amount of DNA template



Riman et al. 2020

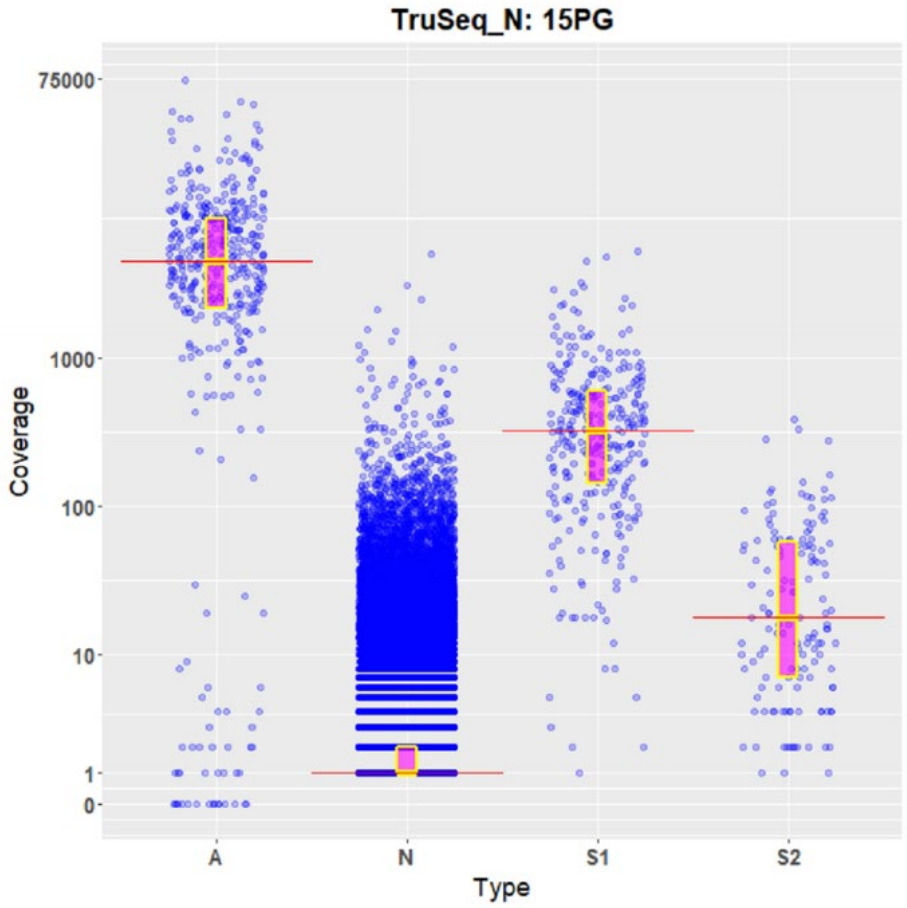
Average Locus Coverage relative to DNA template amounts
Normalized vs. Non-normalized libraries

