An Introduction to Forensic DNA Typing

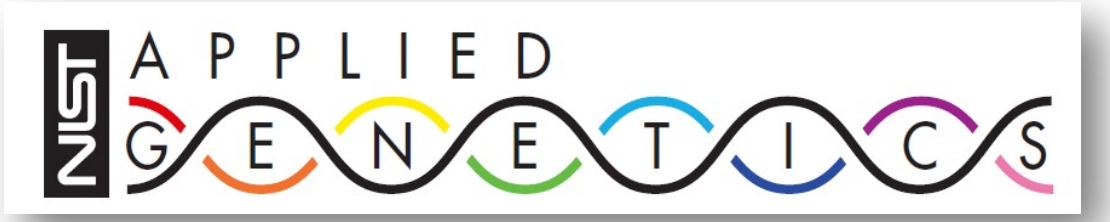
Peter M. Vallone, Ph.D.

Leader, Applied Genetics Group

National Institute of Standards and Technology

APS Mid-Atlantic Senior Physicists Group

June 16, 2021



Outline for today

- Forensic DNA typing workflow
- How work at NIST supports the Forensic DNA Typing community
  - *NIST does not perform forensic casework, but rather provide standards and research to support the forensic community*
- Recent advancements and applications in human identity testing

Assumptions about DNA

General Characteristics of Genomic DNA

- Each individual has a unique DNA profile
  - with exception of monozygotic siblings[1]
- Each person's DNA is the same in every cell
  - DNA from skin cells will match DNA from blood cells
- An individual's DNA profile remains the same throughout life
- Half of your DNA comes from your mother and half from your father
  - implications for determining kinship

DNA transfers and persists and can be collected and analyzed

[1] *Weber-Lehmann et al., Finding the needle in the haystack: Differentiating ''identical'' twins in paternity testing and forensics by ultra-deep next generation sequencing Forensic Science International: Genetics 9 (2014) 42–46*

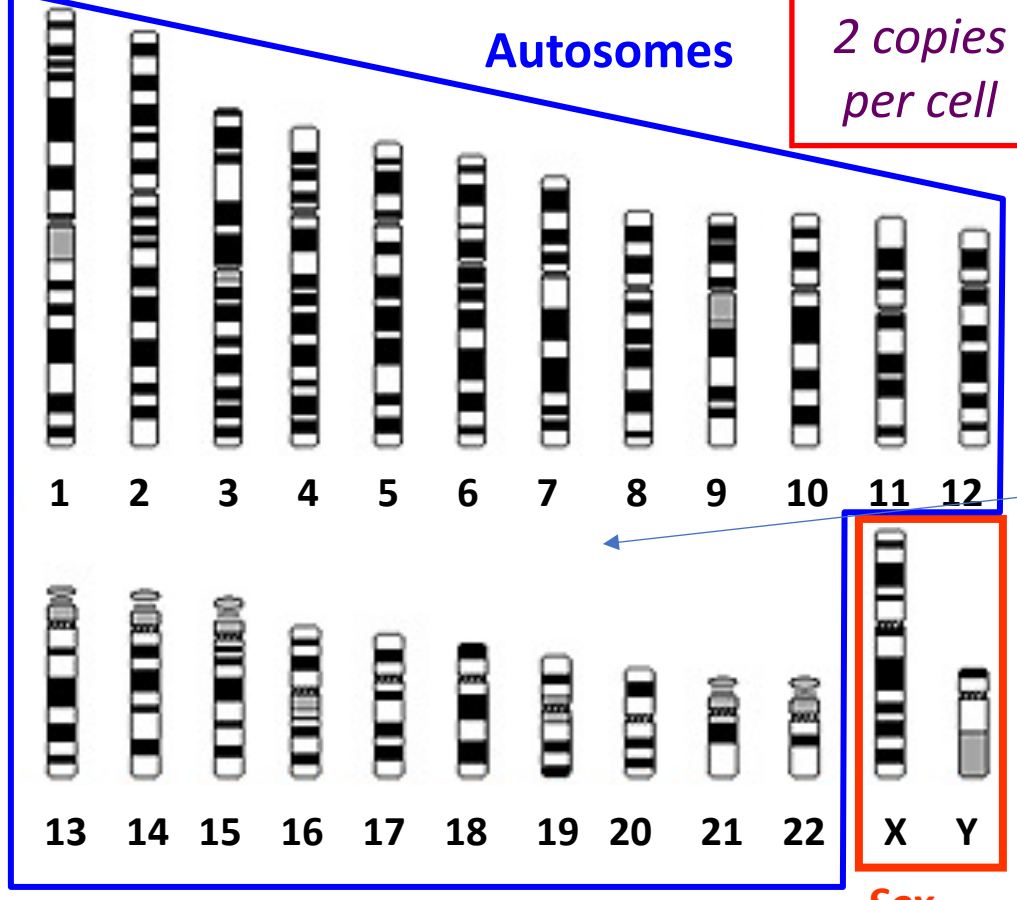
Human Genome

23 Pairs of Chromosomes + mtDNA

Located in cell nucleus

Autosomes

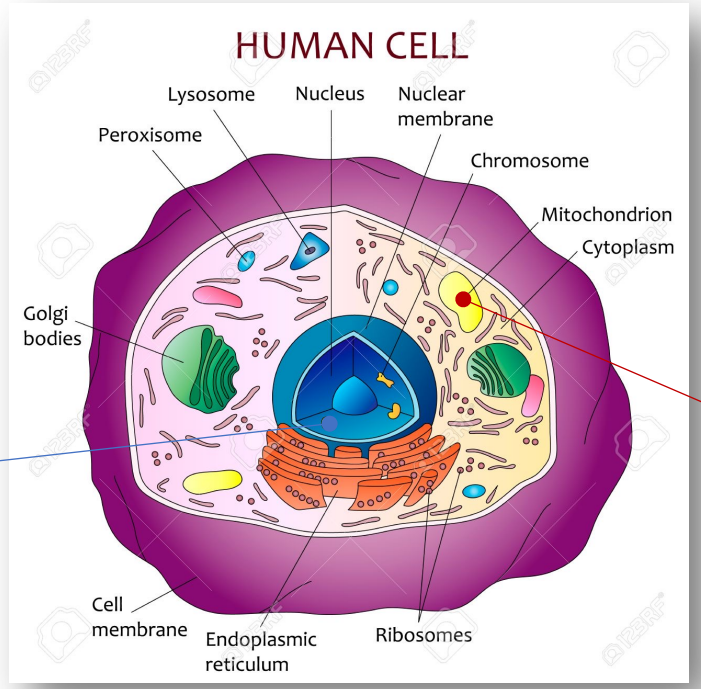
2 copies per cell



Sex-chromosomes

Nuclear DNA

3.2 billion bp



https://www.123rf.com/photo_55148693_stock-vector-human-cell-diagram-.html

Located in mitochondria (multiple copies in cell cytoplasm)

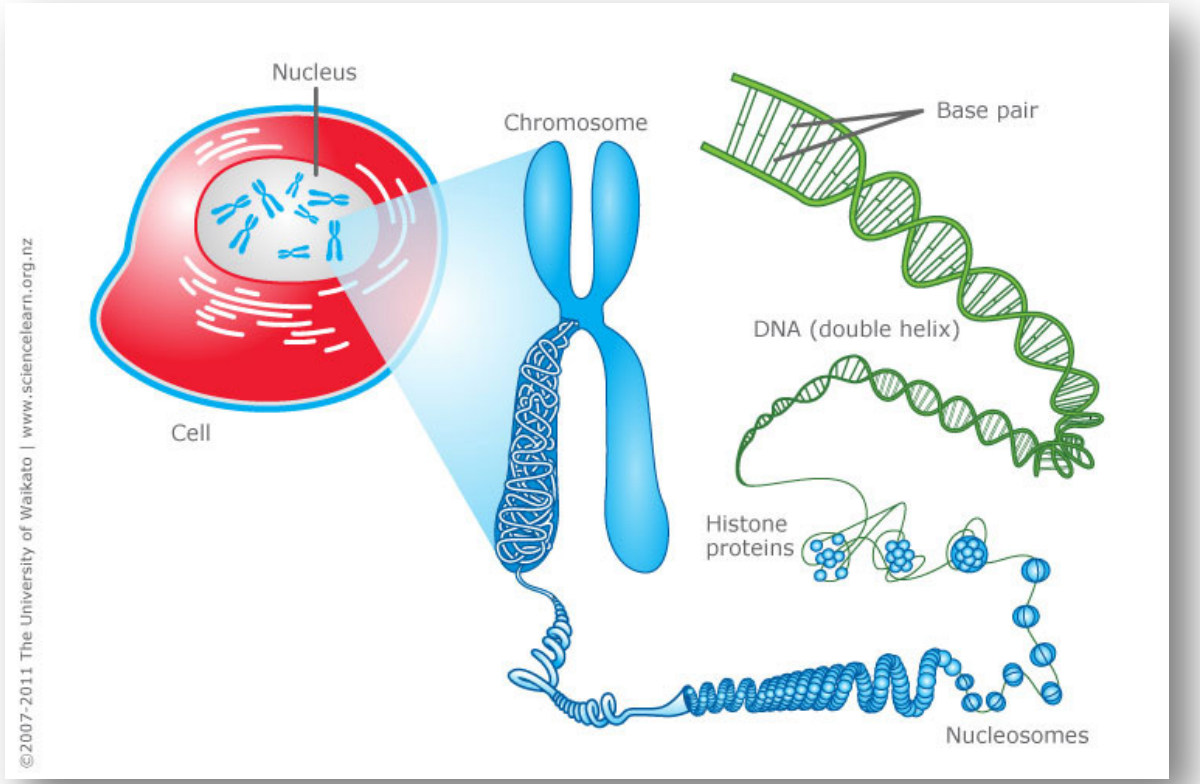
mtDNA

16,569 bp

Mitochondrial DNA

100s of copies per cell

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 2.3, ©Elsevier Science/Academic Press



Unique identifier or ‘barcode’ inherent in your DNA

Genetic Variation

- Length Variation

  short tandem repeats (STRs)

  CTAGTCGT**[GATA][GATA][GATA]**GCGATCGT

- Sequence Variation

  single nucleotide polymorphisms (SNPs)

  insertions/deletions

  GCTAGTCGATGCTC**[G/A]**GCGTATGCTGTAGC

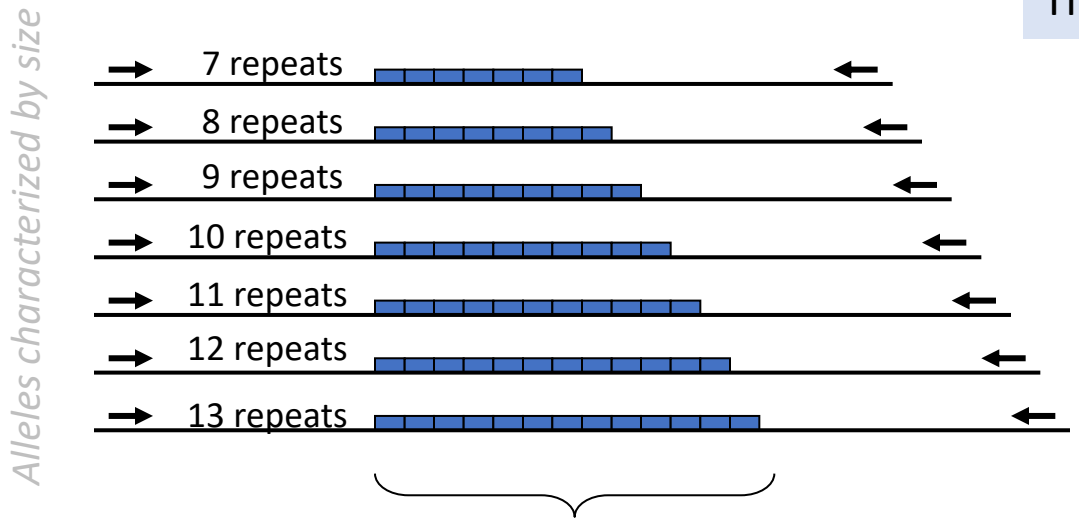
Short Tandem Repeat (STR) Markers

A sequence length variation that occurs between genes

= 12 GATA repeats (“12” allele is reported)

CCTAGATCAATACAGACAGAAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAG
ACAAAACAGAGATGGATGATAGATACATGCTTACAGATGCACAC

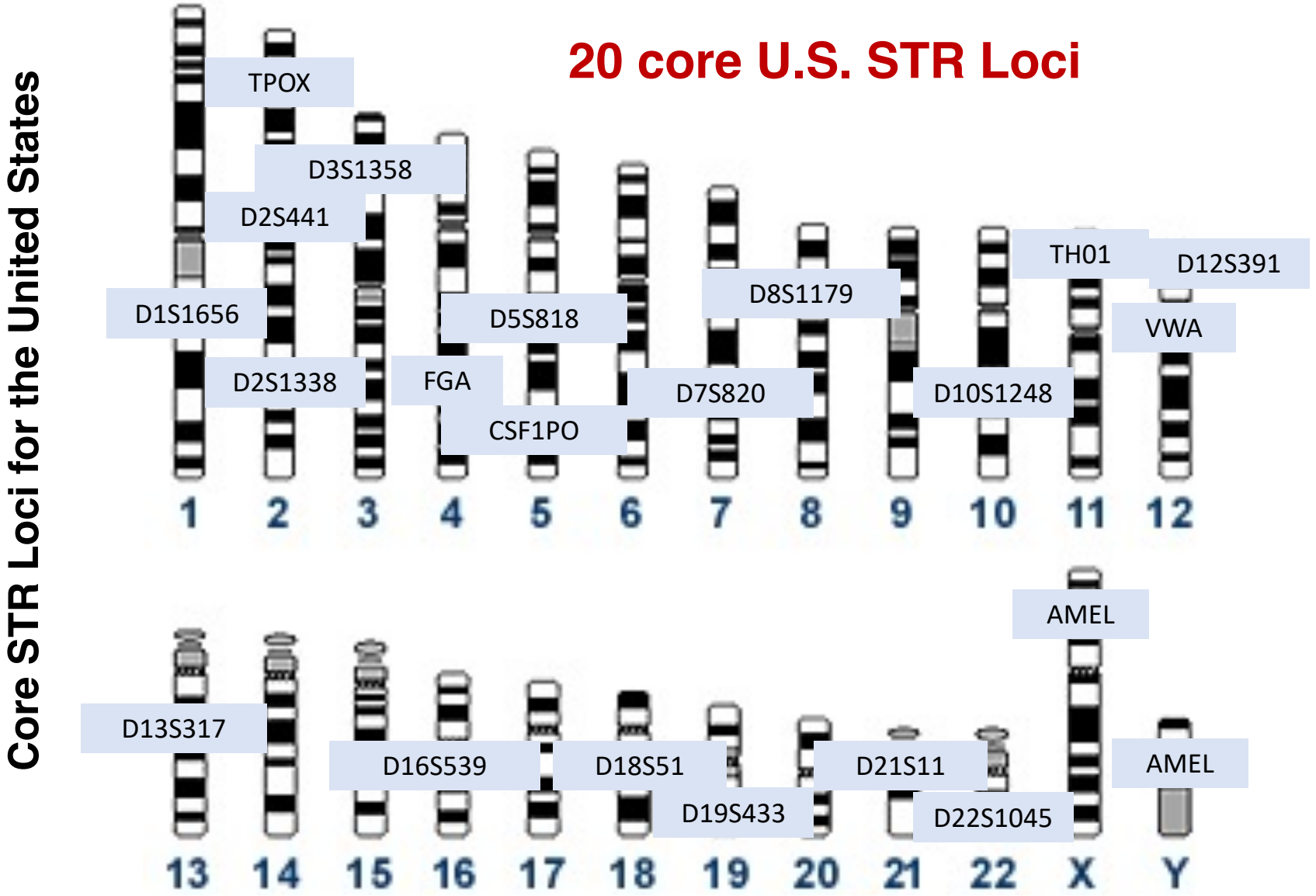
Sampling 100 individuals



Times observed	Frequency	1 in …
5	5%	20
12	12%	8.3
22	22%	4.5
32	32%	3.1
20	20%	5
7	7%	12.3
2	2%	50
= 100	= 100%	

The frequency of these repeats observed in the relevant population have been sampled and are used for the statistical representation of a DNA profile (weight of evidence)

Position of Forensic STR Markers on Human Chromosomes



Forensic DNA Testing

- Probe subsets of genetic variation in order to differentiate between individuals
  - 20 required regions in the human genome (in the U.S.)
- DNA typing must be done efficiently and reproducibly
  - The information must hold up in court
  - FBI Quality Assurance Standards

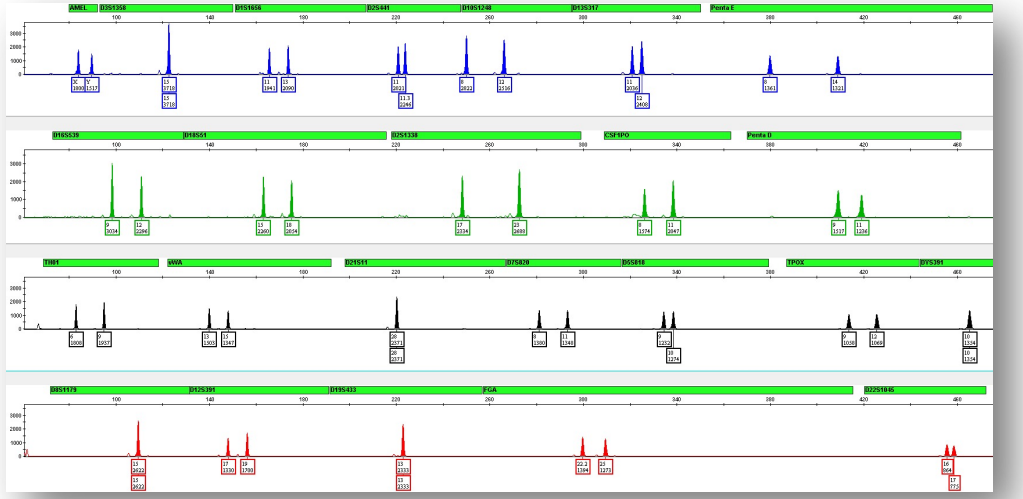
In the U.S. National DNA Database[1]	
Offender profiles	14.5 M
Arrestee profiles	4.3 M
Forensic profiles	1.1 M

- Typically, we are *not* looking at genes – **STR** markers contain little/no information about ancestry, predisposition to disease, or phenotypic information (facial features, eye color, height, hair color) → evolving with **SNP** markers

[1]http://https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics (April 2021)

A STR profile alone is not useful for identification without a reference for comparison

Unknown (evidence)