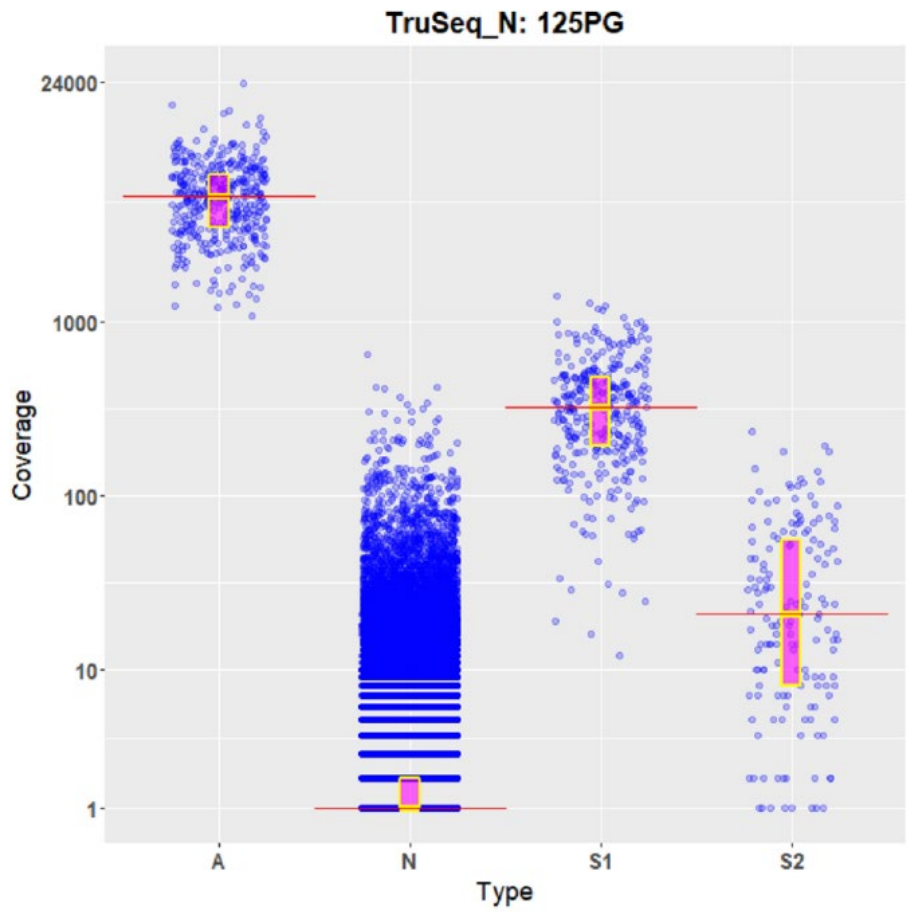




Riman et al. 2020

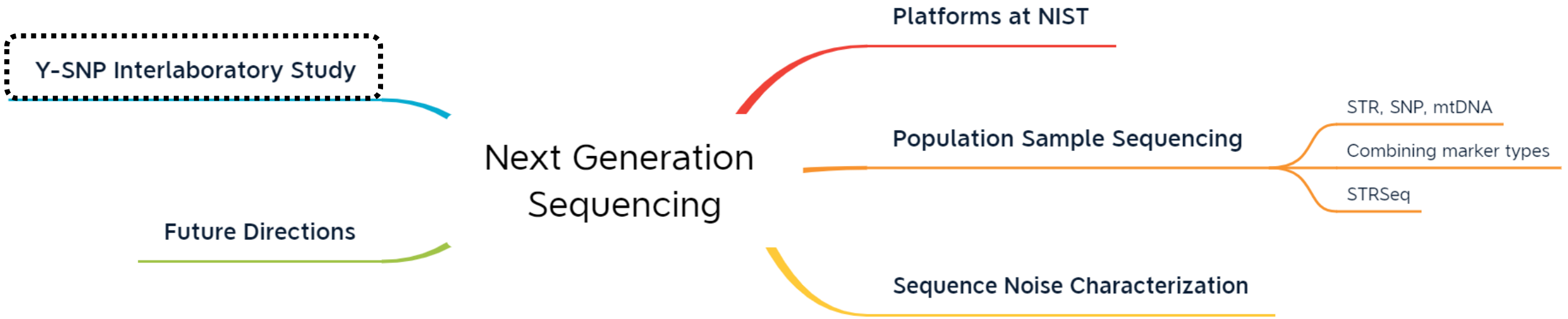
Impact of DNA template amount on the distribution of known allele, stutter, and noise sequences





Riman et al. 2020

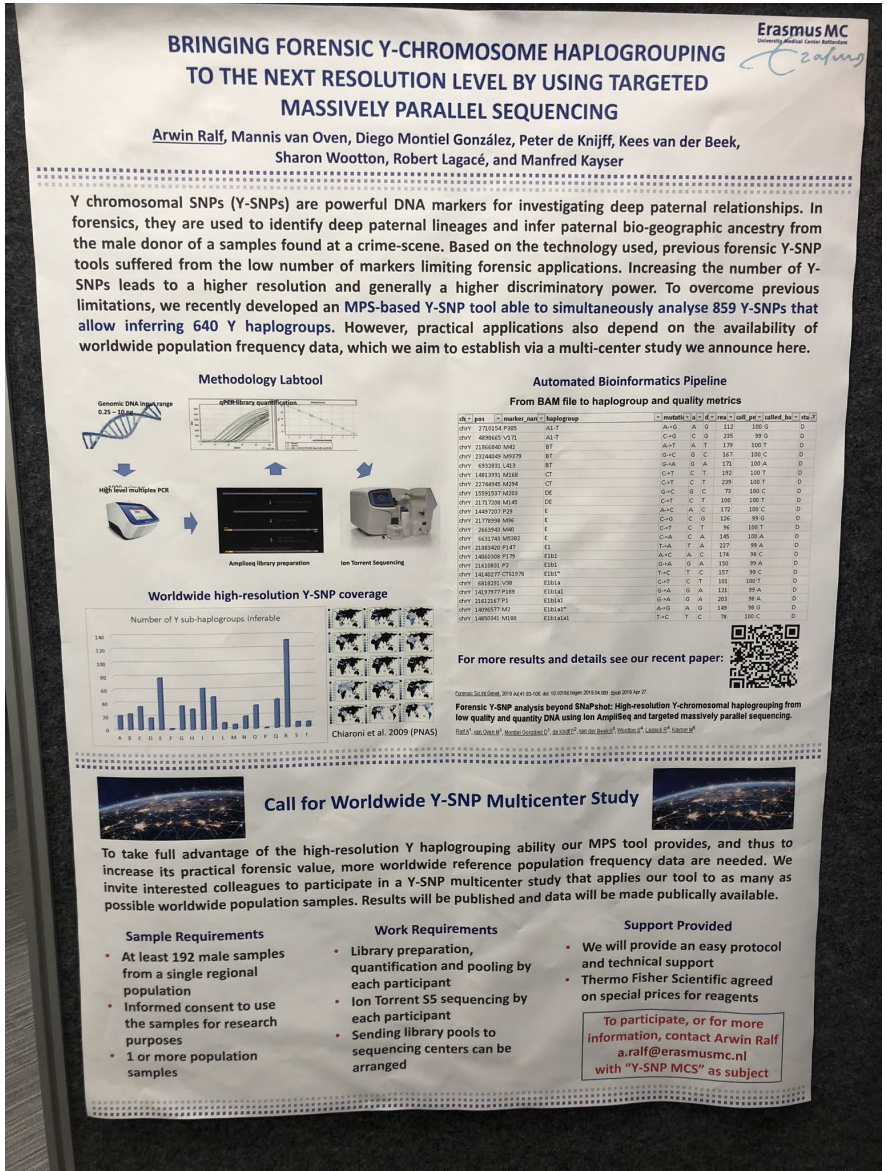
Topics for today



Y-SNP Sequence Study with U.S. Population Samples

Worldwide collaboration with Erasmus Medical Center Rotterdam

Call for Participants at ISFG 2019



Erasmus MC
University Medical Center Rotterdam

BRINGING FORENSIC Y-CHROMOSOME HAPLOGROUPING TO THE NEXT RESOLUTION LEVEL BY USING TARGETED MASSIVELY PARALLEL SEQUENCING

Arwin Ralf, Mannis van Oven, Diego Montiel González, Peter de Knijff, Kees van der Beek, Sharon Wootton, Robert Lagacé, and Manfred Kayser

Y chromosomal SNPs (Y-SNPs) are powerful DNA markers for investigating deep paternal relationships. In forensics, they are used to identify deep paternal lineages and infer paternal bio-geographic ancestry from the male donor of a samples found at a crime-scene. Based on the technology used, previous forensic Y-SNP tools suffered from the low number of markers limiting forensic applications. Increasing the number of Y-SNPs leads to a higher resolution and generally a higher discriminatory power. To overcome previous limitations, we recently developed an MPS-based Y-SNP tool able to simultaneously analyse 859 Y-SNPs that allow inferring 640 Y haplogroups. However, practical applications also depend on the availability of worldwide population frequency data, which we aim to establish via a multi-center study we announce here.

Methodology Labtool

Genomic DNA input range 0.25 – 10 ng

qPCR library quantification

High level multiplex PCR

Ampliseq library preparation

Ion Torrent Sequencing

Automated Bioinformatics Pipeline

From BAM file to haplogroup and quality metrics

ch	pos	marker_nam	haplogroup	mutati	a	d	rea	call_pe	called_ba	rta
chrY	1701154 P305		A1-T	A->G	A	G	112	100 G		D
chrY	4898465 V171		A1-T	C->G	C	G	235	99 G		D
chrY	21866840 M42		BT	A->T	A	T	179	100 T		D
chrY	23244840 M9379		BT	G->C	G	C	167	100 C		D
chrY	6932831 L413		BT	G->A	G	A	171	100 A		D
chrY	14813391 M168		CT	C->T	C	T	192	100 T		D
chrY	21744945 M294		CT	C->T	C	T	239	100 T		D
chrY	15591537 M203		DE	G->C	G	C	73	100 C		D
chrY	2171708 M145		DE	C->T	C	T	108	100 T		D
chrY	14497207 P29		E	A->C	A	C	172	100 C		D
chrY	21778998 M96		E	C->G	C	G	126	99 G		D
chrY	2663943 M40		E	C->T	C	T	96	100 T		D
chrY	6631743 M5362		E	C->A	C	A	145	100 A		D
chrY	21983420 P147		E1	T->A	T	A	227	99 A		D
chrY	14661668 P179		E1b1	A->C	A	C	178	99 C		D
chrY	21618851 P2		E1b1	G->A	G	A	150	99 A		D
chrY	14140277 CTS1976		E1b1*	T->C	T	C	157	99 C		D
chrY	6818291 V38		E1b1a	C->T	C	T	101	100 T		D
chrY	14197977 P189		E1b1a1	G->A	G	A	121	99 A		D
chrY	21612167 P1		E1b1a1	G->A	G	A	203	98 A		D
chrY	14496577 M2		E1b1a1*	A->G	A	G	149	98 G		D
chrY	14850341 M180		E1b1a1a1	T->C	T	C	78	100 C		D

Worldwide high-resolution Y-SNP coverage

Chiaroni et al. 2009 (PNAS)

For more results and details see our recent paper:

Forensic Sci Int Genet. 2019 Jul;41:93-106. doi: 10.1016/j.fsigen.2019.04.001. Epub 2019 Apr 27.

Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing.

Ralf A, van Oven M, Montiel González D, de Knijff P, van der Beek K, Wootton S, Lagacé R, Kayser M.

Call for Worldwide Y-SNP Multicenter Study

To take full advantage of the high-resolution Y haplogrouping ability our MPS tool provides, and thus to increase its practical forensic value, more worldwide reference population frequency data are needed. We invite interested colleagues to participate in a Y-SNP multicenter study that applies our tool to as many as possible worldwide population samples. Results will be published and data will be made publically available.

Sample Requirements

- At least 192 male samples from a single regional population
- Informed consent to use the samples for research purposes
- 1 or more population samples

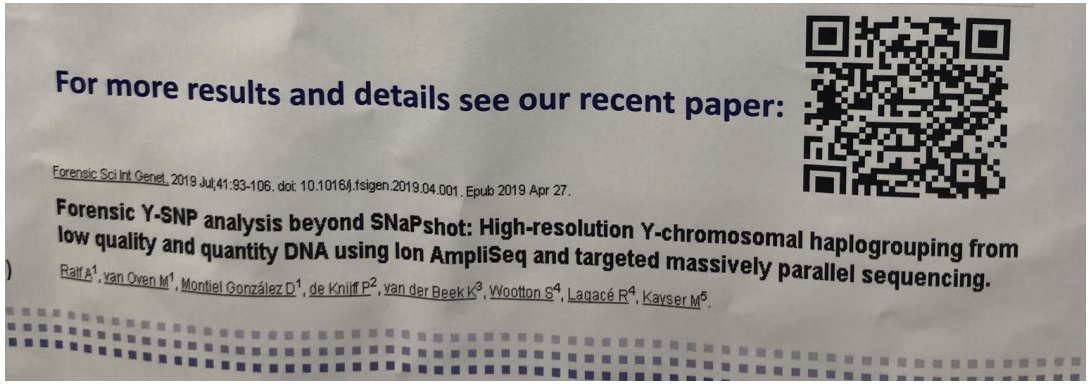
Work Requirements

- Library preparation, quantification and pooling by each participant
- Ion Torrent S5 sequencing by each participant
- Sending library pools to sequencing centers can be arranged