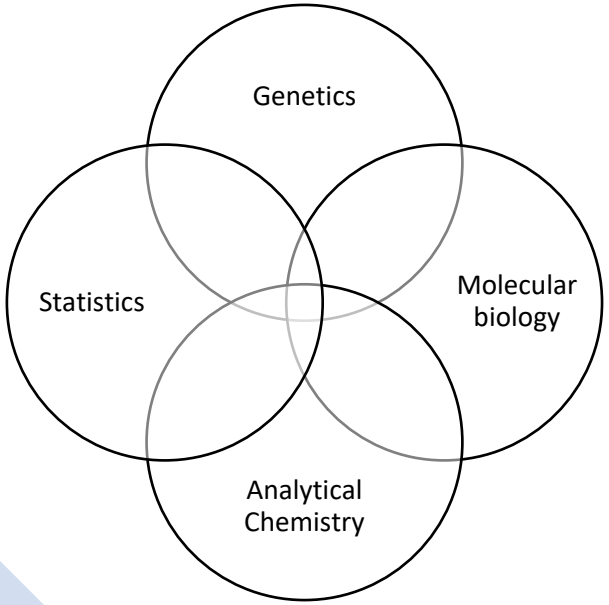
DNA profile developed from evidence

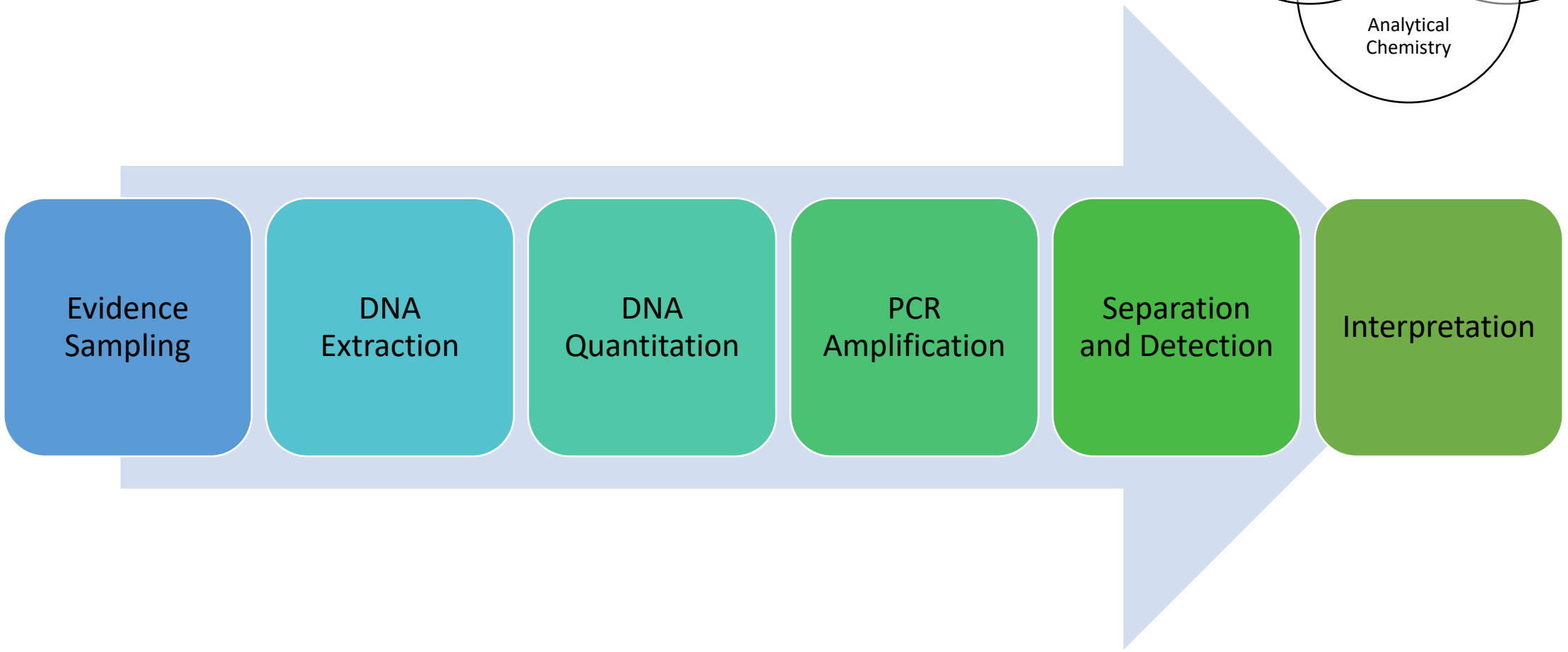
Database

Known profile associated with a name/identity

Reference taken from a suspect

Forensic DNA Typing





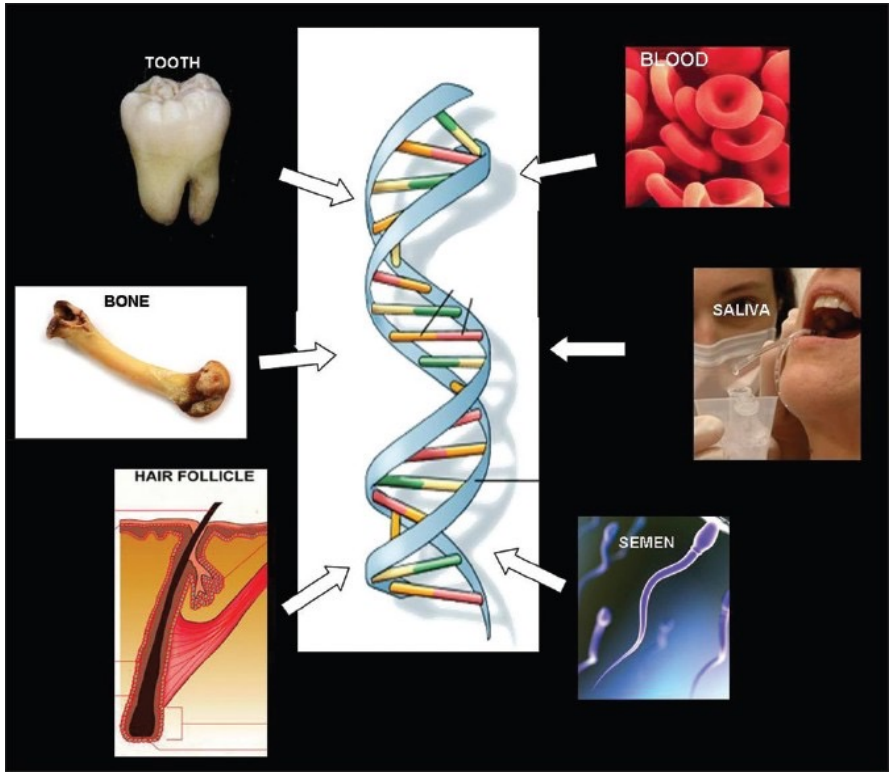
Evidence Sampling

- Goal: recover DNA and cells containing DNA
  - Sampling the evidence

- Swabs (cotton, nylon, other)
  - Wet versus Dry
  - Double swab technique
  - Buccal swab
  - Tape lift – other devices/methods
  - Cuttings and scrapings

- Keep the sample stable until it is ready for analysis
  - Dry and temperature controlled
  - Avoid contamination

Sources of DNA

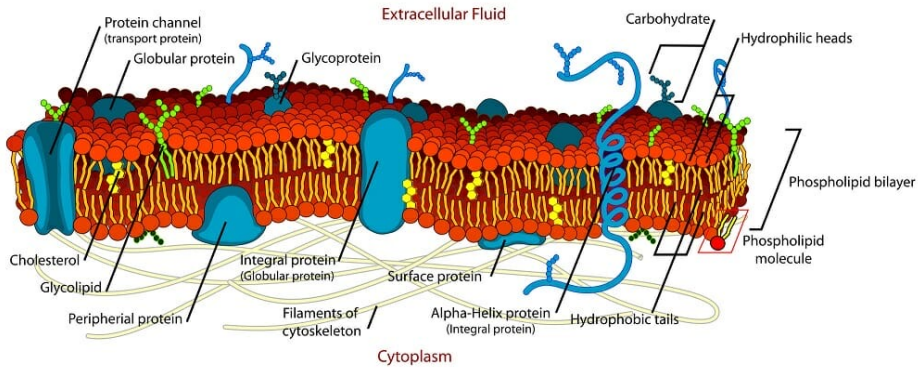


Muruganandhan J, Sivakumar G. Practical aspects of DNA-based forensic studies in dentistry. J Forensic Dent Sci 2011;3:38-45

Evidence Sampling

- Who collects the evidence (analyst, police, dedicated team)?
- What area/region should be swabbed?
- How many swabs should be taken per evidence item?
- Should separate regions be swabbed?

DNA Extraction



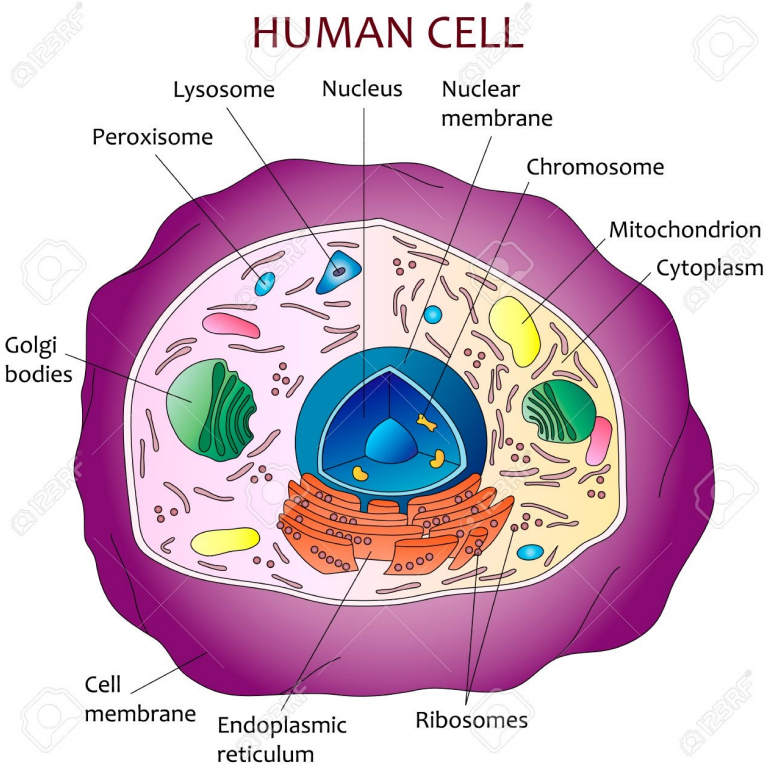
- Goal: lyse the cells, purify, concentrate, and recover genomic DNA
  - Recover material from the collection media

- Chemicals that disrupt the cell membrane
- Conditions that degrade proteins

- Recover the DNA
  - Magnetic bead - silica binding
  - Spin column
  - Organic separation
  - Differential extraction (sperm fraction from total)

Some analyst choices

- Manual versus robotic
- Incubation time (cell lysis)
- Elution volume



https://www.123rf.com/photo_55148693_stock-vector-human-cell-diagram-.html

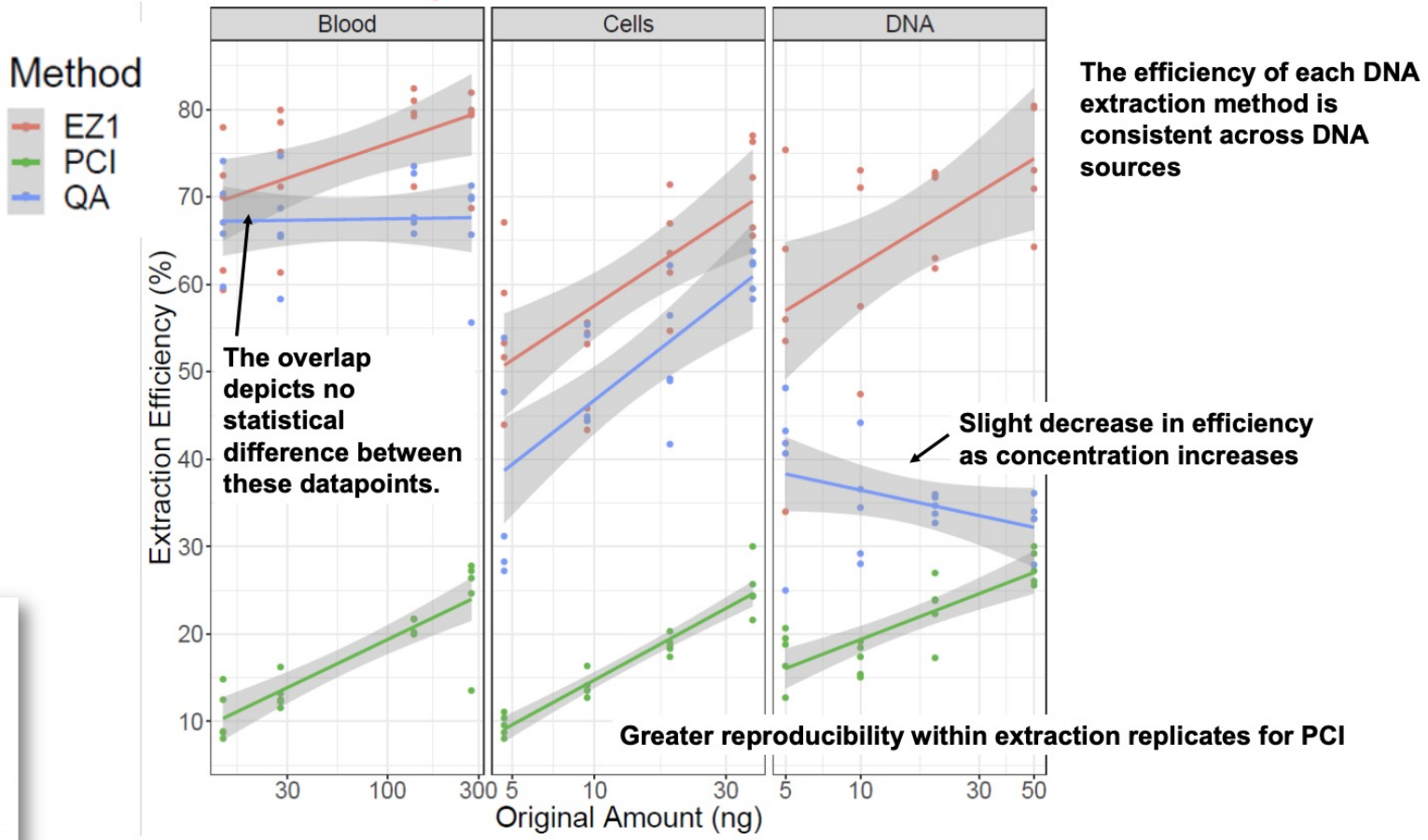
Estimation of Extraction Efficiency by Droplet Digital PCR

Estimation of Extraction Efficiency by Droplet Digital PCR

Erica L. Romsos & Peter M. Vallone
U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA
Email: Erica.Romsos@nist.gov

P391

Extraction Efficiency for three methods across three DNA Sources



Each color represents an extraction method for each of the independent DNA sources. The individual points represent the extraction replicates for each DNA input amount. The slope of the line represents a change in efficiency dependent on original DNA input amount. Only a **slight increase in efficiency is observed across increasing DNA input amounts**, for all extraction methods ***with the exception of DNA with the QIAamp spin columns.***

Research performed to assess the efficiency of different extraction methods across DNA sources

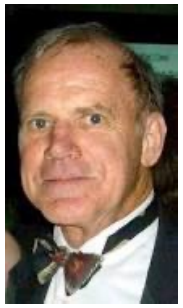
DNA Quantitation

1 cell ≈ 6 pg of DNA (diploid genome)
500 pg is a common target value ≈ 80 cells worth
Would like to be greater than ≈ 100 pg (16 cells worth)

- Goal: quantify the amount of DNA recovered from the extraction process
- Why: the next step (PCR amplification) requires a specific range of input amounts – if not met, interpretation is complex
- Quantitation methods can also inform your workflow
  - Is there enough for one test or many?
  - Low amounts of DNA that can be further concentrated
  - Extent of DNA degradation
  - The ratio of total DNA to male DNA (Y chromosome)
  - Degree of inhibition (agents in the sample that reduce PCR amplification efficiency)
  - Go back and re-sample and/or re-extract
- How is this carried out?

Real-time Quantitative PCR (qPCR)

Polymerase chain reaction (PCR)



Kary Mullis
1944 – 2018
Nobel prize in Chemistry 1993

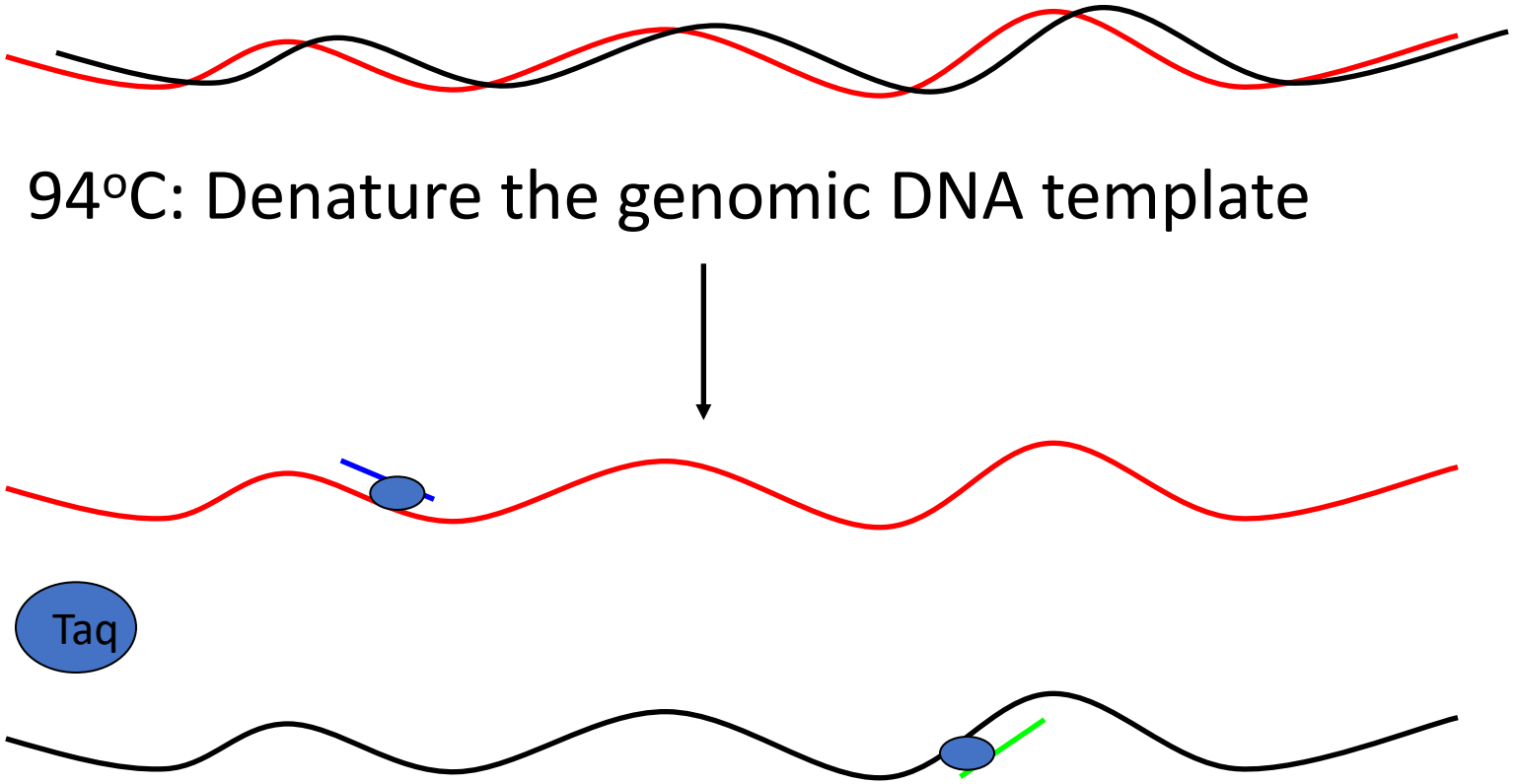
- A means to copy (amplify) the starting DNA template

- Polymerase chain reaction, or PCR, is a laboratory technique used to make multiple copies of a segment of DNA. PCR is very precise and can be used to amplify, or copy, a specific DNA target from a pool of DNA molecules

- Replicate a section of the genome (hundreds of bases long) out of billions of bases

Why: we can't detect a few copies of DNA – need to make billions to detect!

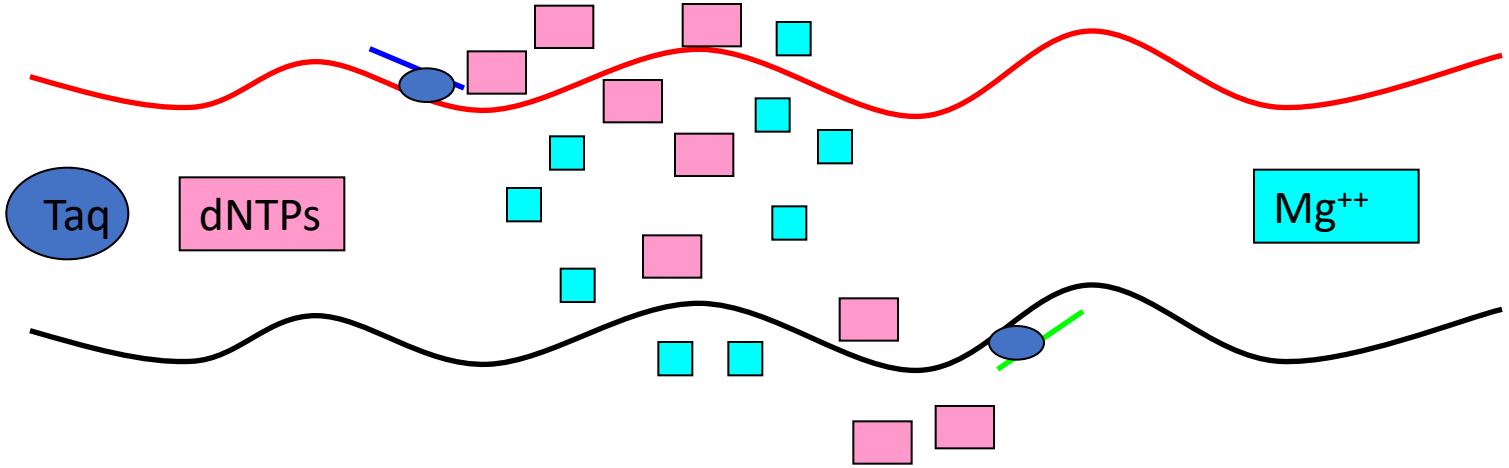
PCR Mechanism



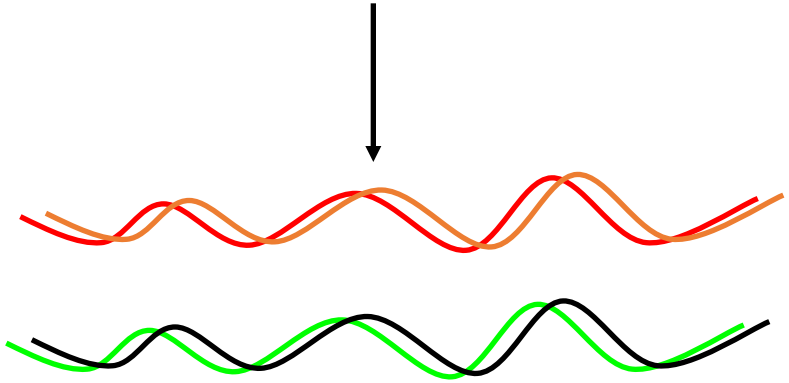
94°C: Denature the genomic DNA template

60°C: PCR primers bind to the DNA template; this will define the size of amplicon

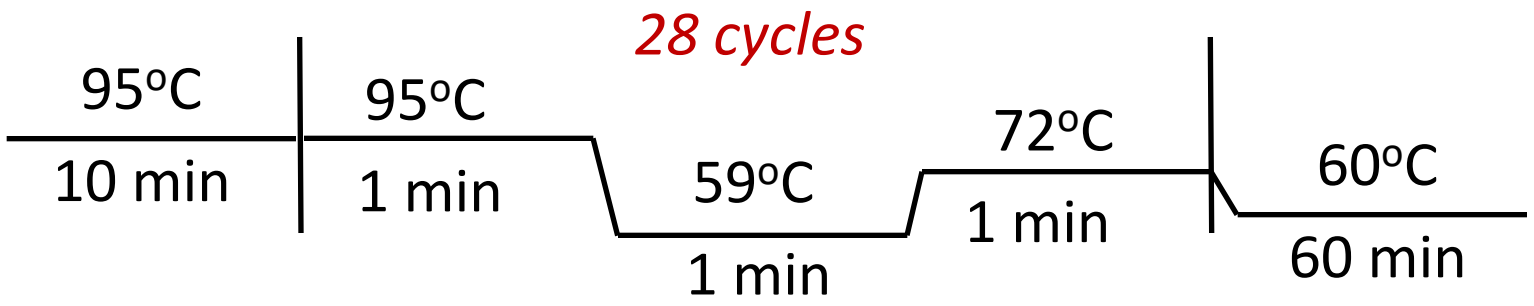
PCR Mechanism



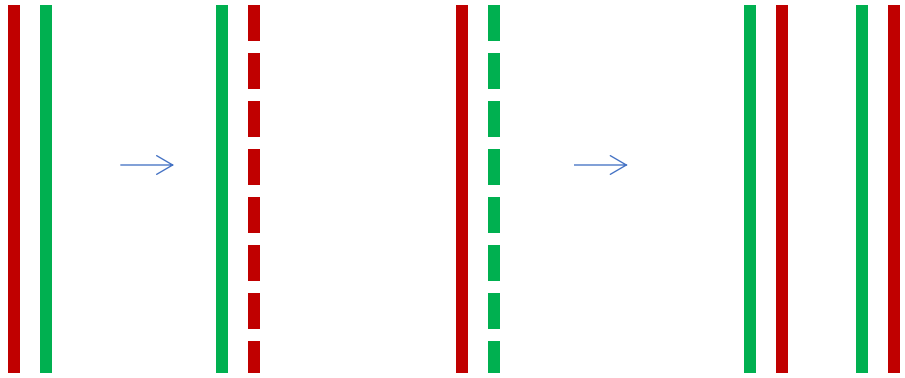
72°C: The Taq primer complex forms and the dNTPs are incorporated into the new strand(s)



PCR Thermal Cycling Profile



Duplex DNA denatures into 2 single strands
Polymerase and NTPs create new duplexes
Repeat!