Support Provided

- We will provide an easy protocol and technical support
- Thermo Fisher Scientific agreed on special prices for reagents

To participate, or for more information, contact Arwin Ralf
a.ralf@erasmusmc.nl
with "Y-SNP MCS" as subject

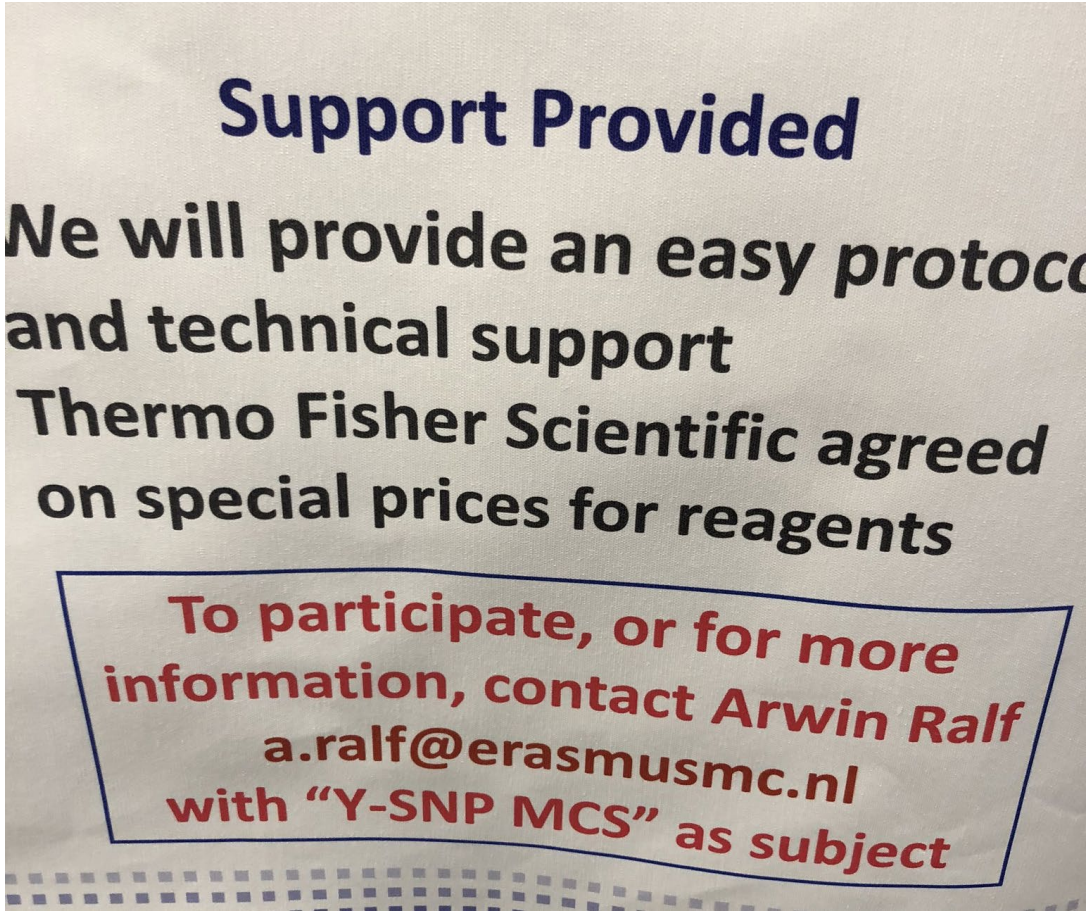


For more results and details see our recent paper:

Forensic Sci Int Genet. 2019 Jul;41:93-106. doi: 10.1016/j.fsigen.2019.04.001. Epub 2019 Apr 27.

Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing.

Ralf A[1], van Oven M[1], Montiel González D[1], de Knijff P[2], van der Beek K[3], Wootton S[4], Lagacé R[4], Kayser M[5].



Support Provided

We will provide an easy protocc
and technical support
Thermo Fisher Scientific agreed
on special prices for reagents

To participate, or for more information, contact Arwin Ralf
a.ralf@erasmusmc.nl
with "Y-SNP MCS" as subject

Project Title:

Bringing Forensic Y-Chromosome Haplogrouping to the Next Resolution Level by using Targeted Massively Parallel Sequencing

Study Organizers:

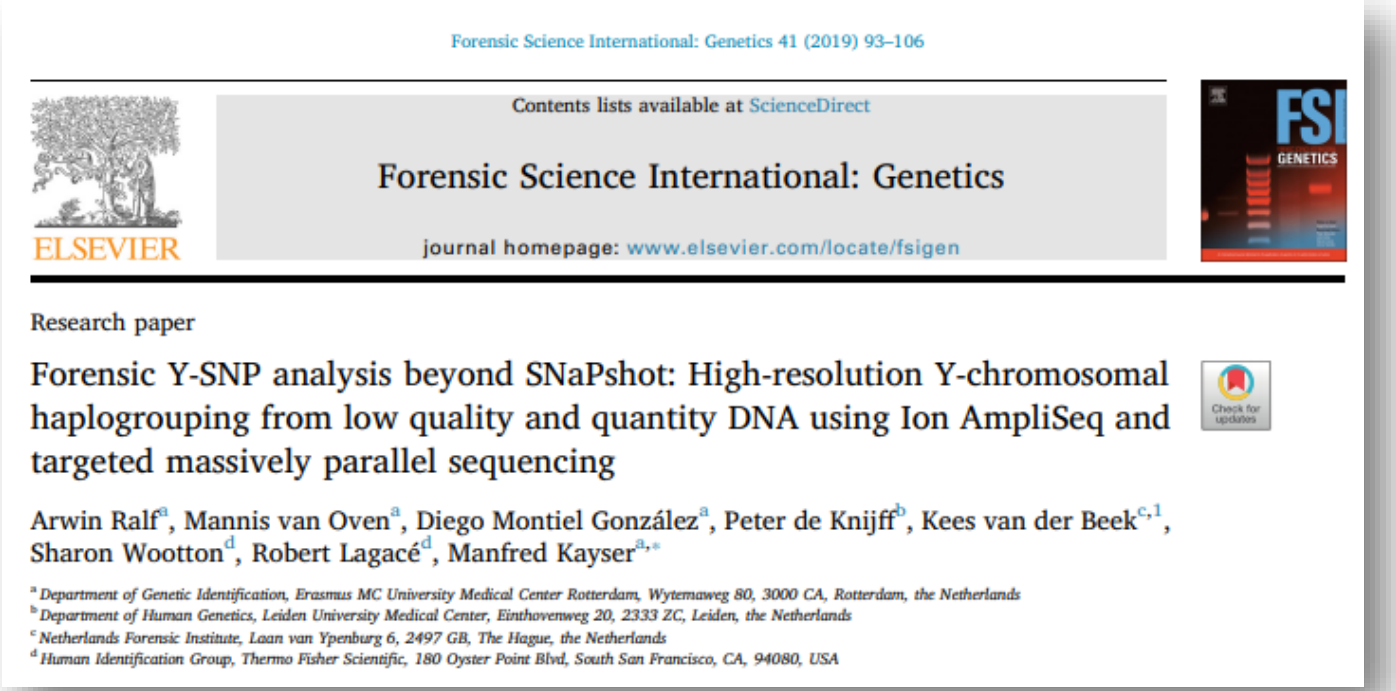
Manfred Kayser and Arwin Ralf

Erasmus MC

Timeframe: Dec 2020
(extended due to COVID-19)

Provided by Organizers

- Protocol
- Y-SNP Panel



Forensic Science International: Genetics 41 (2019) 93–106

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing

Arwin Ralf[a], Mannis van Oven[a], Diego Montiel González[a], Peter de Knijff[b], Kees van der Beek[c,1], Sharon Wootton[d], Robert Lagacé[d], Manfred Kayser[a,*]

[a] Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Wytemaweg 80, 3000 CA, Rotterdam, the Netherlands
[b] Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC, Leiden, the Netherlands
[c] Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB, The Hague, the Netherlands
[d] Human Identification Group, Thermo Fisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA, 94080, USA

Reference Paper

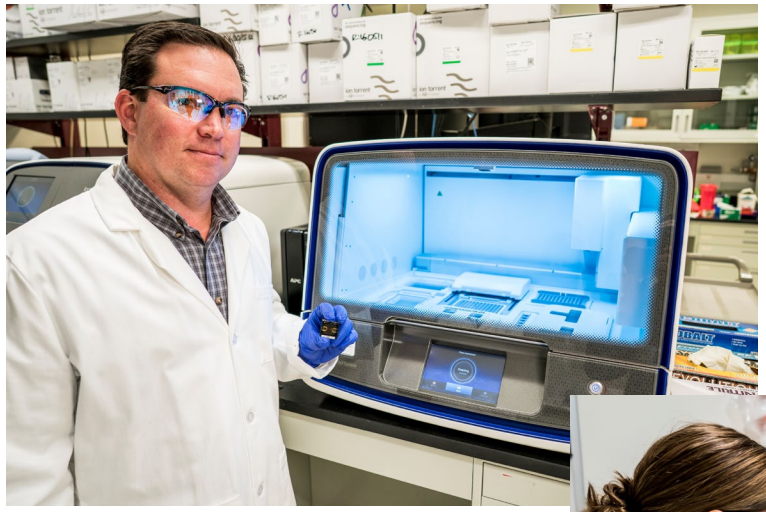
Description of Study – Sequencing Y-SNP markers

To obtain worldwide population frequency data:

- 884 Y-SNP Markers sequenced on Ion S5
- Infer 640 Y haplogroups
- ≥192 male samples per population

NIST 1032 male samples:

- 359 U.S. Caucasians
- 341 African Americans
- 236 U.S. Hispanics
- 96 U.S. Asians (extra data as needed)



Previous Study with Erasmus: Rapidly Mutating Y-STR Markers with U.S. Population Male Samples