Polymerase and NTPs

Cycles	Copies
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024
28	268,435,456

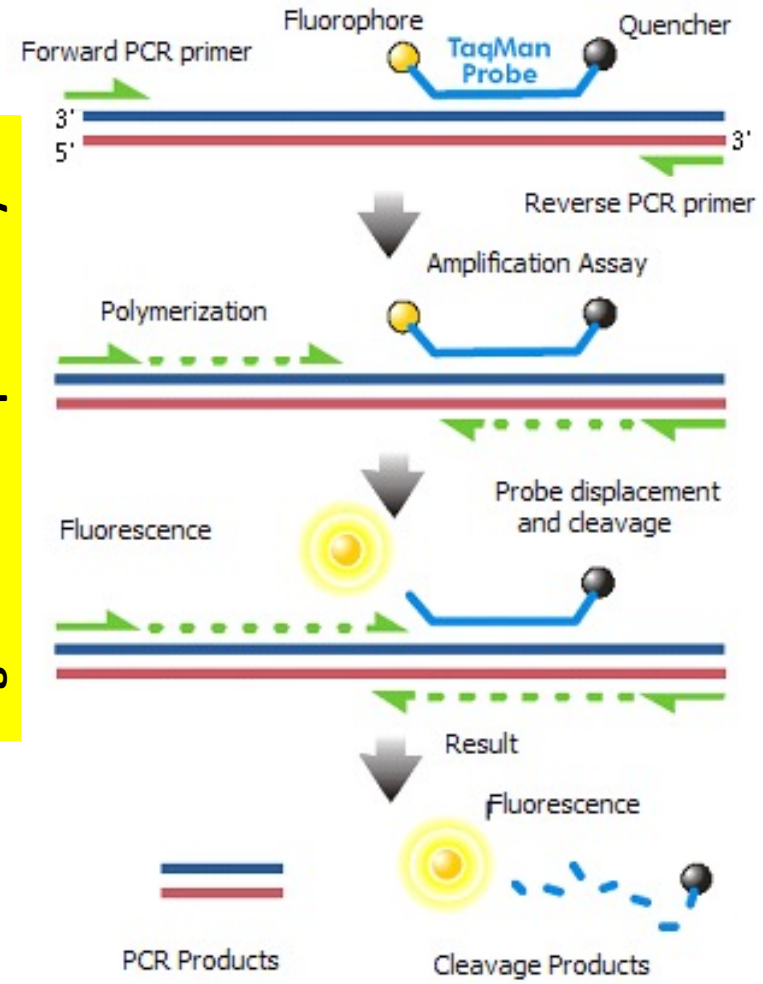
2^N = Copies
N = 'cycles'

Real-time Quantitative PCR (qPCR)

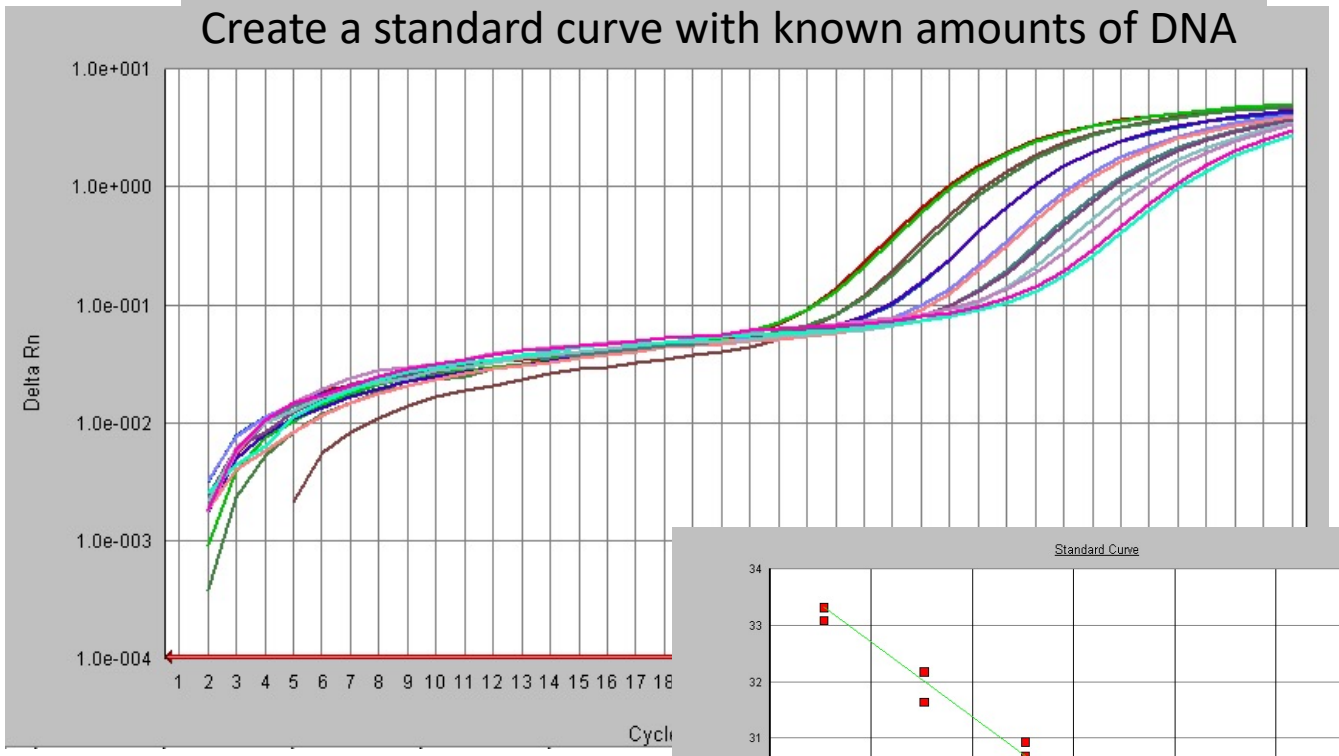
Note that DNA typing is not occurring in this PCR step – just quantification of DNA template

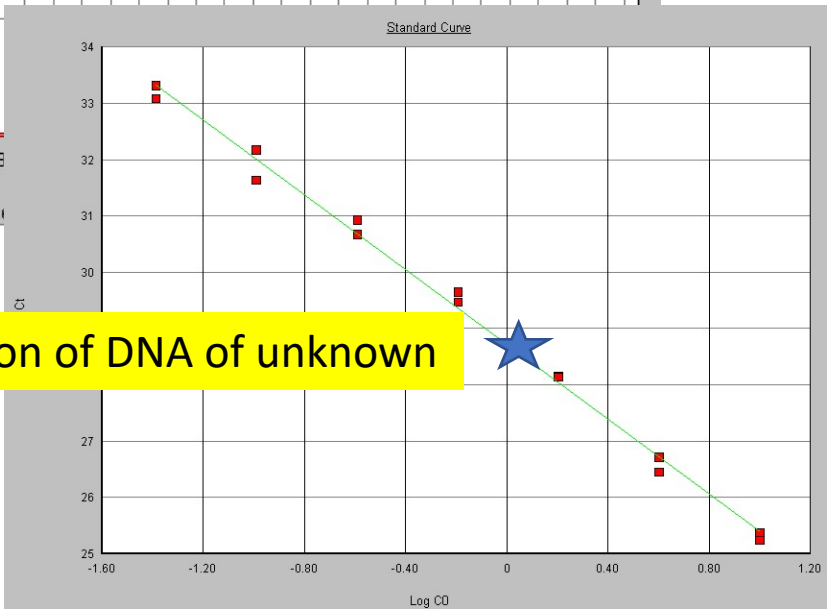
Wait…what is Real-time Quantitative PCR?

Light is emitted per PCR cycle



Create a standard curve with known amounts of DNA



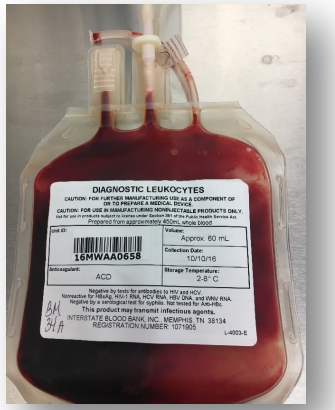


Concentration of DNA of unknown

https://en.wikipedia.org/wiki/TaqMan#/media/File:Taqman.png

SRM 2372a - Human DNA Quantitation Standard

- Certified by dPCR measurements



→

Table 1. Certified Values of Number and Mass Concentration for SRM 2372a[a]

The copy number values are metrologically traceable to the natural units count 1 and ratio 1 and International System of Units (SI) derived units of volume. The DNA mass concentration values are metrologically traceable to the natural units count and ratio 1 and SI derived units of mass and volume.

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

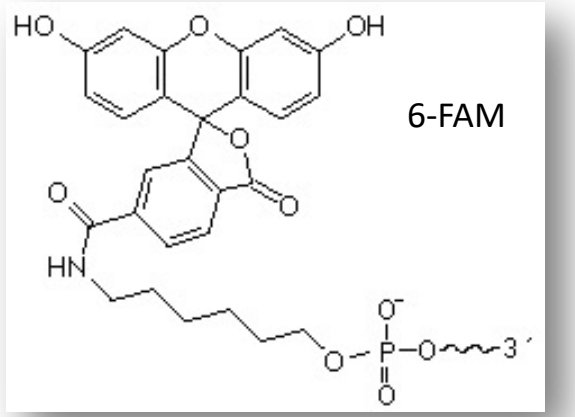
To be used as a qPCR calibrant
OR to assign a value to a 'pot' of DNA – in house or commercial

https://www-s.nist.gov/srmors/view_detail.cfm?srm=2372a

PCR Amplification

- Goal: amplify STR markers and attach fluorescent dye labels (for downstream detection)

- Multiplex PCR – Primers will have a fluorescent dye attached to the 5’ end
- Typically 20 or more STR markers (= 20 sets of compatible PCR primers)
- Well-established commercial PCR kits to carry this out



STR Typing Kits (Multiplex PCR Kits)

Thermo Fisher GlobalFiler

Promega Fusion 6C

QIAGEN Investigator

These all type the U.S. 20 core STR markers

Considerations

- Cost
- Legacy
- Performance
- Availability
- Perceived “sensitivity”
- Locus balance
- Artifacts
- Additional markers (e.g. Y)

Commercial kit testing at NIST

- Testing of commercial products in the **development stage** to provide feedback to the manufacturer ("Beta testing")
- STR tests (CE and NGS); qPCR tests; instrumentation
- Running sets of samples (100 to 1000) to ensure concordance and accuracy

STR kits tested over the past 15 years

Thermo Fisher (10)	Promega (16)	Qiagen (7)	InnoGenomics (1)
MiniFiler	PP CS7	Hexaplex	InnoTyper 21
Identifiler	PP S5	Idplex	
SGM Plus	PP16	ESSplex	
Cofiler	PP16HS	ESSplex SE	
Profiler Plus	PP18D	ESSplex SE Plus	
Sinofiler	PP21	24plex QS	
NGM	SE33 Monoplex	24plex GO!	
NGM Select	PPESI17		
GlobalFiler Express	PPESI17 Pro		
Yfiler Plus	PPESI17 Fast		
	PPESX17		
	PPESX17 Fast		
	PP Fusion		
	PP Fusion 6C		
	PP VersaPlex 27PY		
	PP Y23		

Rapid PCR

Forensic Science International: Genetics 3 (2008) 42–45

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

Demonstration of rapid multiplex PCR amplification involving 16 genetic loci☆

Peter M. Vallone *, Carolyn R. Hill, John M. Butler

National Institute of Standards and Technology, Biochemical Science Division, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899-8311, United States

Electrophoresis 2014, *35*, 3053–3061

Erica L. R. Butts
Peter M. Vallone

National Institute of Standards and Technology, Biomolecular Measurement Division, Gaithersburg, MD, USA

Received April 3, 2014
Revised June 6, 2014
Accepted June 26, 2014

Research Article

Rapid PCR protocols for forensic DNA typing on six thermal cycling platforms

Rapid PCR protocols for the amplification of typing STR multiplexes were evaluated on six different thermal cyclers. Through the use of a faster DNA polymerase coupled with the use of rapid thermal cyclers the amplification cycling times were reduced down to as little as 14 min using PCR primers from the commercially available multiplex STR typing kit Identifiler. Previously described two-step and three-step thermal cycling protocols were evaluated for the six thermal cyclers on 95 unique single-source DNA extracts. CE characterization of the PCR products indicates good peak balance between loci (median values greater than 0.84), and N minus four stutter ratios on averages were 30 to 40% higher than for standard Identifiler PCR conditions. Nonspecific amplification artifacts were observed, but were not observed to migrate within the allele calling bins. With the exception of one locus (D18S51) in a single sample, genotyping results were concordant with manufacturer's recommended amplification conditions utilizing standard thermal cycling procedures. Assay conditions were robust enough to routinely amplify 250 to 500 pg of template DNA. This work describes the protocols for the rapid PCR amplification of STR multiplexes on various PCR thermal cyclers with the future intent to support validation for typing single-source samples in a database laboratory.

Forensic Science International: Genetics 18 (2015) 90–99

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Review

Rapid PCR of STR markers: Applications to human identification

Erica L. Romsos *, Peter M. Vallone

National Institute of Standards and Technology, 100 Bureau Drive, MS 8314, Gaithersburg, MD 20899-8314, USA

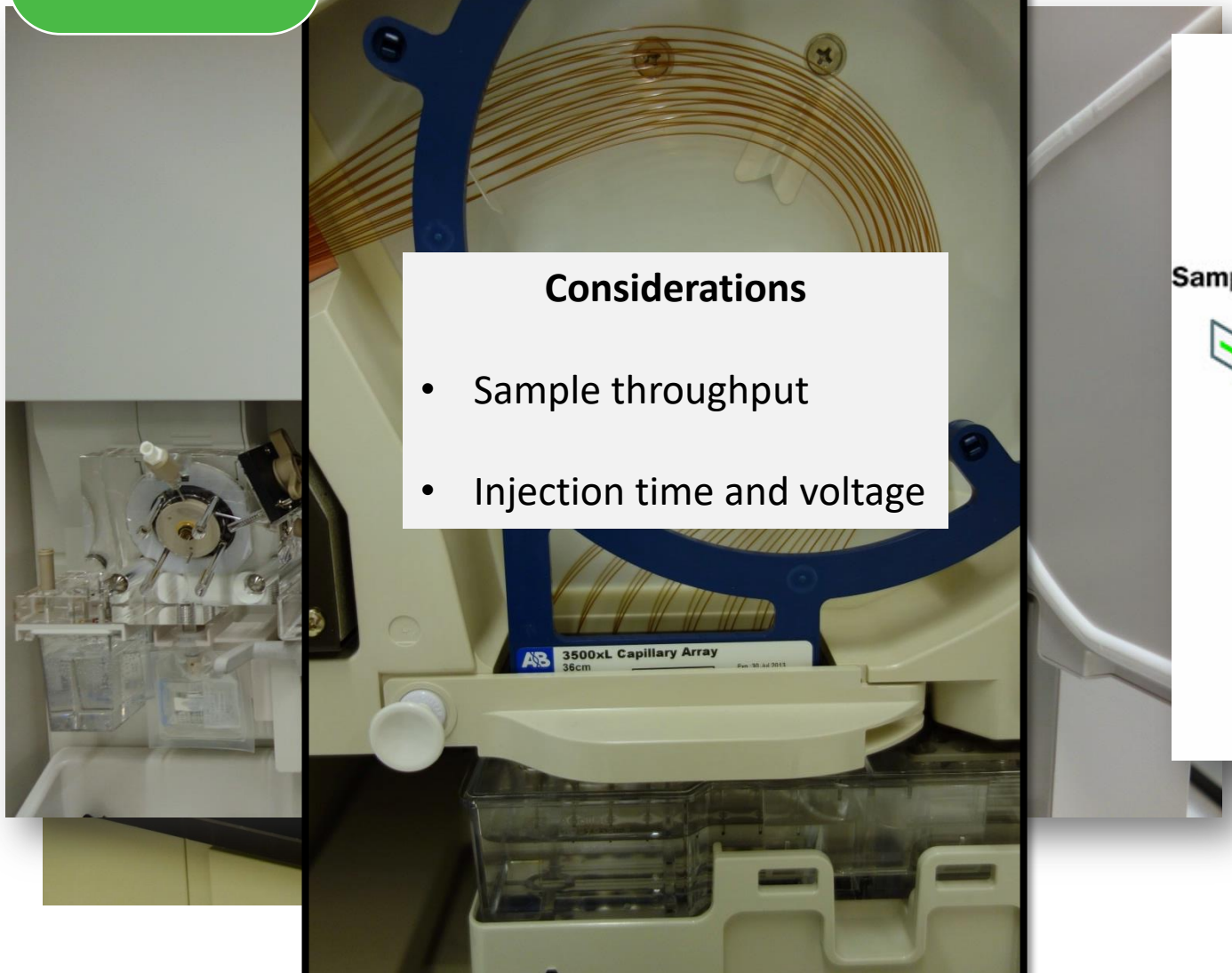
Work performed at NIST a rapid PCR of a STR multiplex in less than 14 min (compared to 2-3 hours)

Separation and Detection

- Goal: Separate and detect PCR fragments of differing lengths
- Each fragment is ‘labeled’ with a fluorescent dye during PCR
- This is performed by Capillary Electrophoresis and Fluorescence Detection

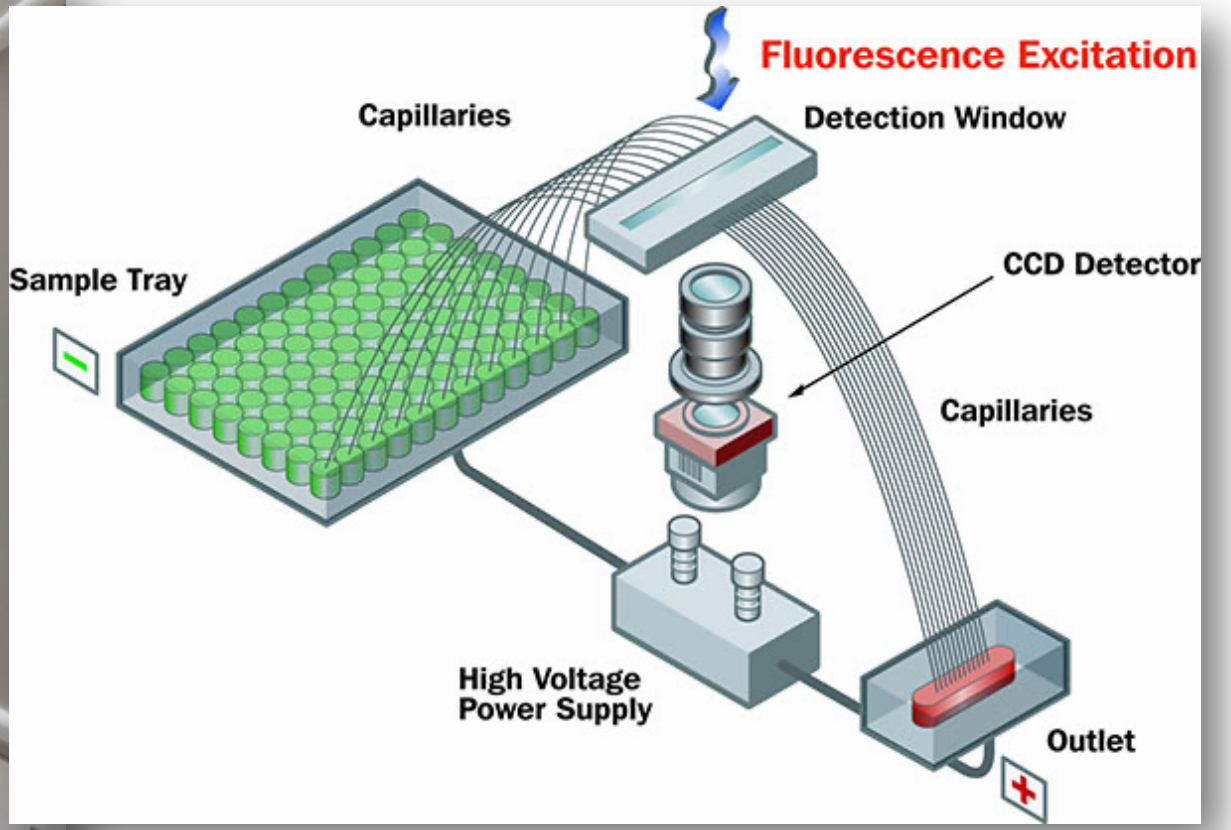
Separation and Detection

Capillary Electrophoresis



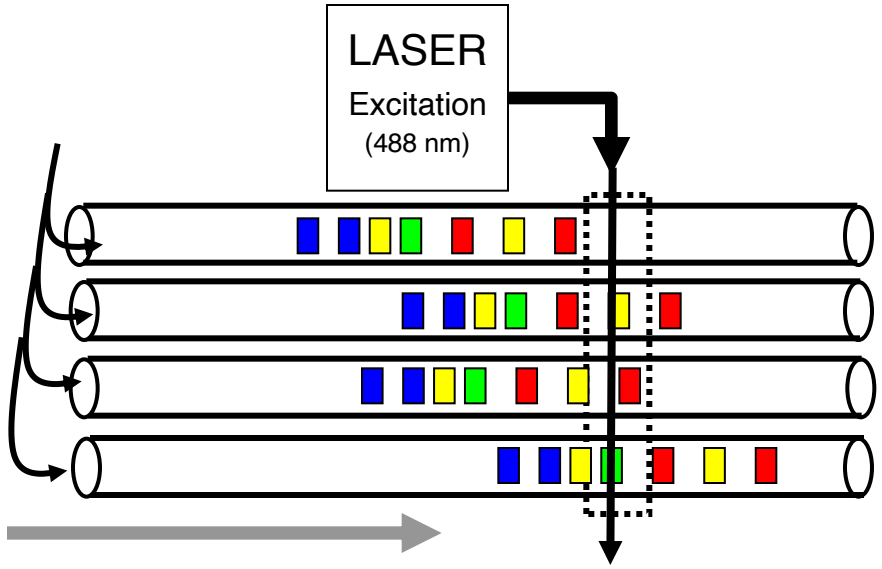
Considerations

- Sample throughput
- Injection time and voltage



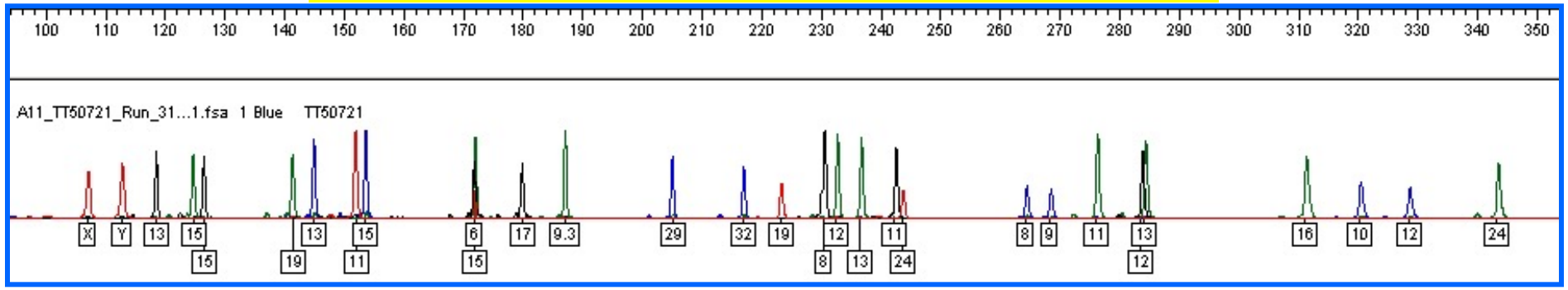
http://www.philekorea.co.kr/kor/data/cheditor4/1305/a49f0e4c8103cb5277053a548368c0d8_nC5e7v8vWJZS.jpg