Forensic Science International: Genetics 12 (2014) 12–23

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

A global analysis of Y-chromosomal haplotype diversity for 23 STR loci

Josephine Purps [a,1], Sabine Siegert [b,1], Sascha Willuweit [a], Marion Nagy [a], Cíntia Alves [c], Renato Salazar [c,d], Sheila M.T. Angustia [e], Lorna H. Santos [e], Katja Anslinger [f], Birgit Bayer [f], Qasim Ayub [g], Wei Wei [g], Yali Xue [g], Chris Tyler-Smith [g], Miriam Baeta Bafalluy [h], Begoña Martínez-Jarreta [h], Balazs Egyed [i], Beate Balitzki [j], Sibylle Tschumi [j], David Ballard [k], Denise Syndercombe Court [k], Xinia Barrantes [l], Gerhard Bäßler [m], Tina Wiest [m], Burkhard Berger [n], Harald Niederstätter [n,o], Walther Parson [n,o], Carey Davis [p], Bruce Budowle [p,q], Helen Burri [r], Urs Borer [r], Christoph Koller [r], Elizeu F. Carvalho [s], Patricia M. Domingues [s], Wafaa Takash Chamoun [t], Michael D. Coble [u], Carolyn R. Hill [u], Daniel Corach [v], Mariela Caputo [v], Maria E. D'Amato [w], Sean Davison [w], Ronny Decorte [x], Maarten H.D. Larmuseau [x], Claudio Ottoni [x], Olga Rickards [y], Di Lu [z], Chengtao Jiang [z], Tadeusz Dobosz [aa], Anna Jonkisz [aa], William E. Frank [ab], Ivana Furac [ac], Christian Gehrig [ad], Vincent Castella [ad], Branka Grskovic [ae], Cordula Haas [af], Jana Wobst [af], Gavrilo Hadzic [ag], Katja Drobnic [ag], Katsuya Honda [ah], Yiping Hou [ai], Di Zhou [ai], Yan Li [ai], Shengping Hu [aj], Shenglan Chen [aj], Uta-Dorothee Immel [ak], Rüdiger Lessig [ak], Zlatko Jakovski [al], Tanja Ilievska [al], Anja E. Klann [am], Cristina Cano García [am], Peter de Knijff [an], Thirsa Kraaijenbrink [an], Aikaterini Kondili [ao], Penelope Miniati [ao], Maria Vouropoulou [ao], Lejla Kovacevic [ap], Damir Marjanovic [ap], Iris Lindner [aq], Issam Mansour [ar], Mouayyad Al-Azem [ar], Ansar El Andari [ar], Miguel Marino [as], Sandra Furfuro [as], Laura Locarno [as], Pablo Martín [at], Gracia M. Luque [at], Antonio Alonso [at], Luís Souto Miranda [au], Helena Moreira [au], Natsuko Mizuno [av], Yasuki Iwashima [av], Rodrigo S. Moura Neto [aw], Tatiana L.S. Nogueira [ax], Rosane Silva [ay], Marina Nastainczyk-Wulf [az], Jeanett Edelmann [az], Michael Kohl [az], Shengjie Nie [ba], Xianping Wang [bb], Baowen Cheng [bc], Carolina Núñez [bd], Marian Martínez de Pancorbo [bd], Jill K. Olofsson [be], Niels Morling [be], Valerio Onofri [bf], Adriano Tagliabracci [bf], Horolma Pamjav [bg], Antonia Volgyi [bg], Gusztav Barany [bg], Ryszard Pawlowski [bh], Agnieszka Maciejewska [bh], Susi Pelotti [bi], Witold Pepinski [bj], Monica Abreu-Glowacka [bk], Christopher Phillips [bl], Jorge Cárdenas [bl], Danel Rey-Gonzalez [bl], Antonio Salas [bl], Francesca Brisighelli [bl,bm], Cristian Capelli [bl,bn], Ulises Toscanini [bl,bo], Andrea Piccinini [bp], Marilidia Piglionica [bq], Stefania L. Baldassarra [bq], Rafal Ploski [br], Magdalena Konarzewska [br], Emila Jastrzebska [bs], Carlo Robino [bt], Antti Sajantila [p,bu], Jukka U. Palo [bu], Evelyn Guevara [bu], Jazelyn Salvador [bv], Maria Corazon De Ungria [bv], Jae Joseph Russell Rodriguez [bv,bw], Ulrike Schmidt [bx], Nicola Schlauderer [bx], Pekka Saukko [by], Peter M. Schneider [bz], Miriam Sirker [bz], Kyoung-Jin Shin [ca], Yu Na Oh [ca], Iulia Skitsa [cb], Alexandra Ampati [cb], Tobi-Gail Smith [cc], Lina Solis de Calvit [cd], Vlastimil Stenzl [ce], Thomas Capal [ce], Andreas Tillmar [cf],

* Corresponding author at: Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité – Universitätsmedizin Berlin (Campus Virchow Klinikum), Augustenburger Platz 1, 13353 Berlin, Germany. Tel.: +49 30 450 525060; fax: +49 30 450 525912.
E-mail address: lutz.roewer@charite.de (L. Roewer).
[1] These authors contributed equally to this work.
[2] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.fsigen.2014.04.008
1872-4973/

Helena Nilsson [cf], Stefania Turrina [cg], Domenico De Leo [cg], Andrea Verzeletti [ch], Venusia Cortellini [ch], Jon H. Wetton [ci], Gareth M. Gwynne [ci], Mark A. Jobling [ci], Martin R. Whittle [cj], Denilce R. Sumita [cj], Paulina Wolańska-Nowak [ck], Rita Y.Y. Yong [cl], Michael Krawczak [cm], Michael Nothnagel [b,2], Lutz Roewer [a,2,*]

[a] Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin, Berlin, Germany
[b] Department of Statistical Genetics and Bioinformatics, Cologne Center for Genomics, University of Cologne, Germany
[c] Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal
[d] Department of Biology, Faculty of Sciences, University of Porto, Portugal
[e] Philippine National Police Crime Laboratory, Quezon City, Philippines
[f] Institut für Rechtsmedizin, Ludwig-Maximilians-Universität, München, Germany
[g] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
[h] Department of Forensic Medicine, University of Zaragoza, Spain
[i] GenoID Forensic DNA Laboratory, Department of Genetics, Eötvös Loránd University, Budapest, Hungary
[j] Institut für Rechtsmedizin, Universität Basel, Switzerland
[k] Department of Forensic and Analytical Science, King's College London, London, UK
[l] Forensic Sciences Department, Poder Judicial, Heredia, Costa Rica
[m] Landeskriminalamt Baden-Württemberg, Germany
[n] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[o] Penn State Eberly College of Science, University Park, PA, USA
[p] Institute of Applied Genetics and Department of Molecular and Medical Genetics, Ft. Worth, USA
[q] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia
[r] Forensische Genetik, Kantonsspital Aarau AG, Switzerland
[s] Laboratorio de Diagnósticos por DNA, Instituto de Biologia, Universidade do Estado do Rio de Janeiro, Brazil
[t] Human DNA Diagnostics Laboratory, Beirut, Lebanon
[u] National Institute of Standards and Technology, Gaithersburg, USA
[v] Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Servicio de Huellas Digitales Genetica and CONICET (National Scientific and Technical Research Council), Buenos Aires, Argentina
[w] University of the Western Cape, Biotechnology Department, Forensic DNA Laboratory, Cape Town, South Africa
[x] KU Leuven, Department of Imaging & Pathology, Laboratory of Forensic Genetics and Molecular Archaeology, Leuven, Belgium
[y] Centre of Molecular Antropology For Ancient DNA Studies, Department of Biology, University of Rome Tor Vergata, Italy
[z] Collaborative Innovation Center of Judicial Civilization, Institute of Evidence Law and Forensic Science, China University of Political Science and Law, Beijing, China
[aa] Institute of Forensic Medicine, Medical University, Wroclaw, Poland
[ab] Illinois State Police, Research & Development Laboratory, Springfield, USA
[ac] Department of Forensic Medicine and Criminology, University of Zagreb, Croatia
[ad] University Center of Legal Medicine, Lausanne-Geneva, Lausanne, Switzerland
[ae] Forensic Science Centre "Ivan Vucetic", General Police Directorate, Ministry of Interior, Zagreb, Croatia
[af] Institut für Rechtsmedizin, Universität Zürich, Switzerland
[ag] National Forensic Laboratory, Ljubljana, Slovenia
[ah] Department of Legal Medicine, Faculty of Medicine, University of Tsukuba, Japan
[ai] Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine Sichuan University, Chengdu, China
[aj] Molecular Biology and Forensic Genetics Laboratory, Shantou University Medical College, Shantou, China
[ak] Institut für Rechtsmedizin, Universität Halle, Germany
[al] Institute for Forensic Medicine and Criminalistics, Medical Faculty, University "Ss. Cyril and Methodius", Skopje, Macedonia
[am] Institut für Rechtsmedizin, Universitätsmedizin Greifswald, Germany
[an] Forensic Laboratory for DNA Research, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
[ao] Subdivision of Biological and Biochemical Examinations and Analyses F.S.D. – Hellenic Police, Athens, Greece
[ap] Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina
[aq] Institut für Rechtsmedizin, Universität Rostock, Germany
[ar] Molecular Biology Laboratory, American University of Science and Technology Beirut, Lebanon and School of Criminal Justice, University of Lausanne, Switzerland
[as] Laboratorio de Análisis de ADN, FCM - National University of Cuyo, Mendoza, Argentina
[at] Instituto Nacional de Toxicología y Ciencias Forenses, Madrid, Spain
[au] Departamento de Biologia, Universidade de Aveiro, Portugal
[av] National Research Institute of Police Science, Chiba, Japan
[aw] Instituto de Biologia, Universidade Federal do Rio de Janeiro and DIMAV/INMETRO, Brazil
[ax] Instituto de Biologia do Exército, Rio de Janeiro, Brazil
[ay] Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Brazil
[az] Institut für Rechtsmedizin, Universität Leipzig, Germany
[ba] School of Forensic Medicine, Kunming Medical University, Kunming, China
[bb] Department of Criminal Investigation, Xuanwei Public Security Bureau, Xuanwei, China
[bc] Department of Criminal Investigation, Yunnan Provincial Public Security Bureau, Kunming, China
[bd] BIOMICs Research Group, Universidad del País Vasco, Vitoria, Spain
[be] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[bf] Section of Legal Medicine, Università Politecnica delle Marche, Ancona, Italy
[bg] DNA Laboratory, Institute for Forensic Medicine, Network of Forensic Science Institutes, Ministry of Public Administration and Justice, Budapest, Hungary
[bh] Forensic Genetics Laboratory, Institute of Forensic Medicine, Medical University of Gdansk, Poland
[bi] Department of Medical and Surgical Sciences (DIMEC), Institute of Legal Medicine, School of Medicine, University of Bologna, Italy
[bj] Department of Forensic Medicine, Medical University of Bialystok, Poland
[bk] Department of Forensic Medicine, Medical University Poznan, Poland
[bl] Unidade de Xenética Forense, Instituto de Ciencias Forenses, Grupo de Medicina Xenómica, Facultade de Medicina, Universidade de Santiago de Compostela, Spain
[bm] Forensic Genetics Laboratory, Institute of Legal Medicine, Università Cattolica del Sacro Cuore, Rome, Italy
[bn] Department of Zoology, University of Oxford, Oxford, UK
[bo] PRICAI-Fundación Favaloro, Buenos Aires, Argentina
[bp] Forensic Genetics Laboratory, Department of Human Morphology and Biomedical Sciences, Università degli Studi di Milano, Italy
[bq] Interdisciplinary Department of Medicine, Section of Legal Medicine, University of Bari, Italy