Separation and Detection



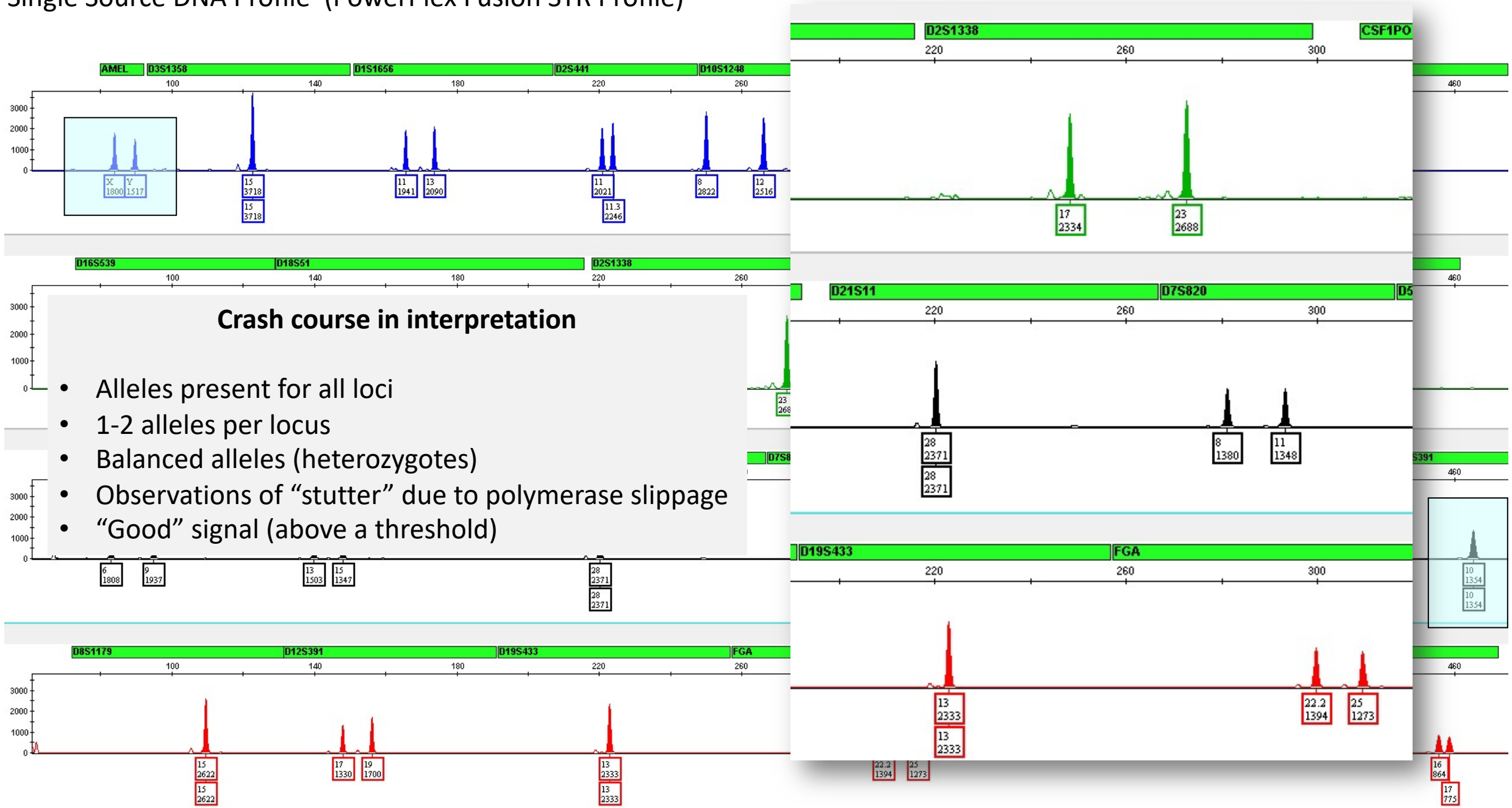
The labeled fragments are separated based on size and detected on a gel or capillary electrophoresis instrument
~1 hour or less

Fragment size ranges from 100 - 500 base pairs



Peaks represent labeled DNA fragments separated by electrophoresis
This ‘profile of peaks’ is unique for an individual – a DNA type

Single Source DNA Profile (PowerPlex Fusion STR Profile)

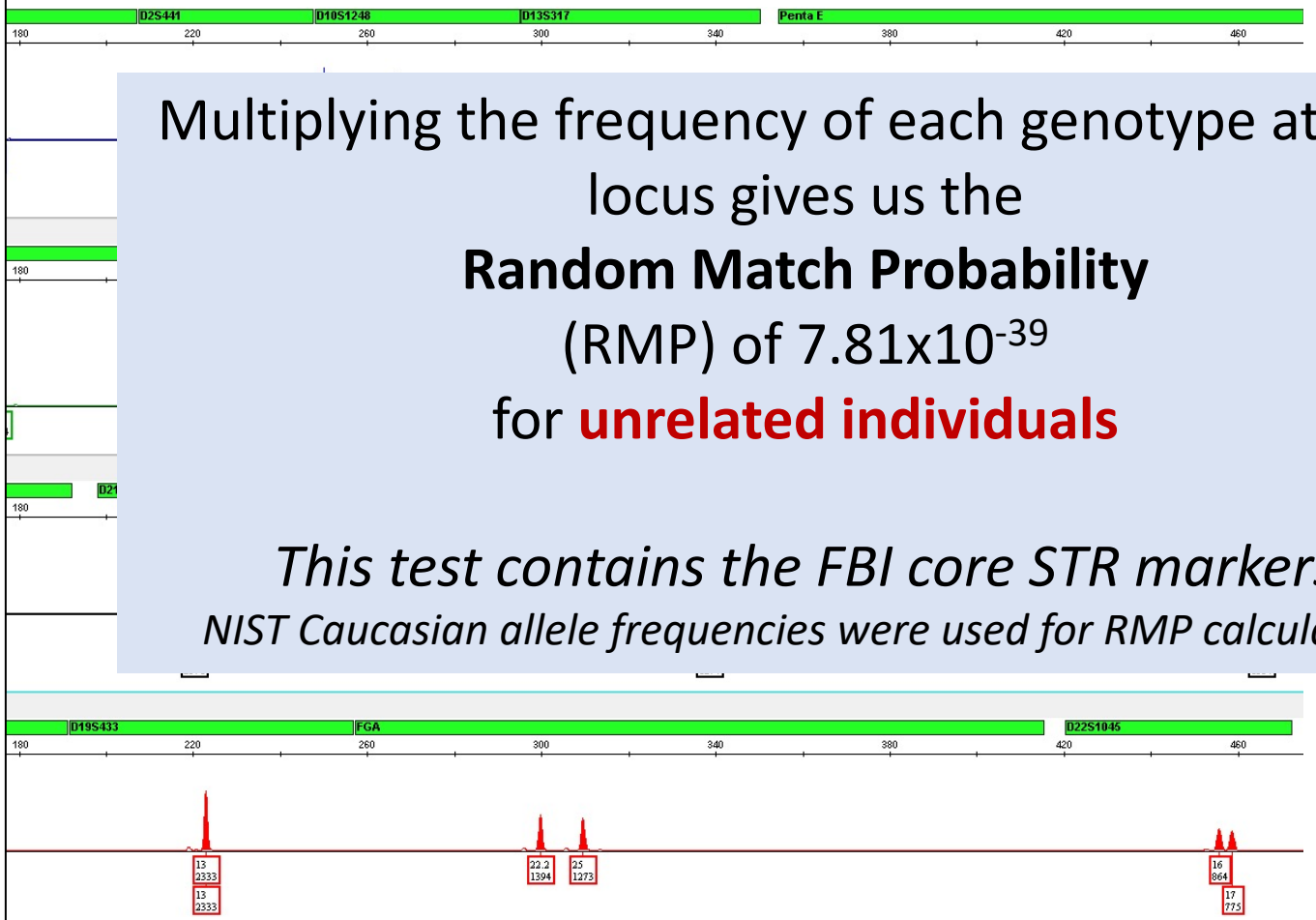


Crash course in interpretation

- Alleles present for all loci
- 1-2 alleles per locus
- Balanced alleles (heterozygotes)
- Observations of “stutter” due to polymerase slippage
- “Good” signal (above a threshold)

This information goes into a database (not the electropherogram)

Amel	{X,Y}
D3S1358	{15,15}
D1S1656	{11,13}
D2S441	{11,11.3}
D10S1248	{8,12}
D13S317	{11,12}
Penta E	{8,14}
D16S539	{9,12}
D18S51	{15,18}
D2S1338	{17,23}
CSF1PO	{8,11}
Penta D	{9,11}
TH01	{6,9}
VWA	{13,15}
D21S11	{28,28}
D7S820	{8,11}
D5S818	{9,10}
TPOX	{9,12}
DYS391	{10}
D8S1179	{15,15}
D12S391	{17,19}
D19S433	{13,13}
FGA	{22.2,25}
D22S1045	{16,17}



Multiplying the frequency of each genotype at each locus gives us the

Random Match Probability

(RMP) of 7.81×10^{-39}

for **unrelated individuals**

This test contains the FBI core STR markers

NIST Caucasian allele frequencies were used for RMP calculation

Allele Frequencies published by NIST

J Forensic Sci, July 2003, Vol. 48, No. 4
Paper ID JFS2003045_484
Published 19 May 2003
Available online at: www.astm.org

FOR THE RECORD

John M. Butler,[1] Ph.D.; Richard Schoske,[1] M.A.; Peter M. Vallone,[1] Ph.D.; Janette W. Redman[1]; and Margaret C. Kline,[1] M.S.

Allele Frequencies for 15 Autosomal STR Loci on U.S. Caucasian, African American, and Hispanic Populations*

Forensic Science International: Genetics 7 (2013) e82–e83

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Letter to the Editor

U.S. population data for 29 autosomal STR loci

Dear Editor,

run and population statistics were confirmed using the PowerMarker v3.25 statistics program [10].

There were 14 instances where statistically significant deviations from Hardy–Weinberg expectations based on the exact test

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Correspondence

Corrigendum to 'U.S. Population Data for 29 Autosomal STR Loci' [Forensic Sci. Int. Genet. 7 (2013) e82–e83]

Carolyn R. Steffen, Michael D. Coble, Katherine B. Gettings, Peter M. Vallone*

National Institute of Standards and Technology, Material Measurement Laboratory, Gaithersburg, MD 20899-8314, United States

Data used by U.S. labs to assign weight of evidence to a DNA profile

SRM 2391d – PCR-based DNA Profiling Standard

Brief history of SRM 2391 series…

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 database analysis.

From FBI Quality Assurance Standards

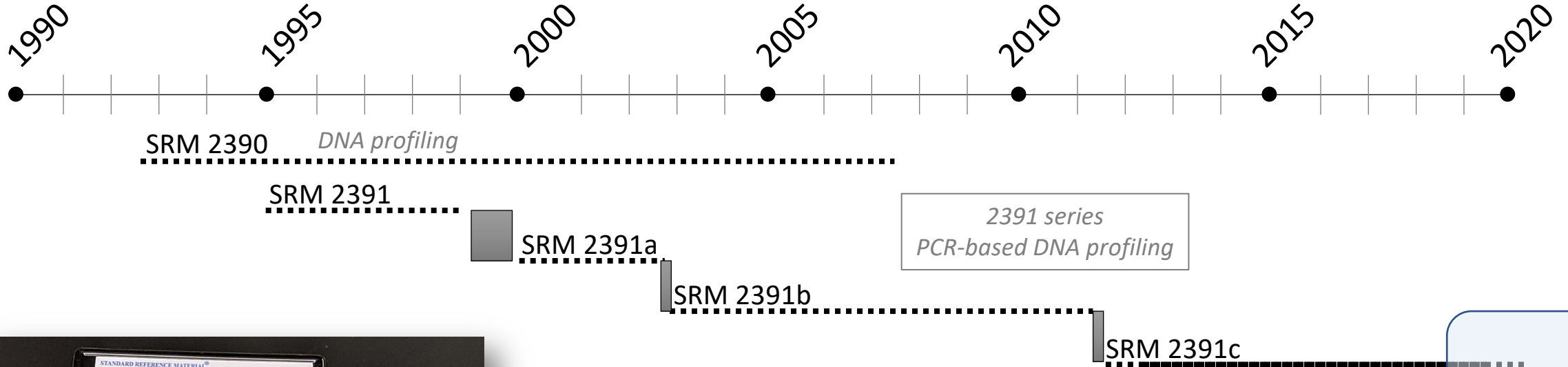
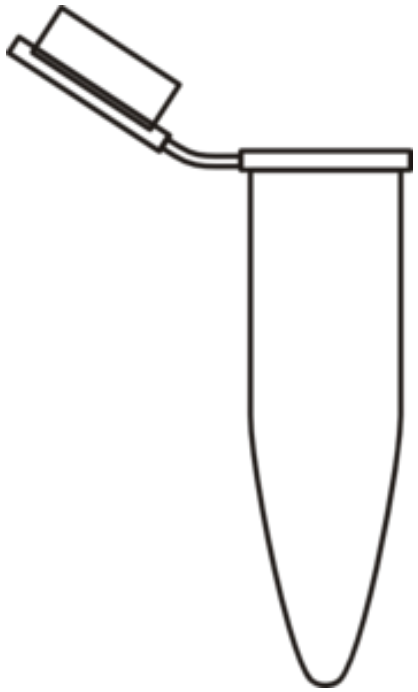


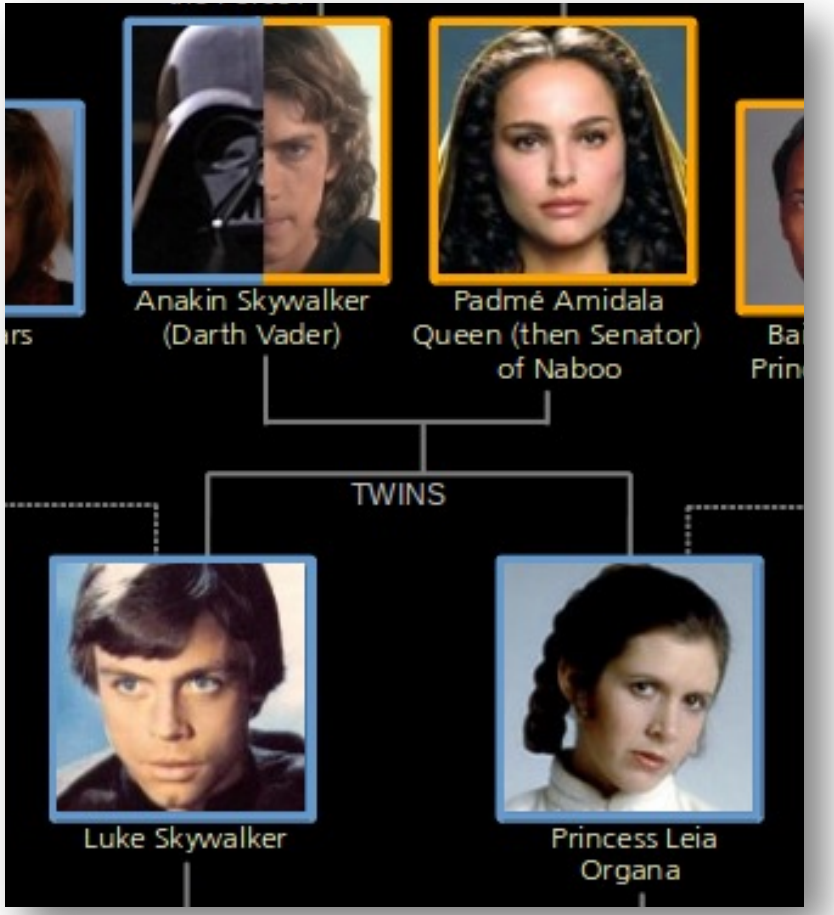
Table 1. Description of Components in SRM 2391d

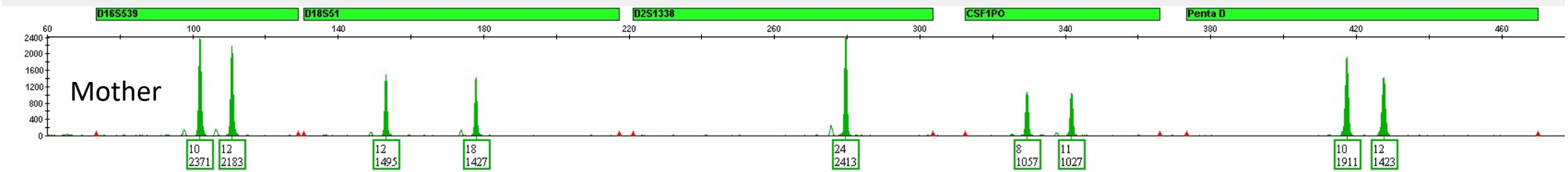
Component	Description	Volume	Concentration[a]
A	Anonymous single-source female genomic DNA in TE^{-4} buffer	55 µL	1.6 ± 0.5 ng/µL
B	Anonymous single-source male genomic DNA in TE^{-4} buffer	55 µL	1.7 ± 0.5 ng/µL
C	Anonymous single-source male genomic DNA in TE^{-4} buffer	55 µL	1.6 ± 0.2 ng/µL
D	Mixed-source, 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

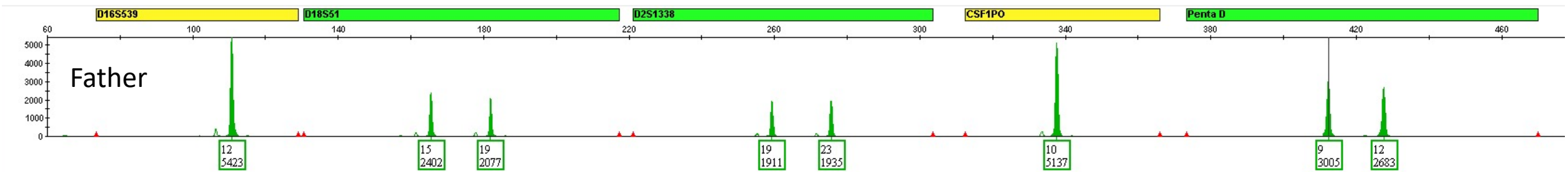
[a] DNA concentrations and cell counts are provided as Information Values.
[b] FTA paper cards contain chemicals that lyse cells, denature proteins and protect nucleic acids from nucleases, oxidation and UV damage. FTA cards rapidly inactivate organisms, including blood-borne pathogens, and prevent the growth of bacteria and other microorganisms.