[br] Department of Medical Genetics, Warsaw Medical University, Poland
[bs] Department of Forensic Medicine, Warsaw Medical University, Poland
[bt] Department of Public Health Sciences and Pediatrics, University of Turin, Italy
[bu] Department of Forensic Medicine, University of Helsinki, Finland
[bv] DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines Diliman, Philippines
[bw] Institute of Biological Sciences, University of the Philippines Los Baños, Laguna, Philippines
[bx] Institut für Rechtsmedizin, Universitätsklinikum Freiburg, Germany
[by] Department of Forensic Medicine, University of Turku, Finland
[bz] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Germany
[ca] Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, South Korea
[cb] Athens Dept. of Legal Medicine, DNA Analysis Laboratory, Athens, Greece
[cc] Department of Basic Medical Sciences, University of the West Indies, Kingston, Jamaica
[cd] Laboratorio Genetix S.A., Panamá, Panama
[ce] Laboratory of Forensic Genetics, Institute of Criminalistics, Prague, Czech Republic
[cf] Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
[cg] Sezione di Medicina Legale, Dipartimento di Medicina e Sanità Pubblica, Università degli Studi di Verona, Italy
[ch] Istituto di Medicina Legale, Università degli Studi di Brescia, Italy
[ci] Department of Genetics, University of Leicester, UK
[cj] Genomic Engenharia Molecular Ltda., Sao Paulo, Brazil
[ck] Institute of Forensic Research, Krakow, Poland
[cl] Defence Medical & Environmental Research Institute, DSO National Laboratories, Singapore
[cm] Institute of Medical Informatics and Statistics, Christian-Albrechts University Kiel, Germany

ARTICLE INFO

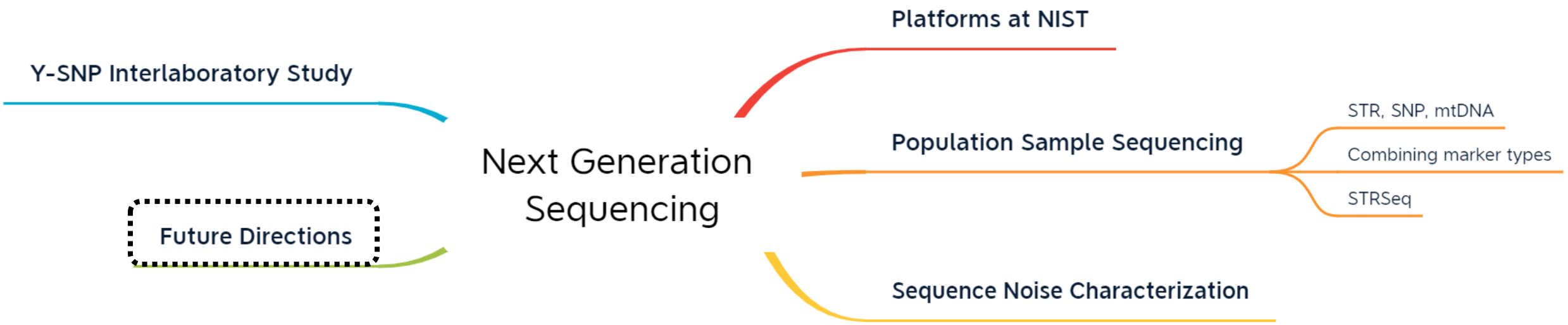
Article history:
Received 28 March 2014
Accepted 19 April 2014

Keywords:
Gene diversity
Discriminatory power
AMOVA
Population structure
Database

ABSTRACT

In a worldwide collaborative effort, 19,630 Y-chromosomes were sampled from 129 different populations in 51 countries. These chromosomes were typed for 23 short-tandem repeat (STR) loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385ab, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATAH4, DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643) and using the PowerPlex Y23 System (PPY23, Promega Corporation, Madison, WI). Locus-specific allelic spectra of these markers were determined and a consistently high level of allelic diversity was observed. A considerable number of null, duplicate and off-ladder alleles were revealed. Standard single-locus and haplotype-based parameters were calculated and compared between subsets of Y-STR markers established for forensic casework. The PPY23 marker set provides substantially stronger discriminatory power than other available kits but at the same time reveals the same general patterns of population structure as other marker sets. A strong correlation was observed between the number of Y-STRs included in a marker set and some of the forensic parameters under study. Interestingly a weak but consistent trend toward smaller genetic distances resulting from larger numbers of markers became apparent.

Topics for today



Future Directions

Implementation… what are the barriers?

Future Directions

Implementation… what are the barriers?

Vendors create/adapt assays and analysis methods

Developmental validation

Future Directions

Implementation… what are the barriers?

Vendors create/adapt assays and analysis methods

Developmental validation

Casework Laboratories evaluate cost/benefit

Internal validation

Future Directions

Implementation… what is needed?

Vendors create/adapt assays and analysis methods

Developmental validation

Casework Laboratories evaluate cost/benefit

Internal validation

Sequence Nomenclature

Match Statistics

Probabilistic Genotyping for NGS

Proficiency Tests

Databases of the Future

Acknowledgements

katherine.gettings@nist.gov


Peter Vallone


Becky Steffen


Erica Romsos


Katherine Gettings


Kevin Kiesler


Margaret Kline


Lisa Borsuk


Sarah Riman


David Duewer


Hari Iyer

Statistical Support


Tunde Huszar PostDoc

Collaborators

David Ballard, Kings College London

Martin Bodner & Walther Parson, Medical University of Innsbruck

Jonathan King, University of North Texas Health Science Center

Chris Phillips, University of Santiago de Compostela

George Riley, National Center for Biotechnology Information

Arwin Rolf & Manfred Kayser, Erasmus University Medical Center

Cassandra Taylor, Kimberly Sturk-Andreaggi, Charla Marshall, AFDIL

Andreas Tillmar, National Board of Forensic Medicine

Funding NIST Special Programs Office and the FBI Biometric Center of Excellence Unit: DNA as a Biometric.