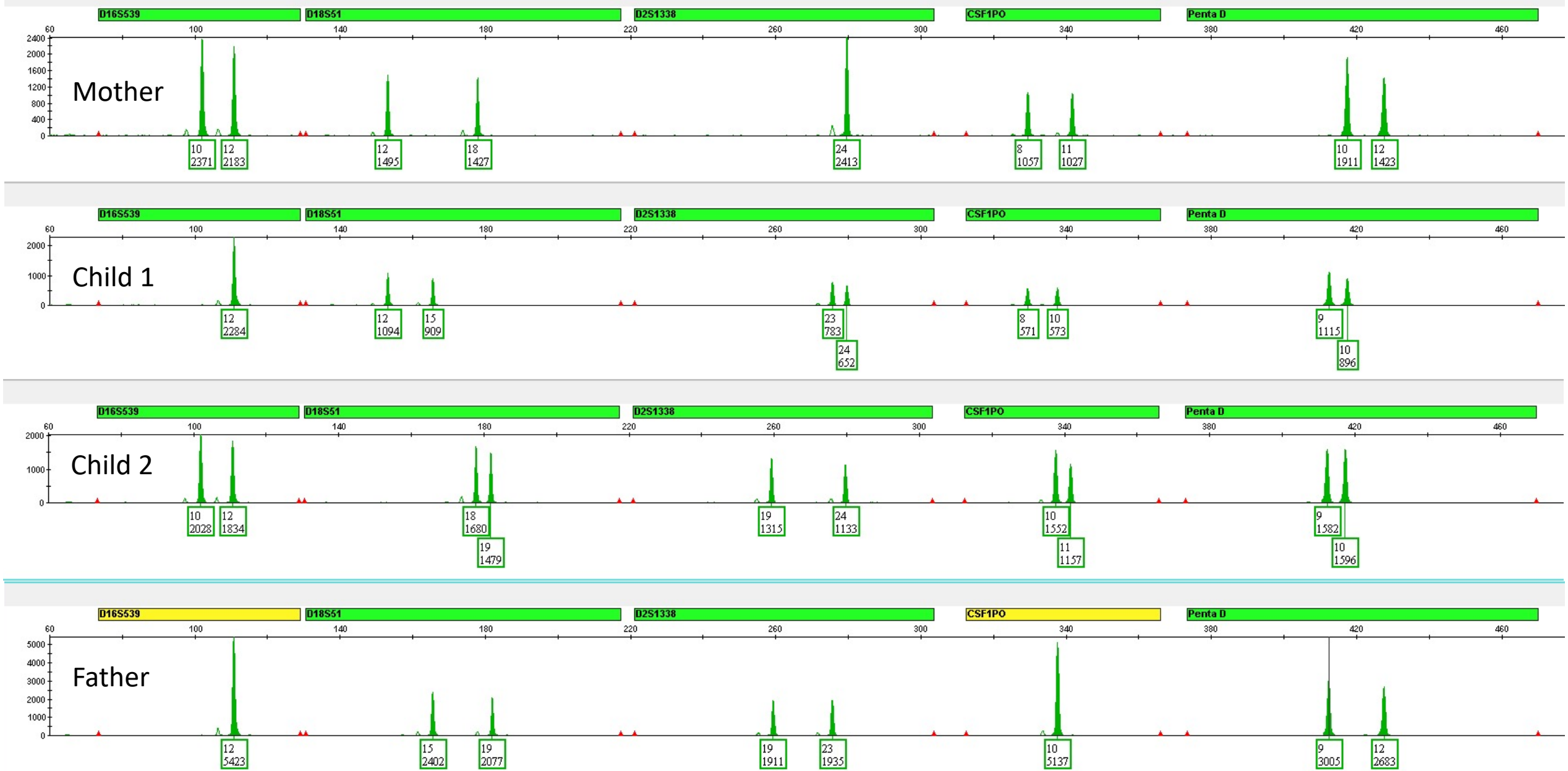


Kinship


Anakin Skywalker
(Darth Vader)
Padmé Amidala
Queen (then Senator)
of Naboo
TWINS
Luke Skywalker
Princess Leia
Organa


D16S539
D18S51
D2S1338
CSF1PO
Penta D
Mother
10 2371
12 2183
12 1495
18 1427
24 2413
8 1057
11 1027
10 1911
12 1423

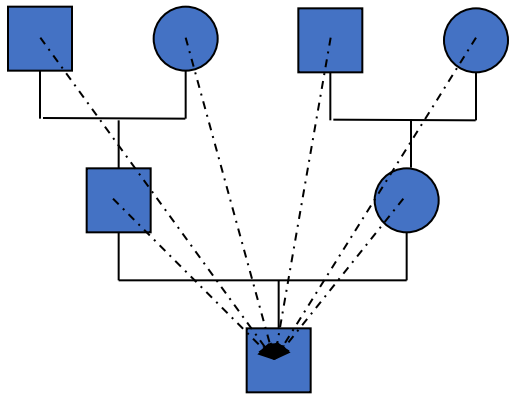

D16S539
D18S51
D2S1338
CSF1PO
Penta D
Father
12 5423
15 2402
19 2077
19 1911
23 1935
10 5137
9 3005
12 2683


D16S539
D18S51
D2S1338
CSF1PO
Penta D
Mother
10 2371
12 2183
12 1495
18 1427
24 2413
8 1057
11 1027
10 1911
12 1423
Child 1
12 2284
12 1094
15 909
23 783
24 652
8 571
10 573
9 1115
10 896
Child 2
10 2028
12 1834
18 1680
19 1479
19 1315
24 1133
10 1552
11 1157
9 1582
10 1596
Father
12 5423
15 2402
19 2077
19 1911
23 1935
10 5137
9 3005
12 2683

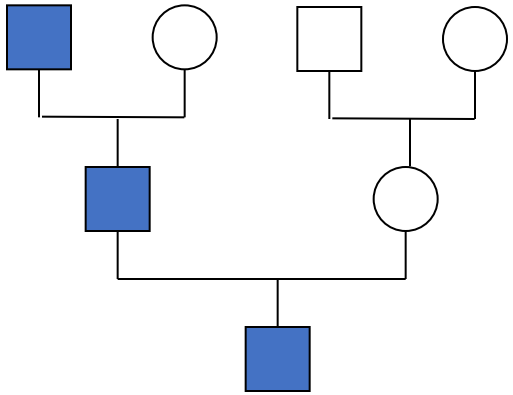
Different Inheritance Patterns

Lineage Markers

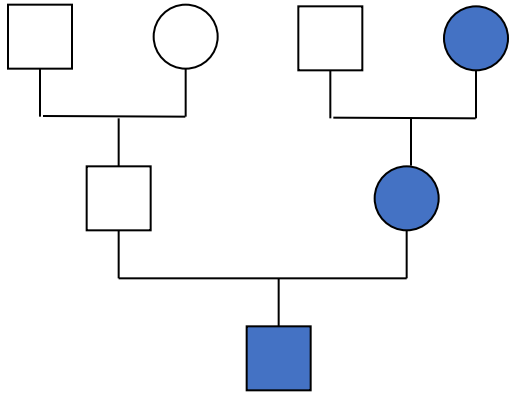
CODIS STR Loci



Autosomal
(passed on in part,
from all ancestors)



Y-Chromosome
(passed on complete,
but only by sons)



Mitochondrial
(passed on complete,
but only by daughters)

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Interpretation

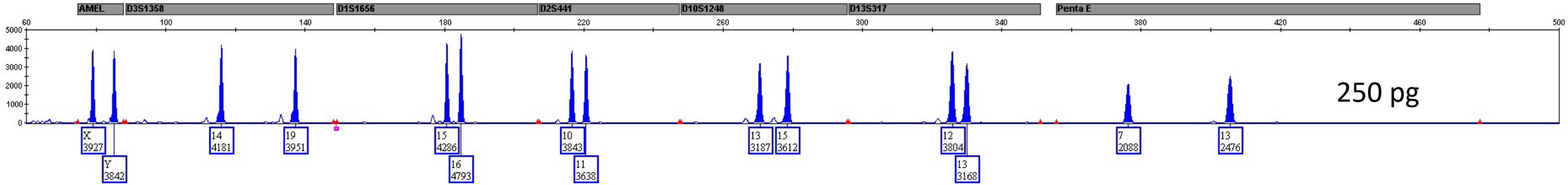
In forensic casework profile interpretation is more involved

Examples…

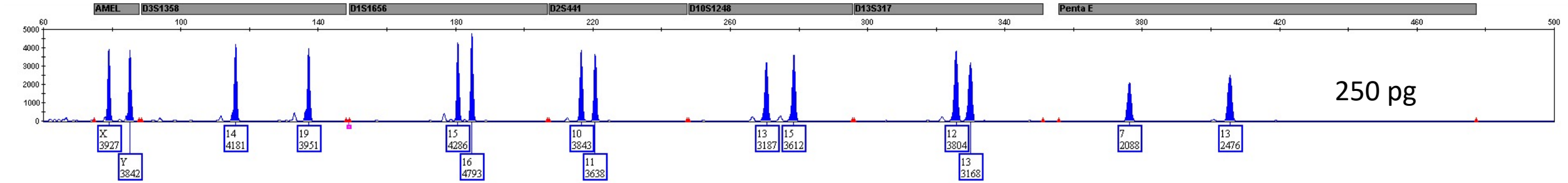
Low template DNA Example

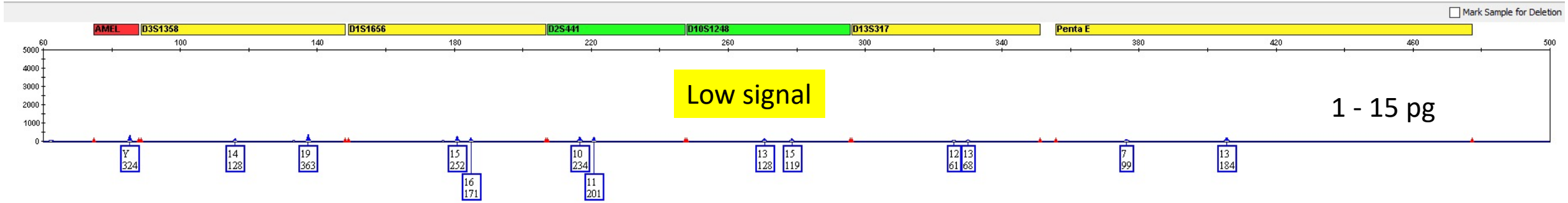
Stochastic variation

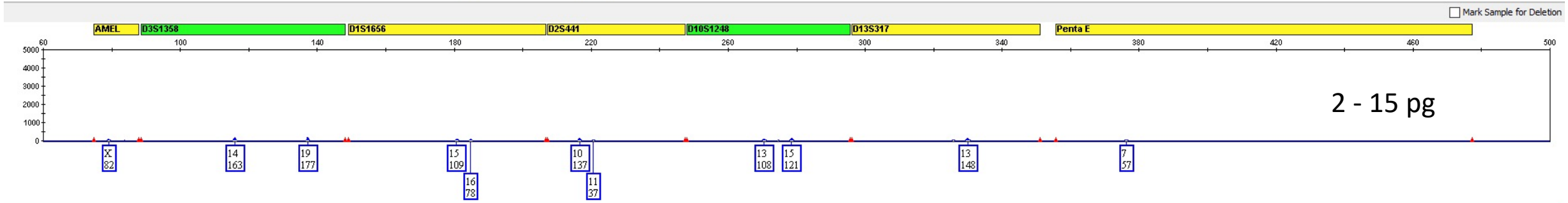
Single source

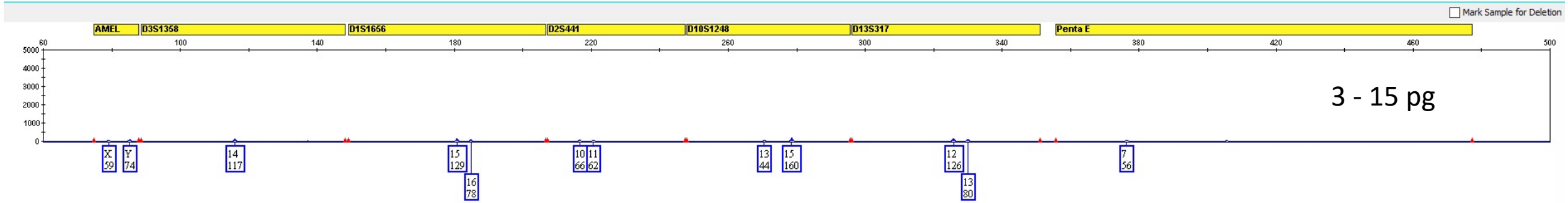


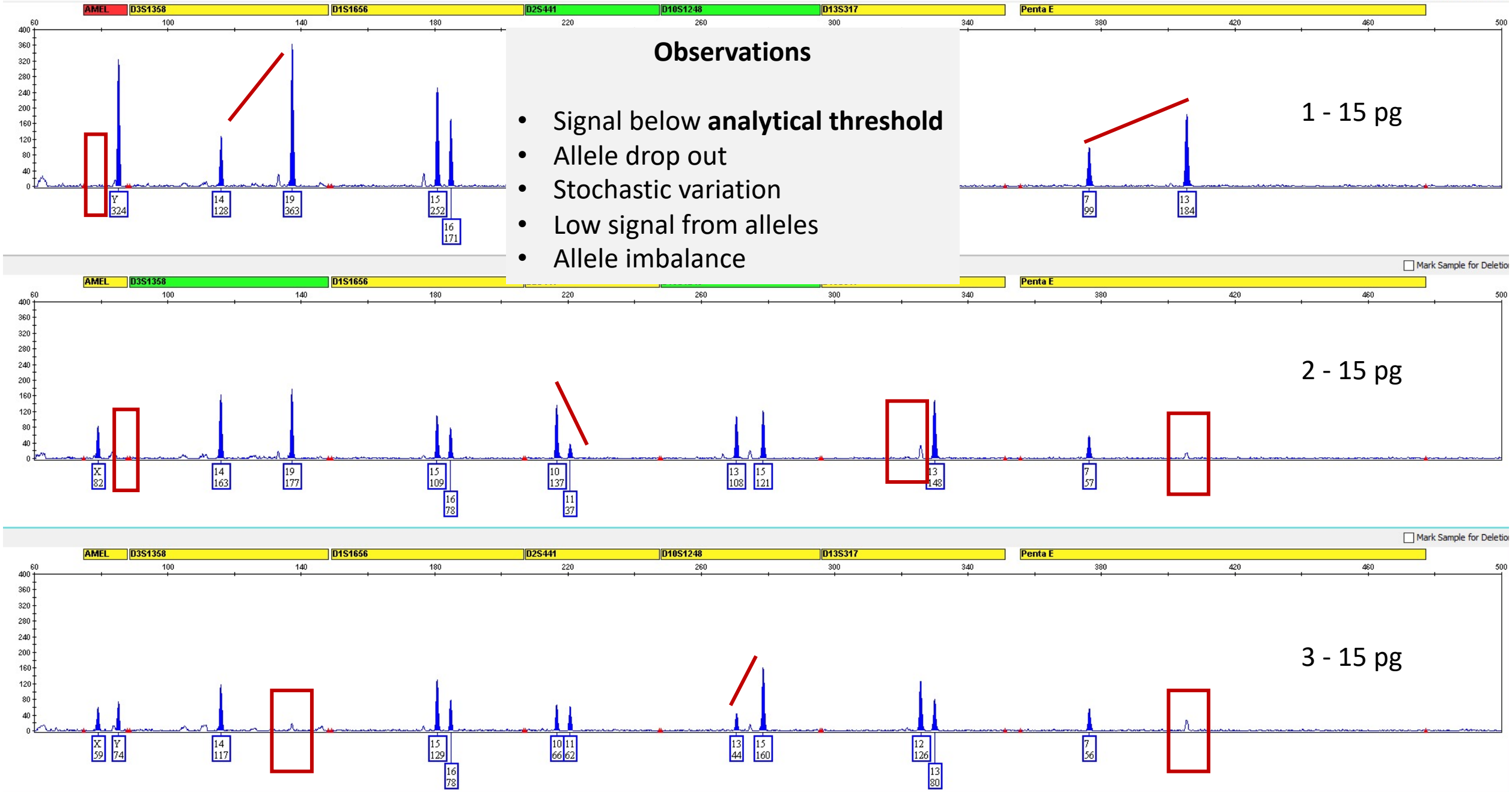
Single source







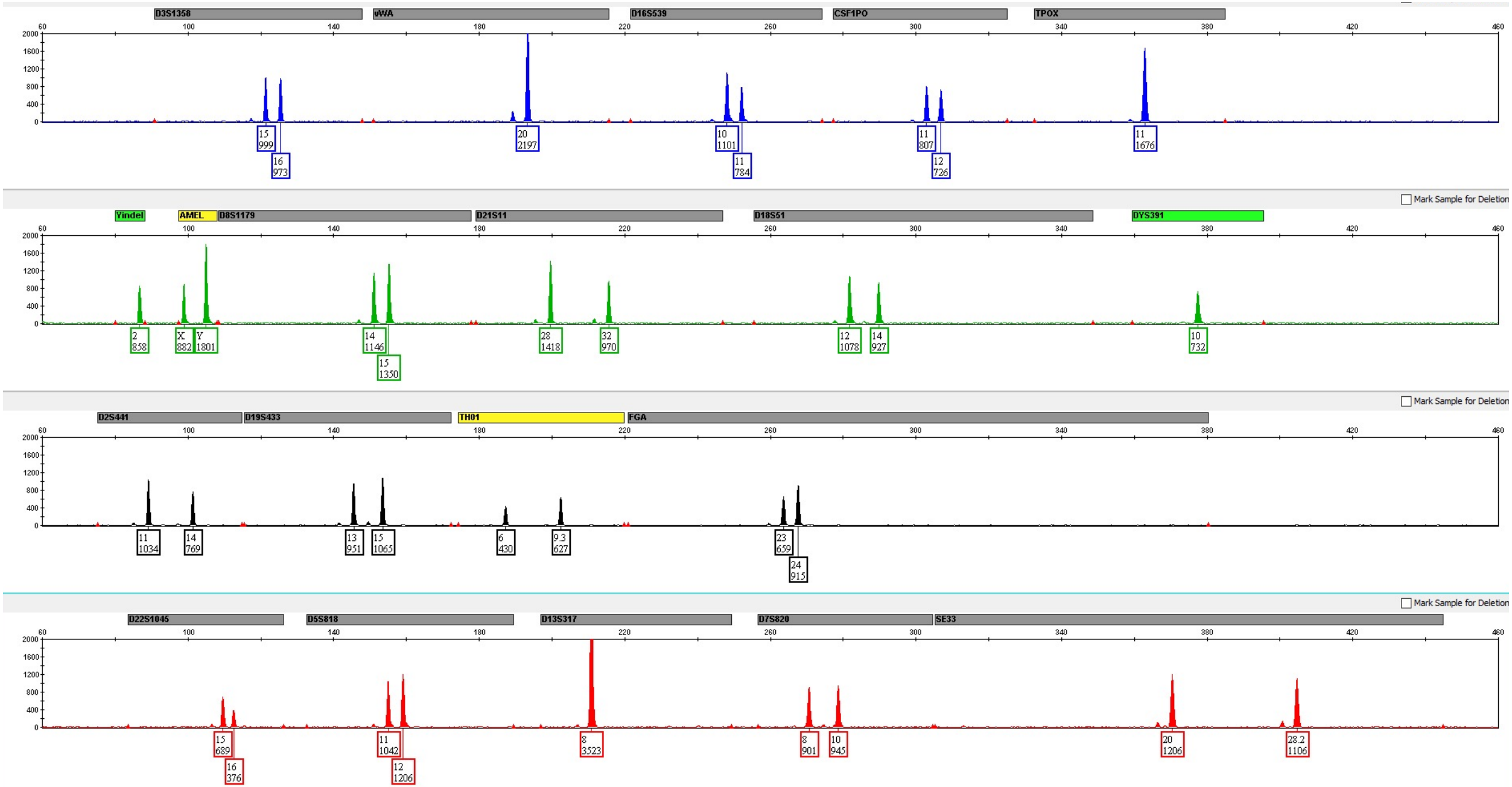



AMEL
D3S1358
D1S1656
D2S441
D10S1248
D13S317
Penta E
Observations
• Signal below **analytical threshold**
• Allele drop out
• Stochastic variation
• Low signal from alleles
• Allele imbalance
1 - 15 pg
2 - 15 pg
3 - 15 pg
Mark Sample for Deletio

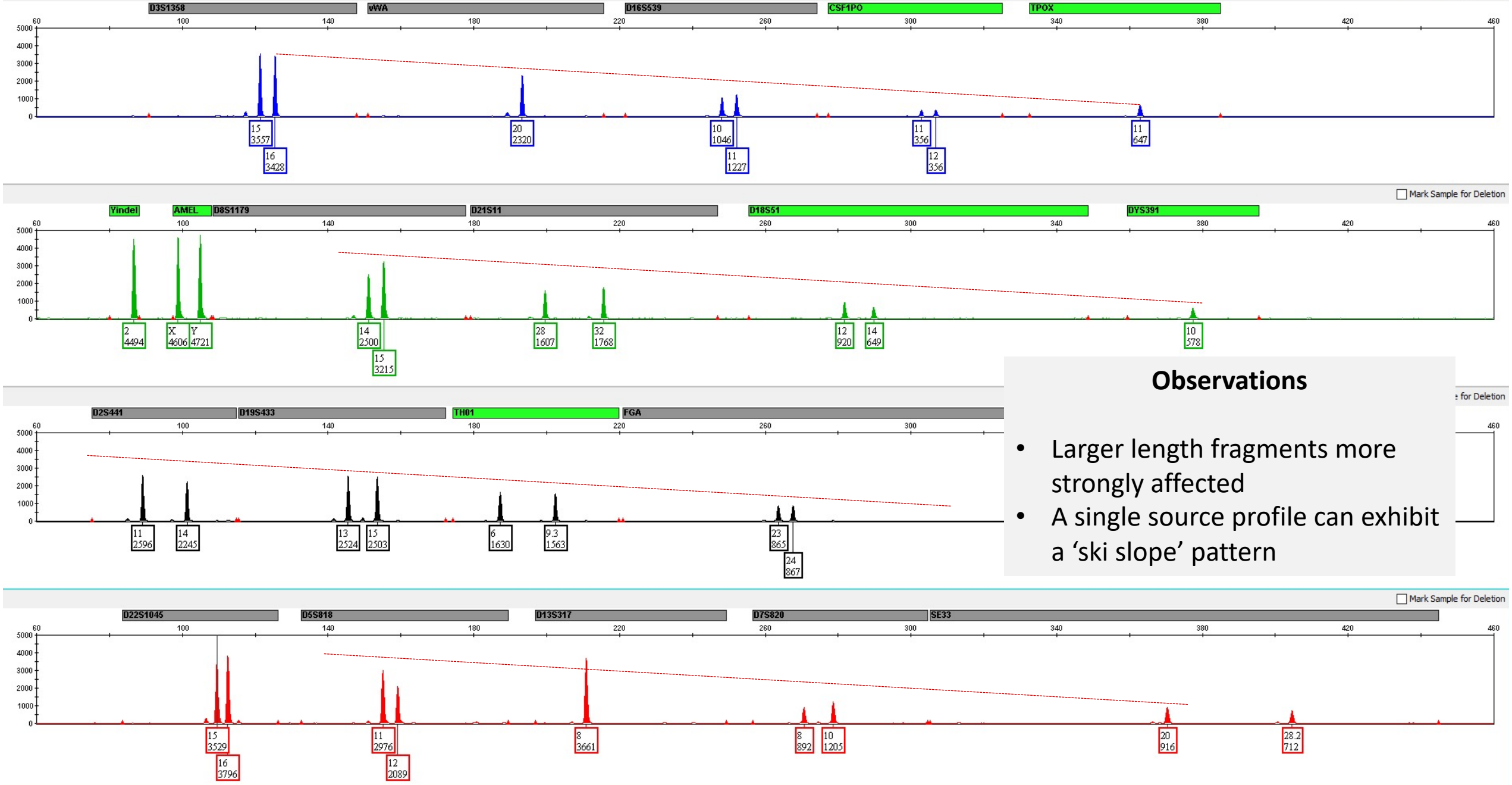
Degraded DNA Example

https://lftdi.camden.rutgers.edu/provedit/

Sample 44 - single source - pristine



Sample 44 - single source - degraded - high template



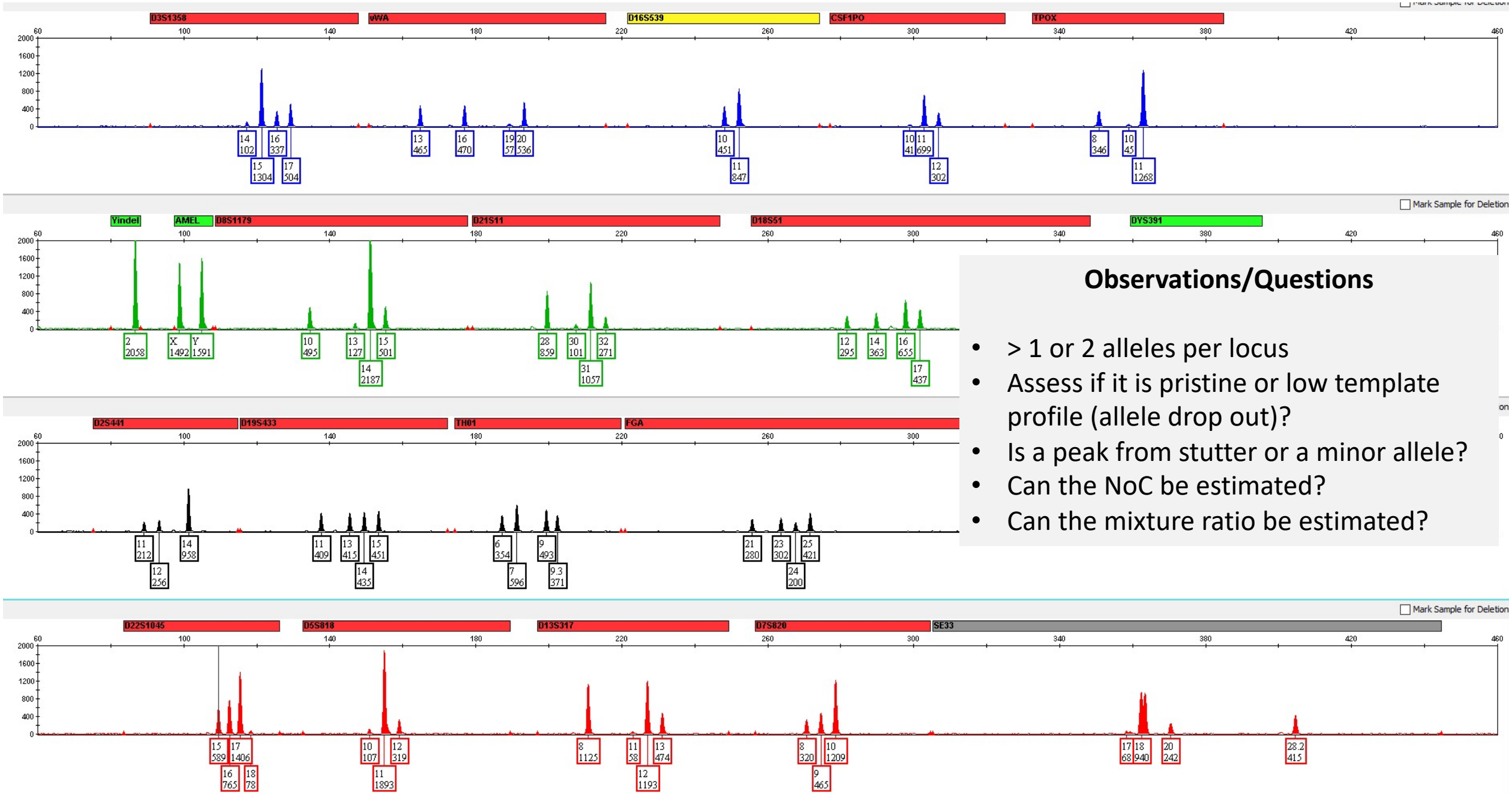
Observations

- Larger length fragments more strongly affected
- A single source profile can exhibit a ‘ski slope’ pattern

Mixture Examples

https://lftdi.camden.rutgers.edu/provedit/

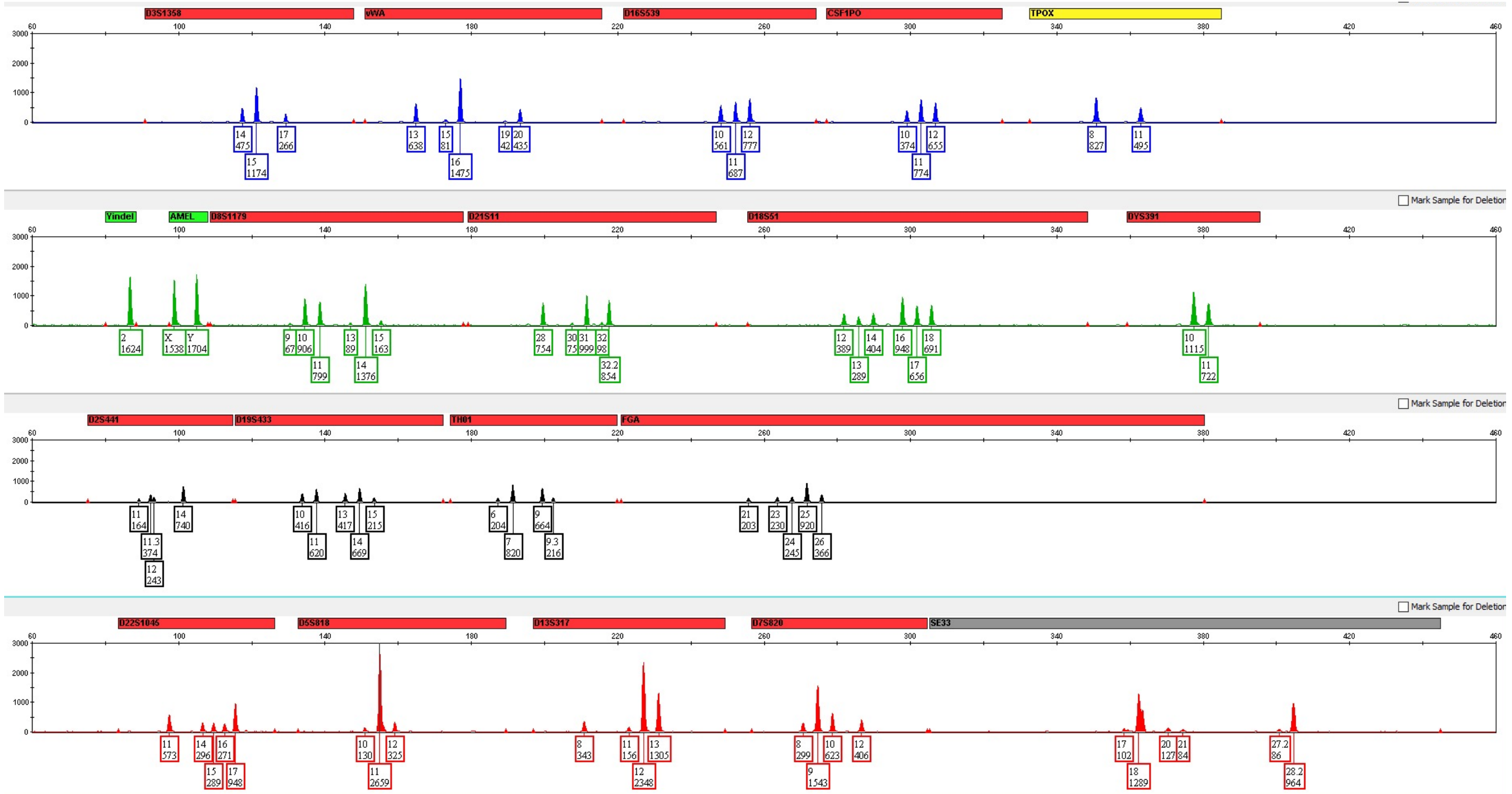
Two person mixture Sample 44 and 45 Ratio 1:1 pristine



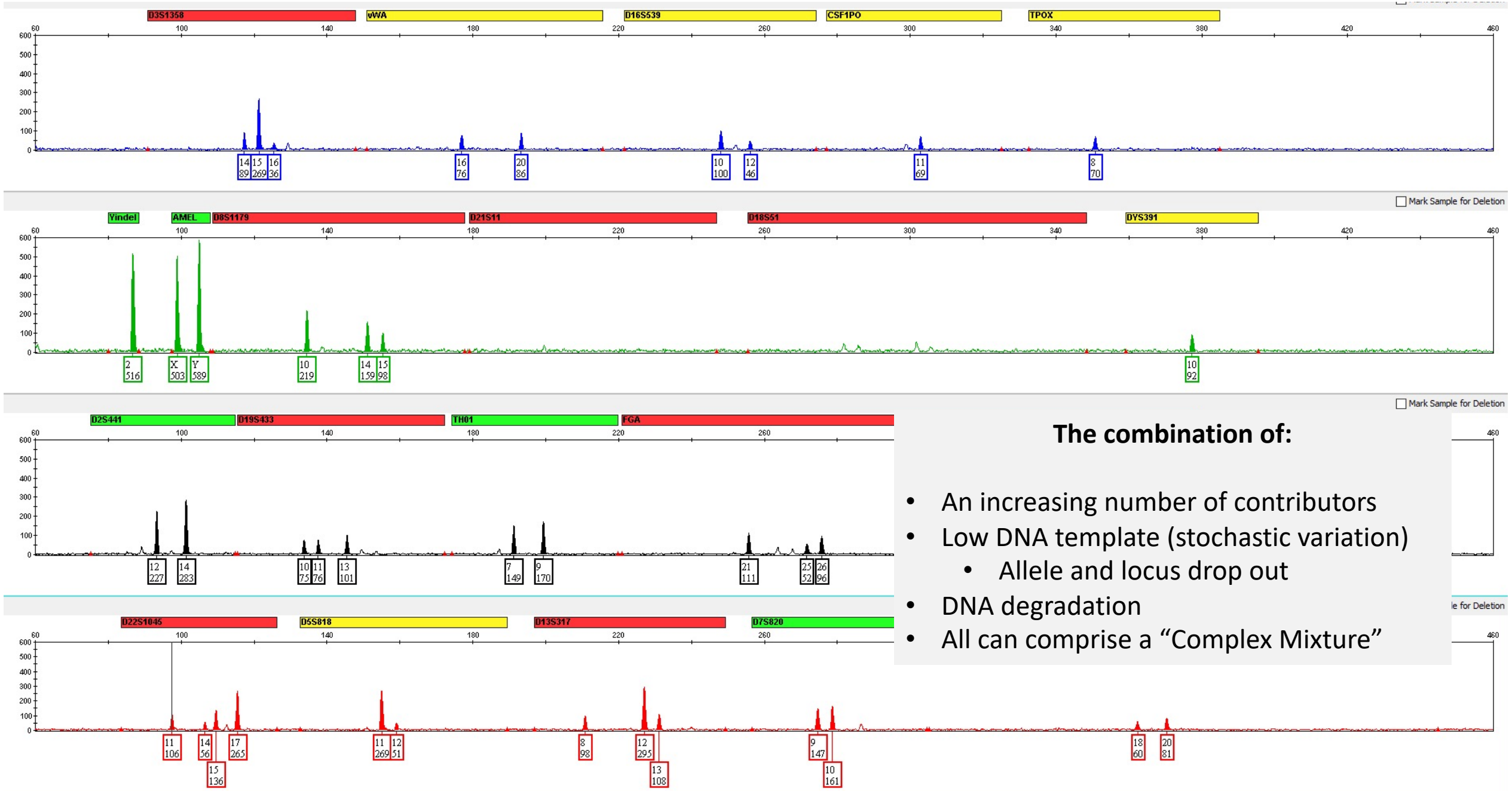
Observations/Questions

- > 1 or 2 alleles per locus
- Assess if it is pristine or low template profile (allele drop out)?
- Is a peak from stutter or a minor allele?
- Can the NoC be estimated?
- Can the mixture ratio be estimated?

Three person mixture Samples 44, 45, and 46 Ratio 1:2:2 pristine



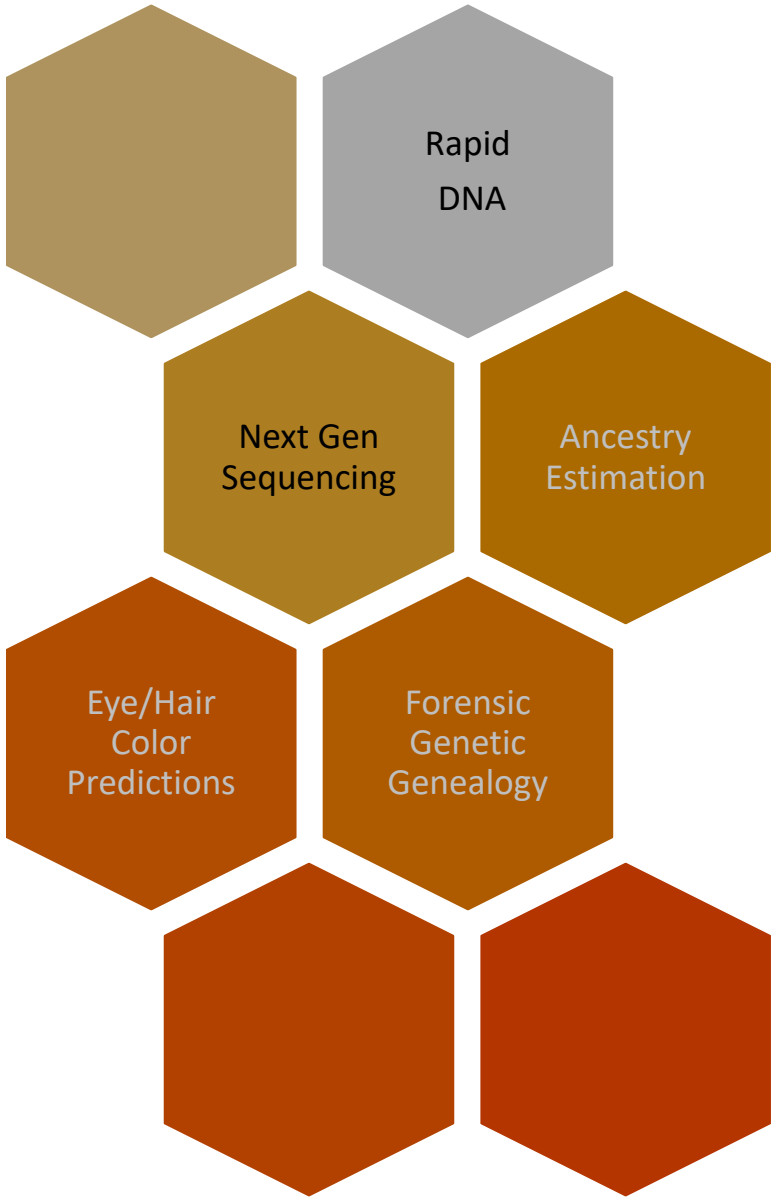
“Complex mixture” Three person mixture - degraded - low template



The combination of:

- An increasing number of contributors
- Low DNA template (stochastic variation)
  - Allele and locus drop out
- DNA degradation
- All can comprise a “Complex Mixture”

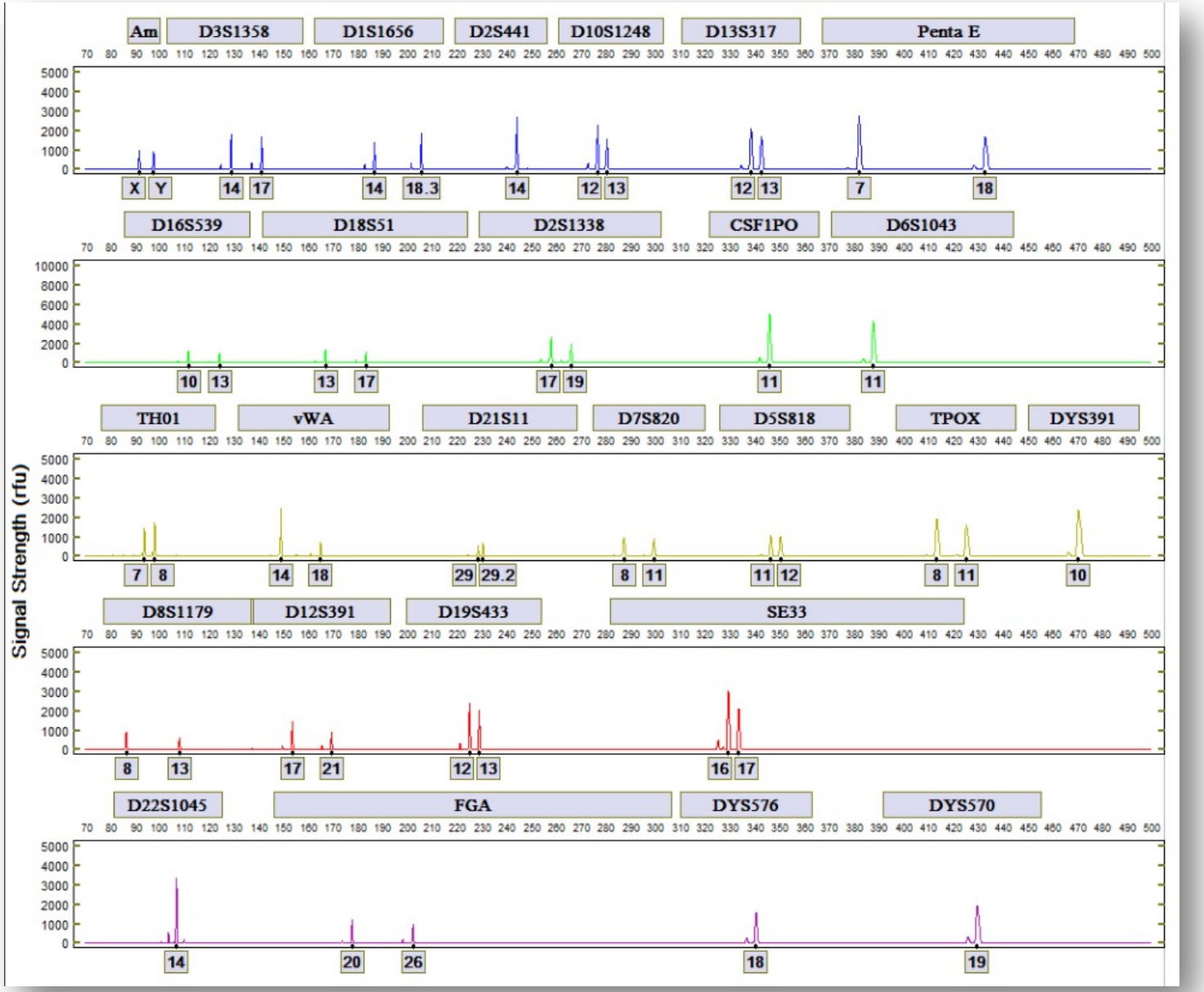
Recent advancements and applications

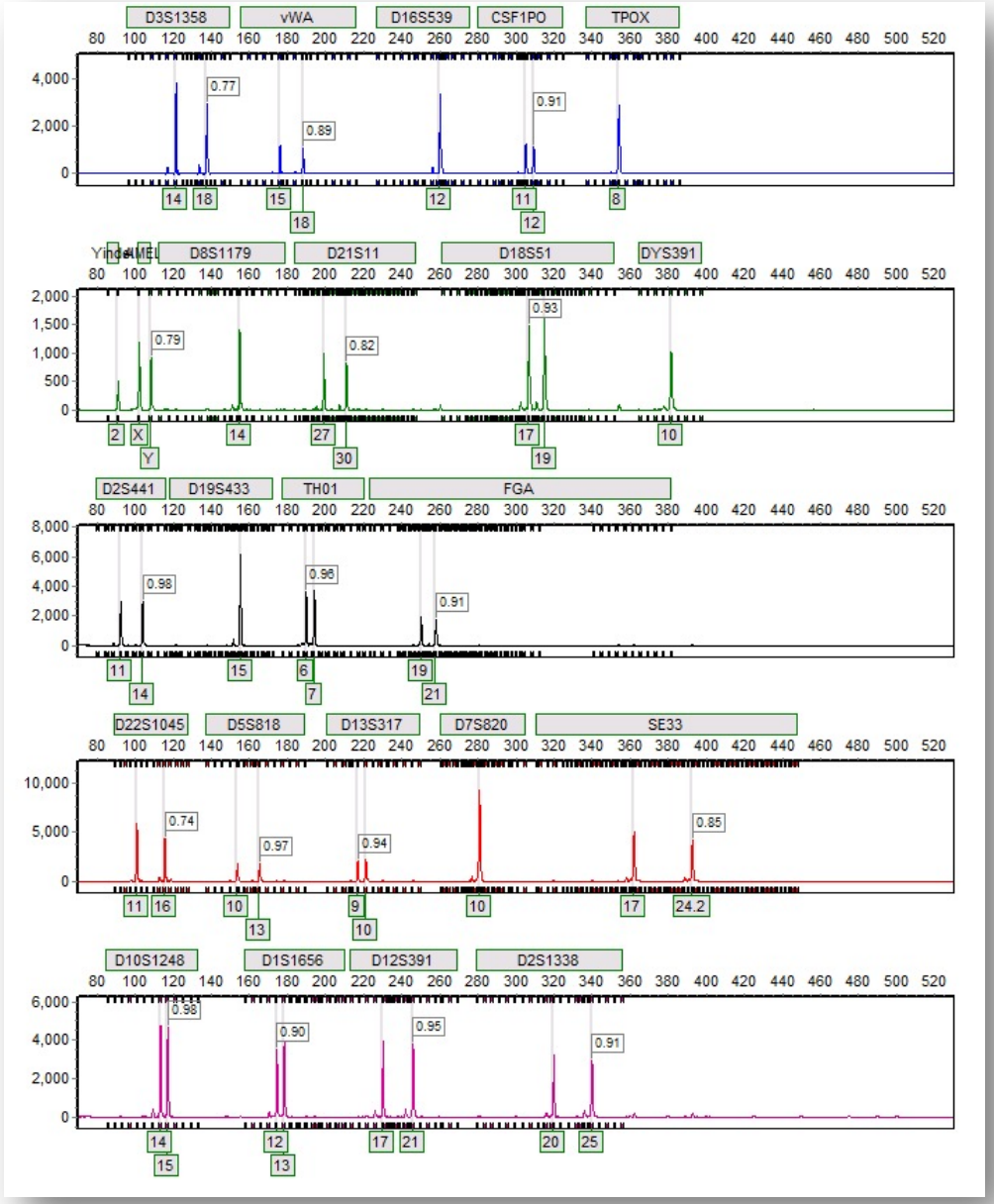


Rapid DNA Systems

- Perform DNA Typing in a closed, integrated system
- Enroll arrestees into the national database
  - Police booking station (buccal swab – an excess of single source DNA)
- Casework?
  - Considerations for field use – is the data quality amenable to automated interpretation?
- Field applications
- Mass disasters
- Kinship
- At a border crossing

Profiles from ***high quality single source*** samples generated by Rapid DNA instruments (generated in 90 - 100 minutes)





Rapid DNA Maturity Assessment 2018

JOURNAL OF FORENSIC SCIENCES

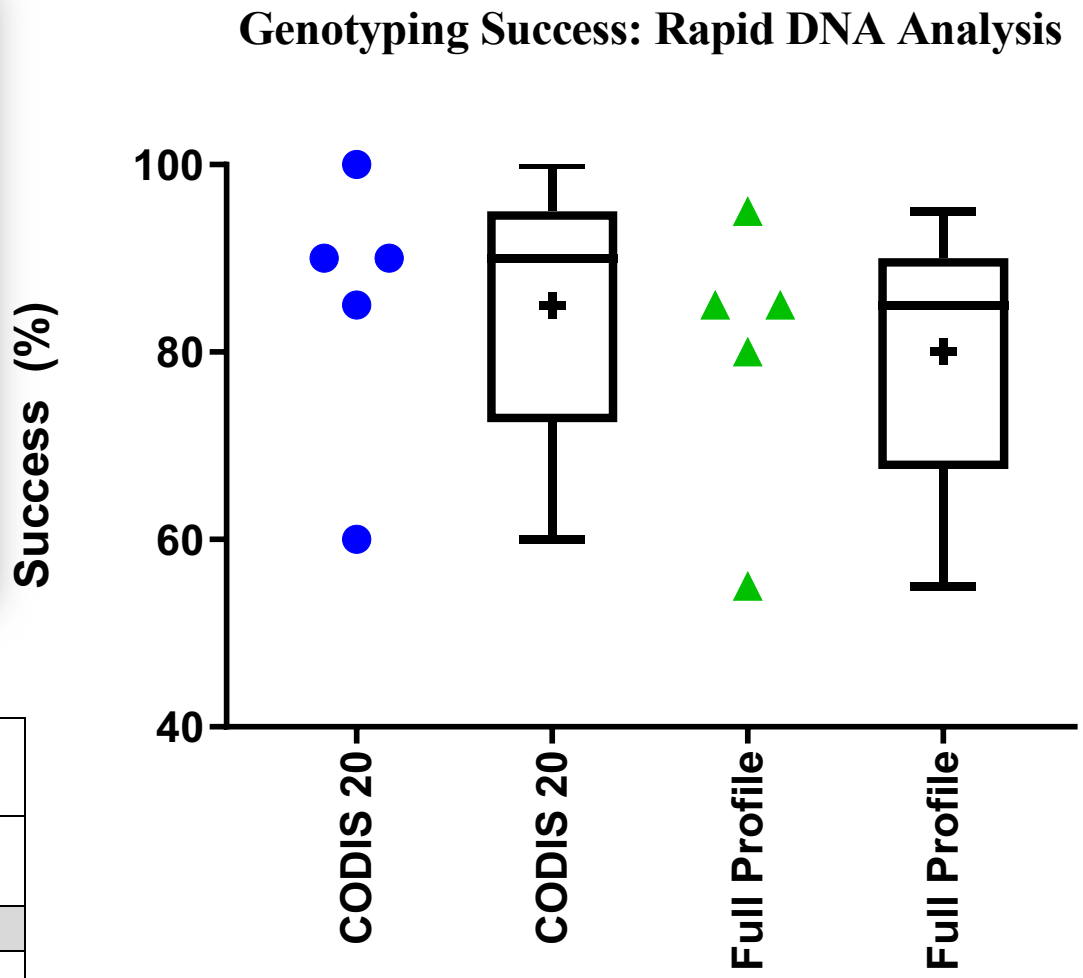
J Forensic Sci, May 2020, Vol. 6
doi: 10.1111/1556-4029.14267
Available online at: onlinelibrary.wiley.com

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.

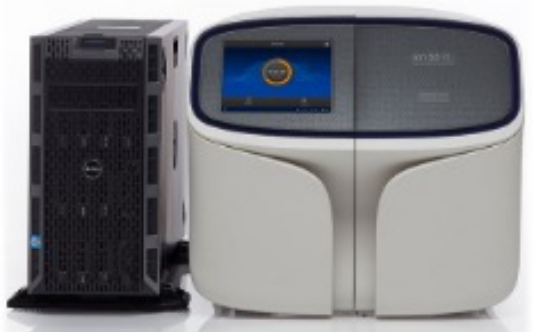
Results of the 2018 Rapid DNA Maturity Assessment*



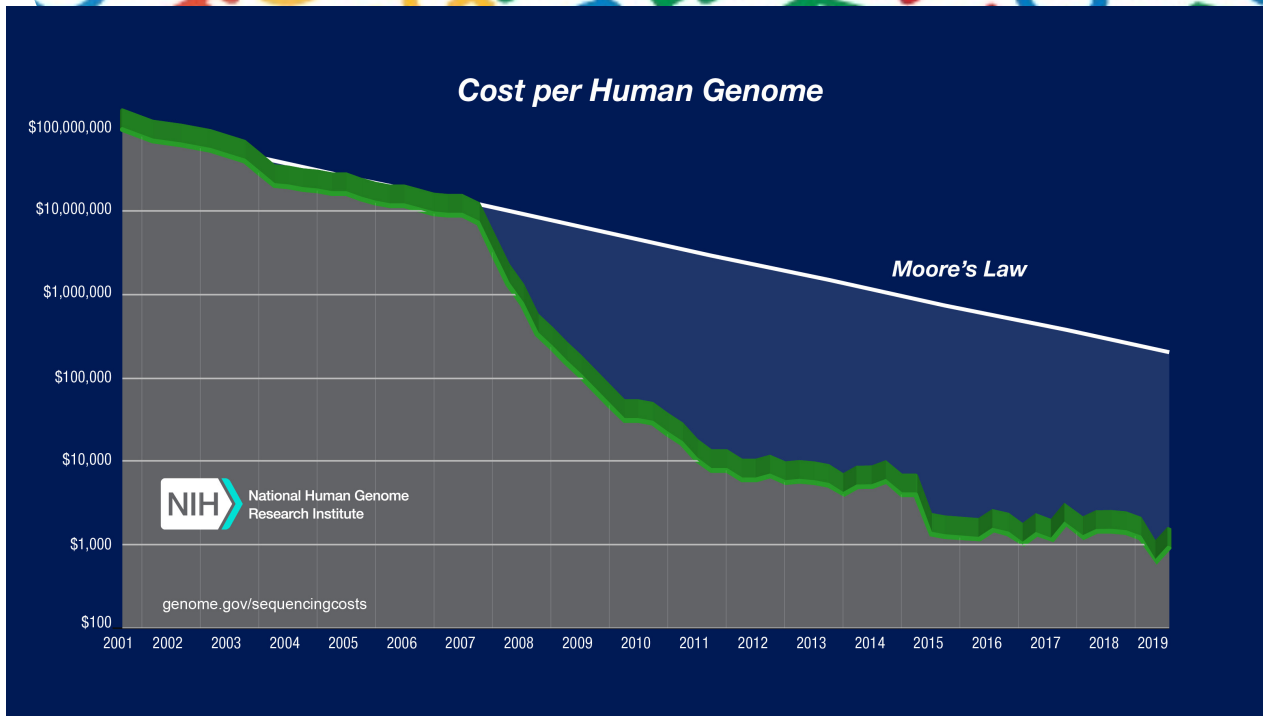
	Prior to Analysis Definitions	Rapid DNA Analysis		Modified Rapid DNA Analysis	
Year of Study	CODIS 13 Success (%)	CODIS 13 Success (%)	CODIS 20 Success (%)	CODIS 13 Success (%)	CODIS 20 Success (%)
2013	88.3				
2014		76.1	70.0	80.0	75.0
2018			85.0		90.0

Next Generation Sequencing (NGS)

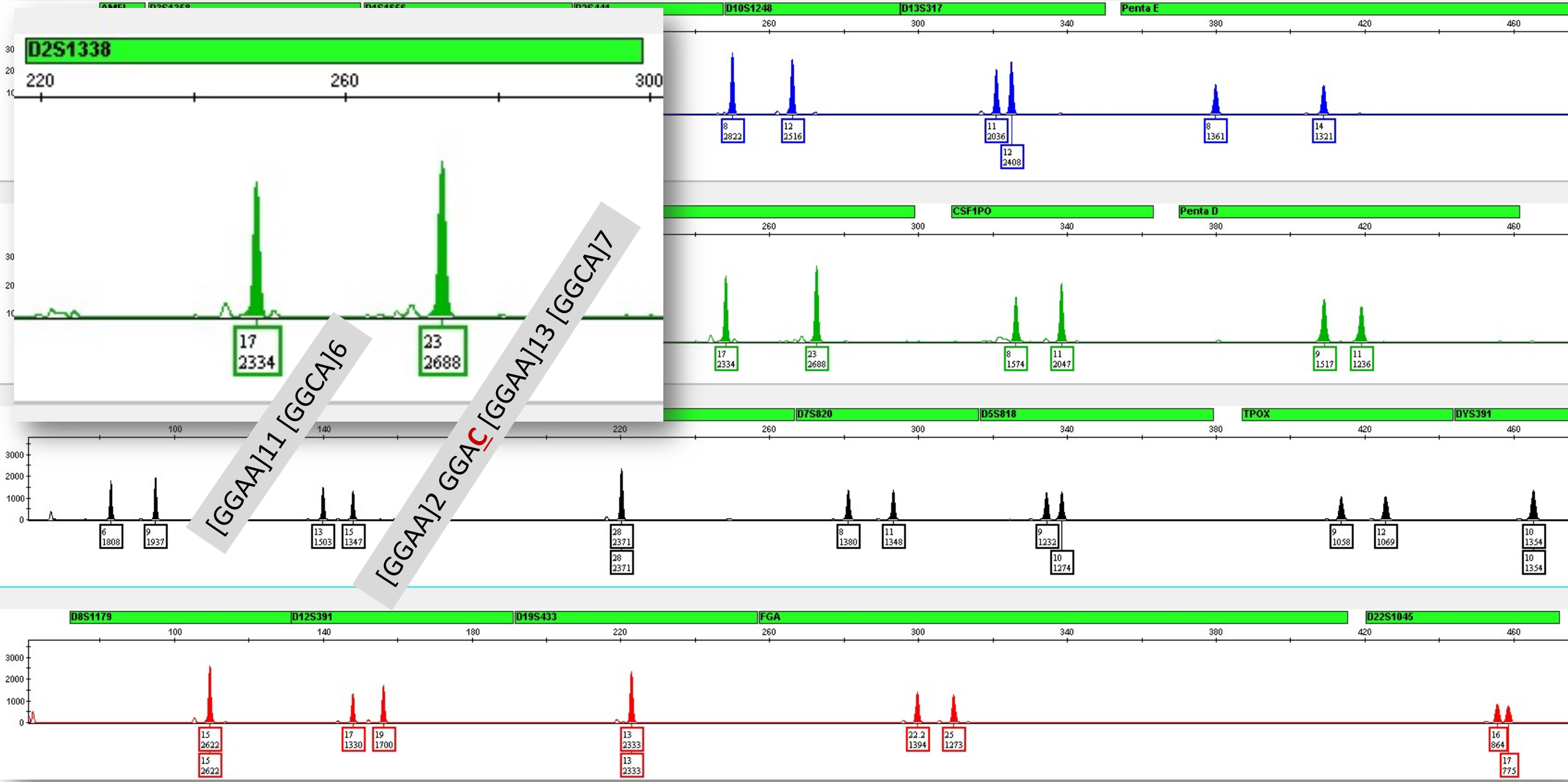
MiSeq FGx



Ion S5 XL

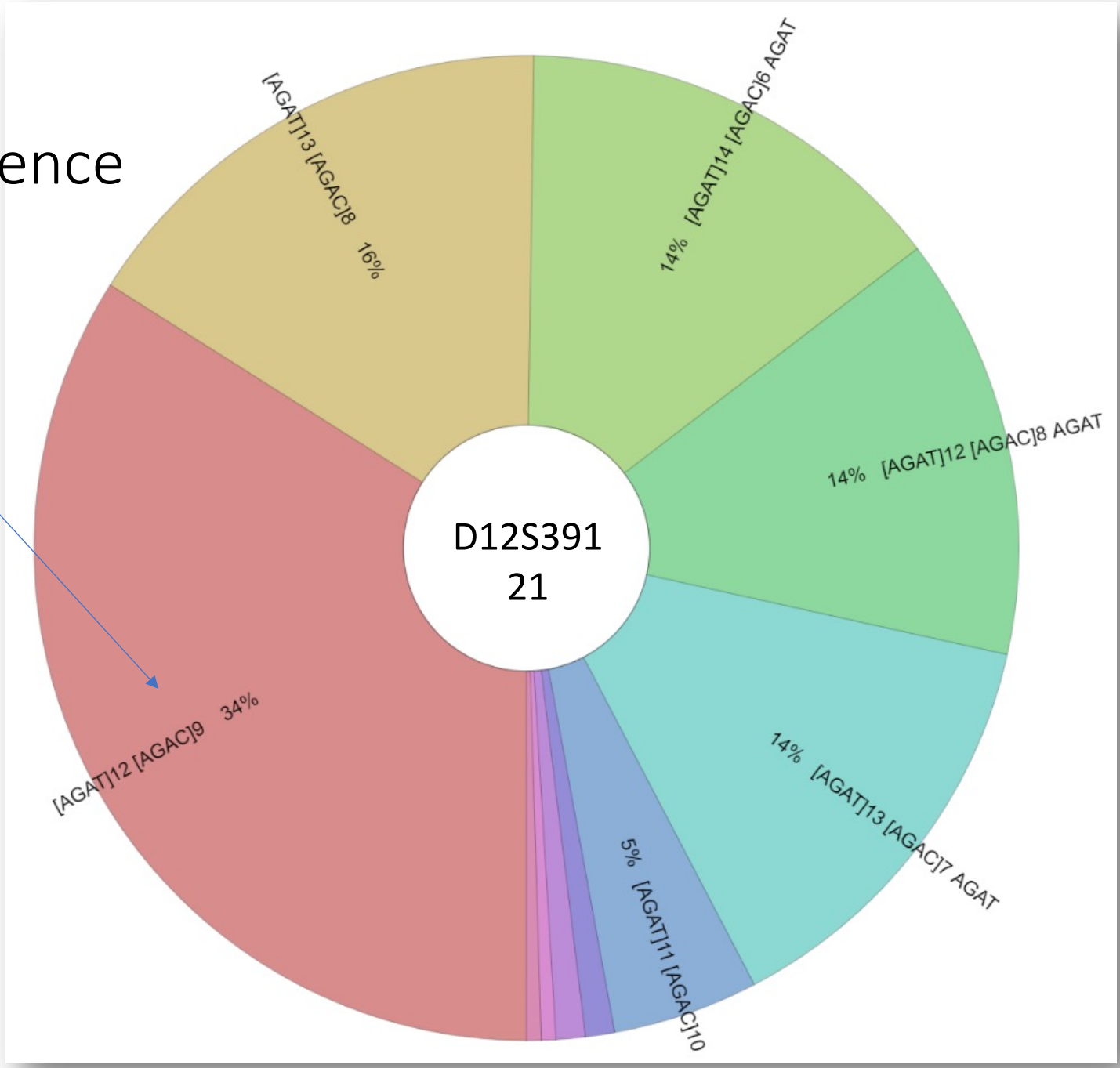


https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data


D2S1338
220
260
300
17
2334
23
2688
[GGAA]11 [GGCA]6
[GGAA]2 GGAC [GGAA]13 [GGCA]7
D10S1248
D13S317
Penta E
CSF1PO
Penta D
D7S820
D5S818
TPOX
DYS391
D8S1179
D12S391
D19S433
FGA
D22S1045

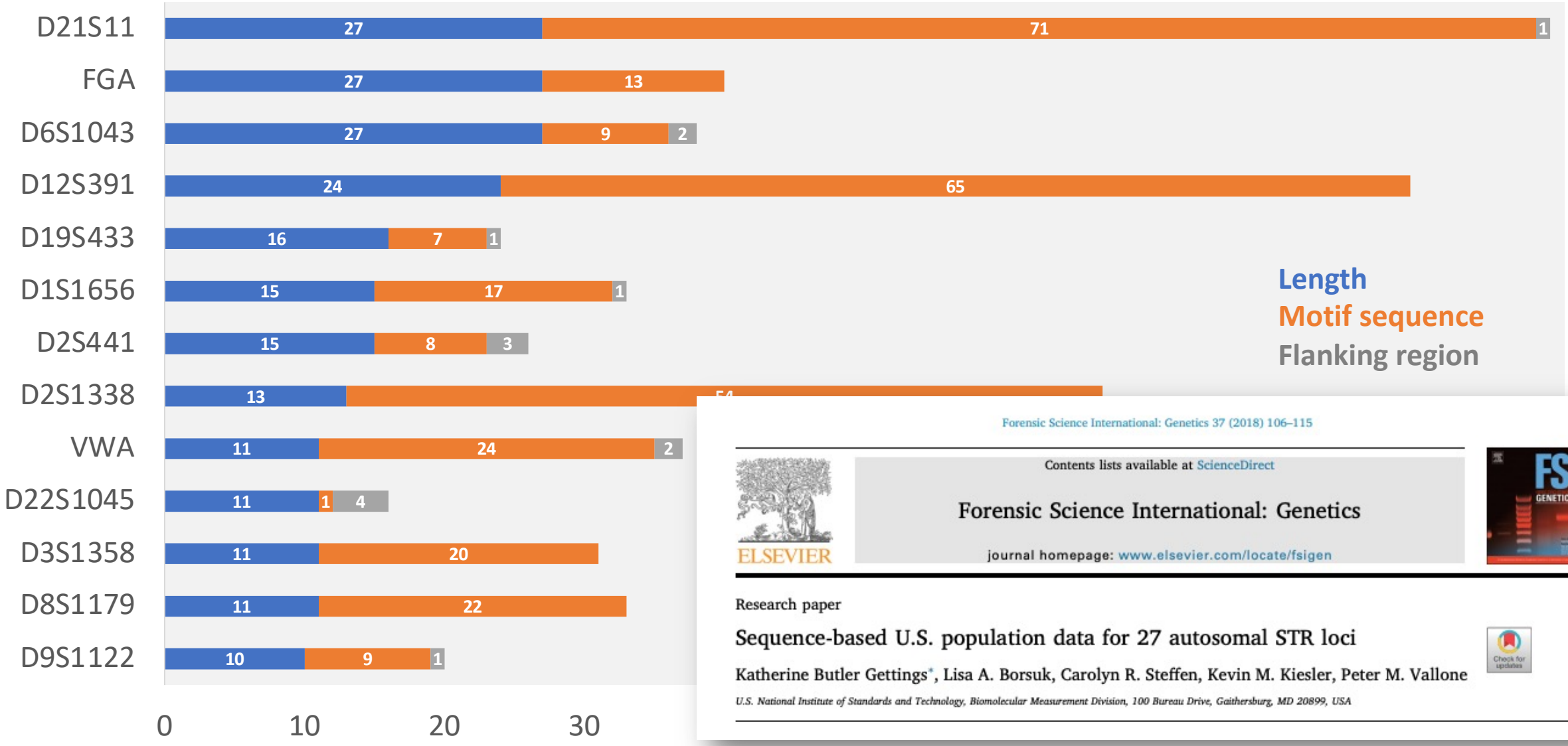
D12S391
Allele “21” frequencies by sequence

	Flavors of D12S391 - 21 allele	Global Freq
1	[AGAT]12 [AGAC]9	3.43%
2	[AGAT]13 [AGAC]8	1.64%
3	[AGAT]14 [AGAC]6 AGAT	1.25%
4	[AGAT]13 [AGAC]7 AGAT	1.40%
5	[AGAT]12 [AGAC]8 AGAT	1.40%
6	[AGAT]11 [AGAC]10	0.48%
7	[AGAT]14 [AGAC]7	0.10%
8	[AGAT]11 [AGAC]9 AGAT	0.10%
9	[AGAT]10 [AGAC]10 AGAT	0.05%
10	[AGAT]13 [AGAC]4 AGGC [AGAC]2 AGAT	0.05%
		10%



Compound/Complex autosomal STRs

Increase in observed alleles through 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

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

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

# of unique alleles

N=1036

Sequencing Forensic STRs in Population Samples

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

Example of length versus sequence-based frequency calculation:

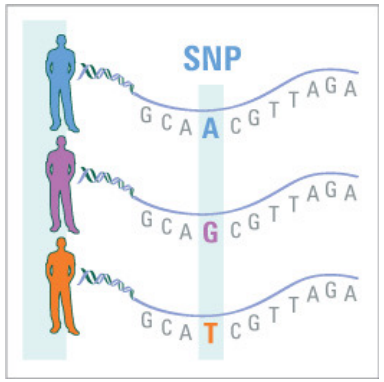
D4S2408

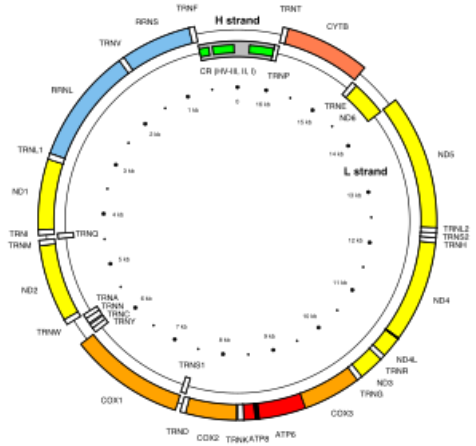
Allele	N	Freq	Sequence Allele	N	Freq
7	1	0.6%	[ATCT]7	1	0.6%
8	23	14.4%	[ATCT]8	23	14.4%
9	60	37.5%	[ATCT]9	18	11.3%
			[ATCT] **G**TCT [ATCT]7	42	26.3%
10	53	33.1%	[ATCT]10	53	33.1%
11	21	13.1%	[ATCT]11	21	13.1%
12	2	1.3%	[ATCT]12	2	1.3%

	Length	Sequence
	8,9	[ATCT]8, [ATCT]9
	2pq	2pq
	2*0.144*0.375	2*0.144*0.113
	0.108	0.033
	1 in 9.3	1 in 30.7

Sequencing also allows for the testing of emerging marker systems

- Mitochondrial genome sequence





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

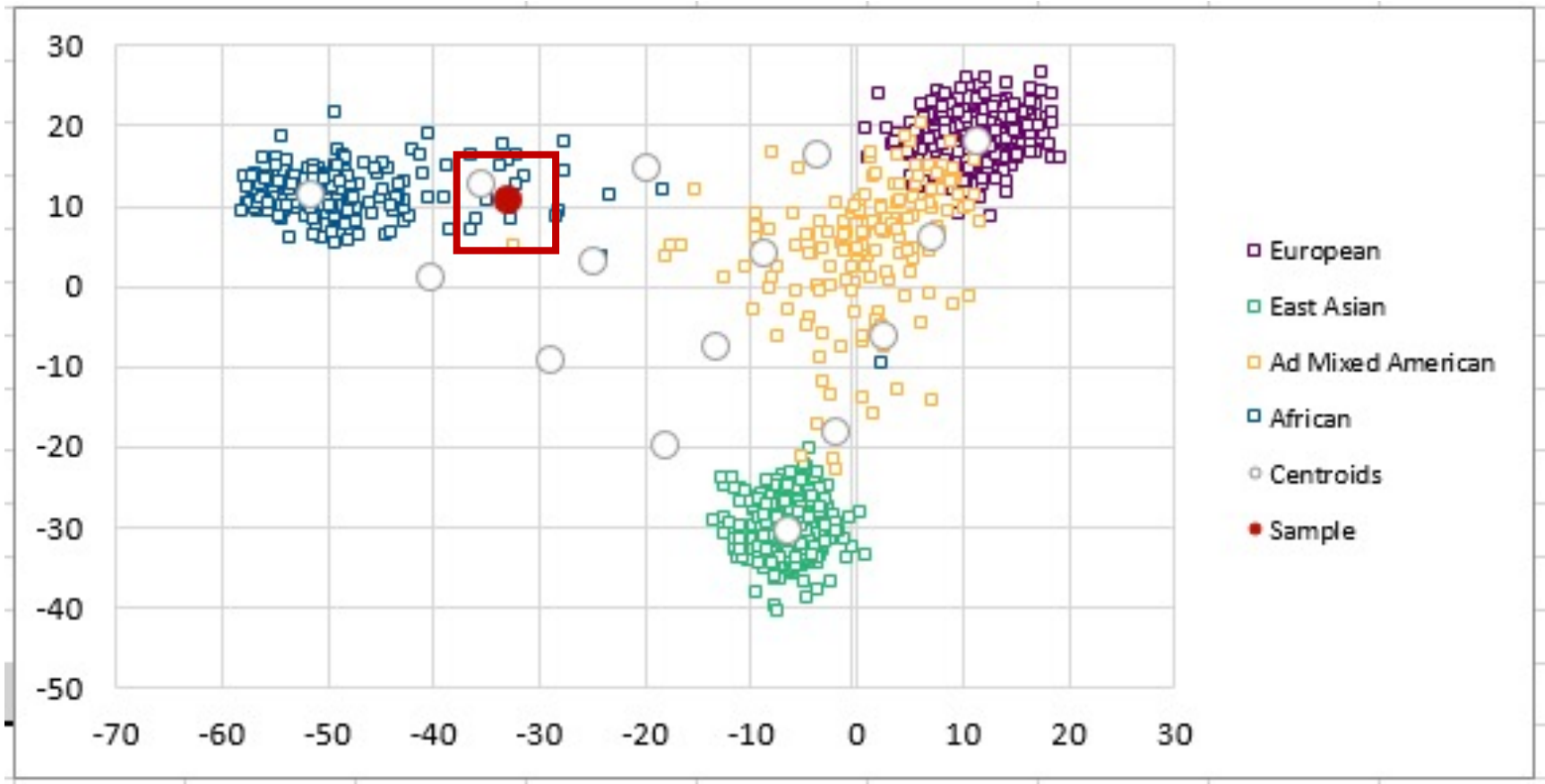
Data Collection for Sample Screening: SNPs

ForenSeq SNP Phenotype and Ancestry Estimation

Hair Color Results	
Brown	0.16
Red	0.00
Black	0.84
Blond	0.00

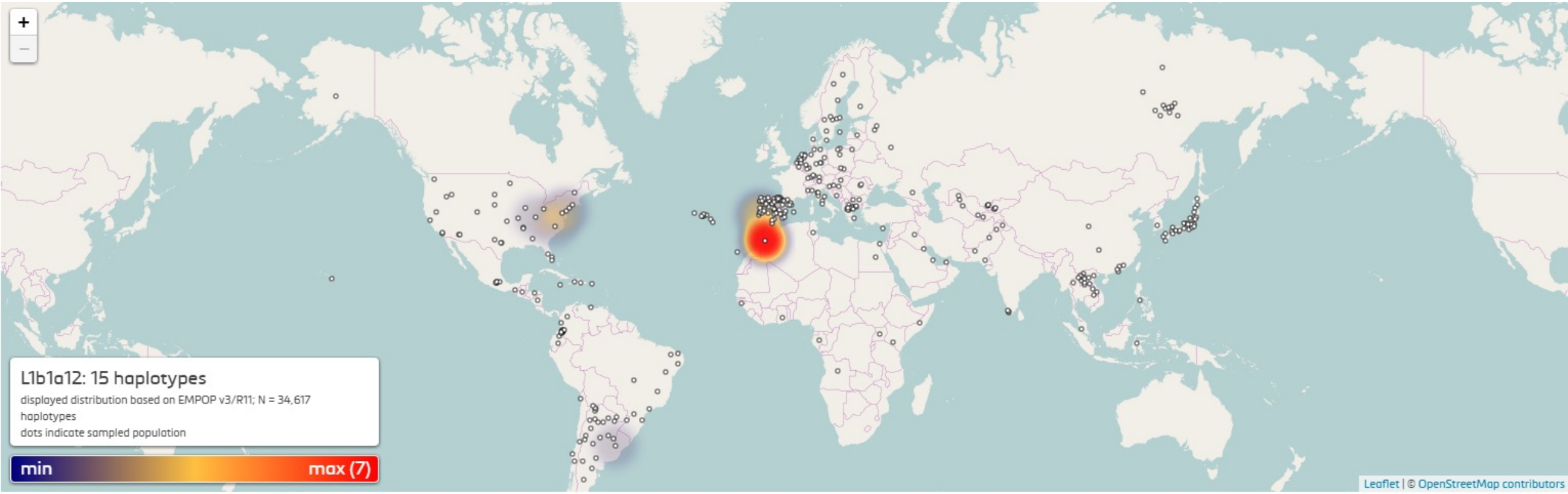
Eye Color Results	
Intermediate	0.00
Brown	1.00
Blue	0.00

Biogeographical Ancestry Results	
Distance to Nearest Centroid	3.36



Data Collection for Sample Screening: **mtDNA**

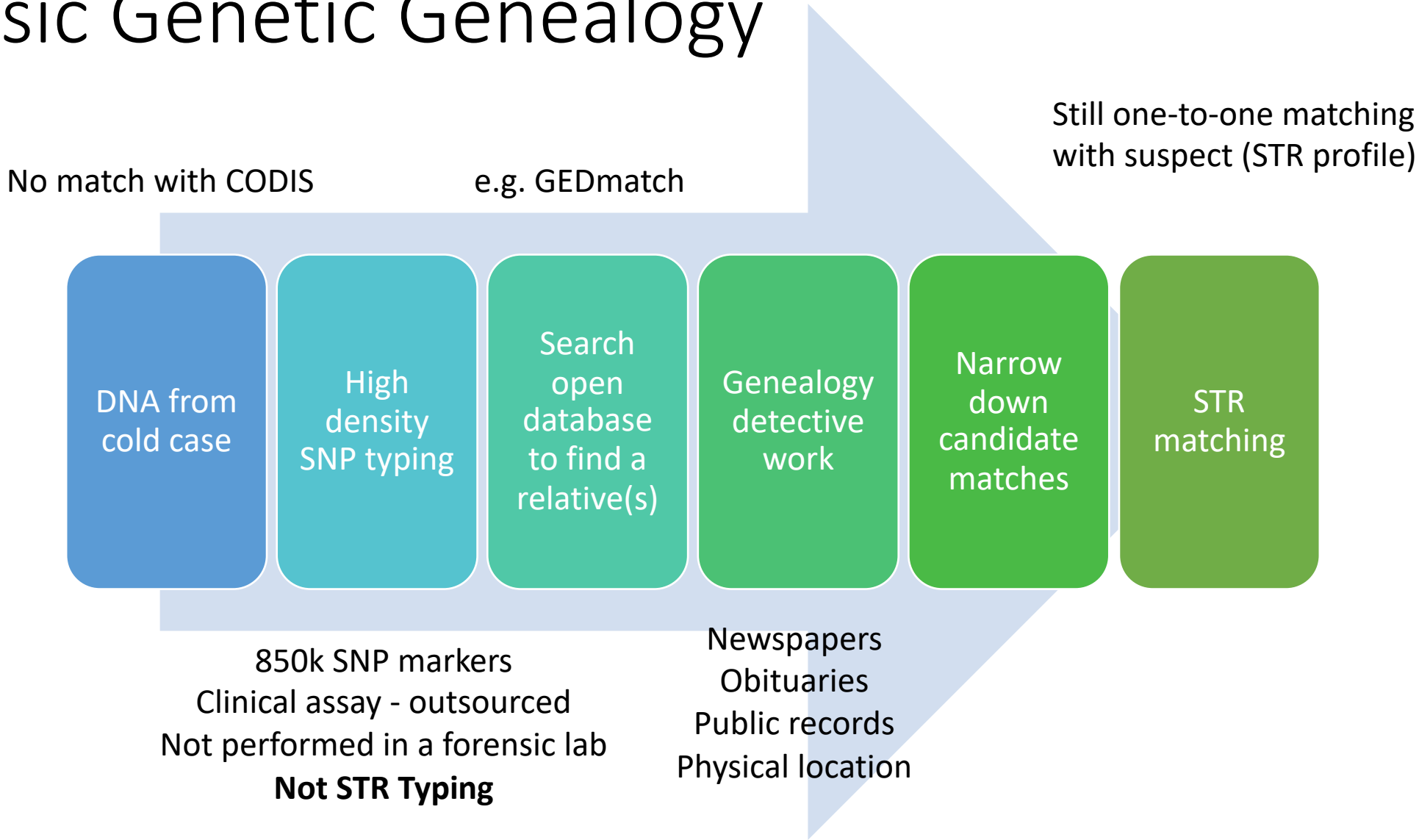
Illumina mtDNA Whole Genome Sequencing protocol with Nextera XT Sample Prep Kit



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

Haplogroup	Ancestry	Match
L1b1a12	African	unique

Forensic Genetic Genealogy



Forensic Genetic Genealogy

No match

DN
co

e matching
STR profile)

ng

Forensic Genetic Genealogy

Expand your network of information

FGG combines genetic and genealogy methods to identify people through relatives. The results provide investigative intelligence that exonerates the innocent, matches adoptees, and creates leads for cold cases. FGG is particularly useful when traditional methods are inconclusive, or all other options are exhausted.

A recent Verogen acquisition, the genealogical database GEDmatch aggregates DNA data files from known, voluntary contributors. Uploading a DNA data file yields a simple measurement of relatedness to help estimate kinship. This type of result makes GEDmatch a valuable tool that can create leads or eliminate suspects.

FGG and the GEDmatch database are gaining traction in real-world scenarios, providing meaningful investigative breakthroughs and resolving cases like the following.

- The charging of Joseph DeAngelo with 13 cases of murder and 13 cases of kidnapping. After decades of dead ends, GEDmatch assisted with the identification of DeAngelo as the alleged Golden State Killer.
- The conviction of William Earl Talbott II of a 1987 double homicide was the first conviction to apply FGG. Data from DNA collected at the crime scene over 30 years ago were uploaded to GEDmatch and ultimately led investigators to Talbot.
- The exoneration of Christopher Tapp for the 1996 homicide of Angie Dodge and the charging of Brian Leigh Dripps marked the first genealogy-based exoneration. A DNA sample from the crime scene was processed and compared in GEDmatch, narrowing the focus to Dripps. A DNA sample from Dripps then confirmed him as the source of the crime scene material.

Leveraging the MiSeq FGx System and working in concert with the forensic community, Verogen is developing FGG as an end-to-end, fully integrated solution. As the only portfolio to include this capability, Verogen is uniquely able to bring the next era of genealogy into your laboratory.

Verogen acquisition of GEDmatch

The Shared cM Project – Version 3.0
August 2017

Blaine T. Bettinger
www.TheGeneticGenealogist.com

For MUCH more information (including histograms and company breakdowns) see: goo.gl/Z1EcJQ

How to read this chart:

Aunt/Uncle
1750
1349 - 2175

Relationship
Average
Range (low-high)
(99% Percentile)

								Great-Great-Great-Grandparent		**GGGG-Aunt/Uncle**
							Great-Great-Grandparent		**GGG-Aunt/Uncle**	
Half GG-Aunt/Uncle 187 12 – 383	**Great-Grandparent** 881 464 – 1486							**Great-Great Aunt/Uncle** 427 191 – 885		
	Half Great-Aunt/Uncle 432 125 – 765	**Grandparent** 1766 1156 – 2311					**Great Aunt/Uncle** 914 251 – 2108			
		Half Aunt/Uncle 891 500 – 1446	**Parent** 3487 3330 – 3720			**Aunt/Uncle** 1750 1349 - 2175				
Half 3c 61 0 – 178	**Half 2c** 117 9 – 397	**Half 1C** 457 137 – 856	**Half-Sibling** 1783 1317 – 2312	**Sibling** 2629 2209 – 3384	**SELF**	**1C** 874 553 – 1225	**2c** 233 46 – 515	**3c** 74 0 – 217	**4c** 35 0 – 127	**5c** 25 0 – 94
Half 3c1R 42 0 – 165	**Half 2c1R** 73 0 – 341	**Half 1C1R** 226 57 – 530	**Half Niece/Nephew** 891 500 – 1446	**Niece/Nephew** 1750 1349 - 2175	**Child** 3487 3330 – 3720	**1C1R** 439 141 – 851	**2c1R** 123 0 – 316	**3C1R** 48 0 – 173	**4C1R** 28 0 – 117	**5C1R** 21 0 – 79
Half 3c2R 34 0 – 96	**Half 2c2R** 61 0 – 353	**Half 1C2R** 145 37 – 360	**Half Great Niece/Nephew** 432 125 – 765	**Great-Niece/Nephew** 910 251 – 2108	**Grandchild** 1766 1156 – 2311	**1C2R** 229 43 – 531	**2c2R** 74 0– 261	**3C2R** 35 0 – 116	**4C2R** 22 0 – 109	**5C2R** 17 0 – 43
Half 3c3R	**Half 2c3R**	**Half 1C3R** 87 0 – 191	**Half GG Niece/Nephew** 187 12 – 383	**Great-Great-Niece/Nephew** 427 191 – 885	**Great-Grandchild** 881 464 – 1486	**1C3R** 123 0 – 283	**2c3R** 57 0 – 139	**3C3R** 22 0 – 69	**4C3R** 29 0 – 82	**5C3R** 11 0 – 44

Other Relationships

6C 21 0 – 86	
6C1R 16 0 – 72	
6C2R 17 0 – 75	
7C 13 0 – 57	
7C1R 13 0 – 53	
8C 12 0 – 50	

Minimum was automatically set to 0 cM for relationships more distant than Half 2C, and averages were determined only for submissions in which DNA was shared

https://thegeneticgenealogist.com/2017/08/26/august-2017-update-to-the-shared-cm-project/

Involvement in community working groups

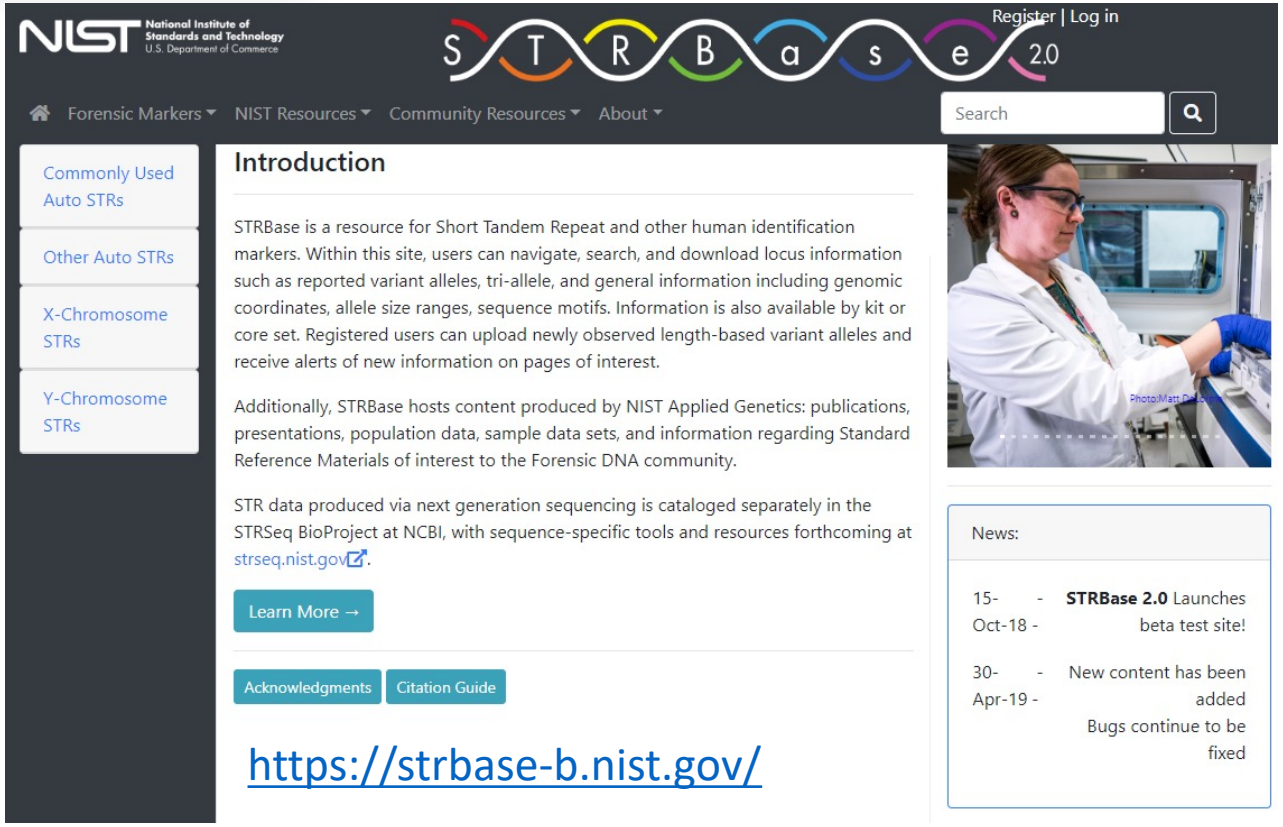
- SWGDAM – FBI Scientific working group for DNA analysis methods
  - Rapid DNA, Validation of STR typing methods, Next Generation Sequencing, Body Fluid Identification
  - Rapid DNA Task Force Groups
- NIST OSAC – Organization of Scientific Area Committees
  - Validation of STR typing methods, Sequencing, Profile interpretation
- NIJ – National Institute of Justice
  - FLN-TWG – Forensic laboratory needs technical working group
  - Grant peer review
  - Technical working group (TWG) for Research and Development

https://strbase.nist.gov/index.htm

Serving the human identity community since 1997

- First round of development
  - STR fact sheets (for 24 loci)
  - Variant allele reporting
- Provide search, sort, and download functionalities
- Automated submission of variant alleles
- Embedded viewer for STR sequence and presentations

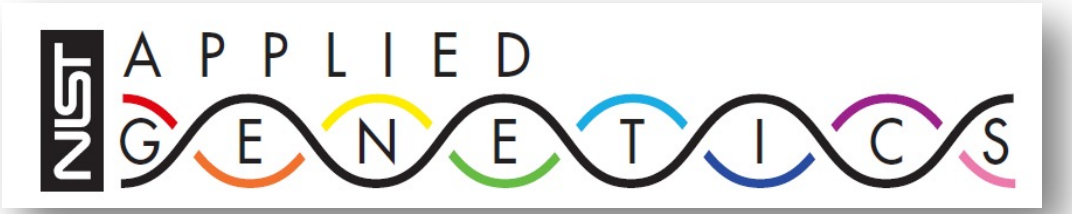
STRBase 2.0



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

Thank you for your attention! Questions?

Contact: Peter.Vallone@nist.gov



- Funding
  - NIST Special Programs Office: *Forensic DNA*
  - FBI Biometrics Center of Excellence: *Forensic DNA Typing as a Biometric tool.*
  - NIJ: *STRSeq and Nomenclature*

- **Disclaimer** - Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

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