STRs in Forensic DNA Analysis

This presentation is made possible through the following financial assistance award #70NANB23H276 awarded to Towson University from the U.S. Department of Commerce, National Institute of Standards and Technology (NIST)

Slides prepared by Kelly M. Elkins, Ph.D., Towson University, 2024

Background

In October 2023, Towson University was awarded a cooperative agreement from NIST to develop a standardized DNA training curriculum for the United States
that address the components in ANSI/ASB Standard 115, *Standards for Training in Forensic Short Tandem Repeat Typing Methods Using Amplification, DNA Separation, and Allele Detection*. 2020. 1st Ed.

This presentation addresses the knowledge-based portion of the training program and covers the topic outlined in 4.2.3a in ANSI/ASB Standard 115.

Learning Objectives

1. history of development and use;
2. structure and nomenclature;
3. methods of analysis;
4. STR typing systems (e.g., commercially produced kits);
5. core STR loci (e.g., CODIS);
6. limitations of the technology

Terms and Definitions (ANSI/ASB 115)

- **Allele** – One of two or more versions of a genetic sequence at a particular location in the genome.
- **Locus** (plural **loci**) – A unique physical location of a gene (or specific sequence of DNA) on a chromosome.
- **Capillary electrophoresis** – An electrophoretic technique for separating DNA molecules by their relative size based on migration through a narrow glass capillary tube filled with a liquid polymer.

History

- Short tandem repeats (STRs) are microsatellite DNA markers discovered in the 1970s
  - Forensic STRs were characterized by Thomas Caskey in Texas and the UK Forensic Science Service (FSS) in the early 1990s
  - Termed *satellite* because they were found to surround the chromosome centromere in early experiments
  - Common throughout the human genome and comprise ~3% of the genome
  - 2-6 bp repeats

History and Development of Use

- Short tandem repeats are used for forensic DNA typing
  - Highly variable
  - Follow Mendelian inheritance
  - Short repeats and overall locus length (80-450 bp) makes them suitable for PCR
  - Tetranucleotide repeats are easily interpreted by discrete size differences using electrophoresis
  - Large numbers of repeats may contain several hundred of the core repeats
  - Publicly available sequences deposited in GenBank

STR Structure and Nomenclature

Originally termed "junk DNA"

Non-coding elements between genes or expressed units

"Words" of nitrogenous bases are repeated one after another like boxcars on a train

Sequence and nomenclature defined by top coding (sense) strand unless historically defined on bottom strand in literature

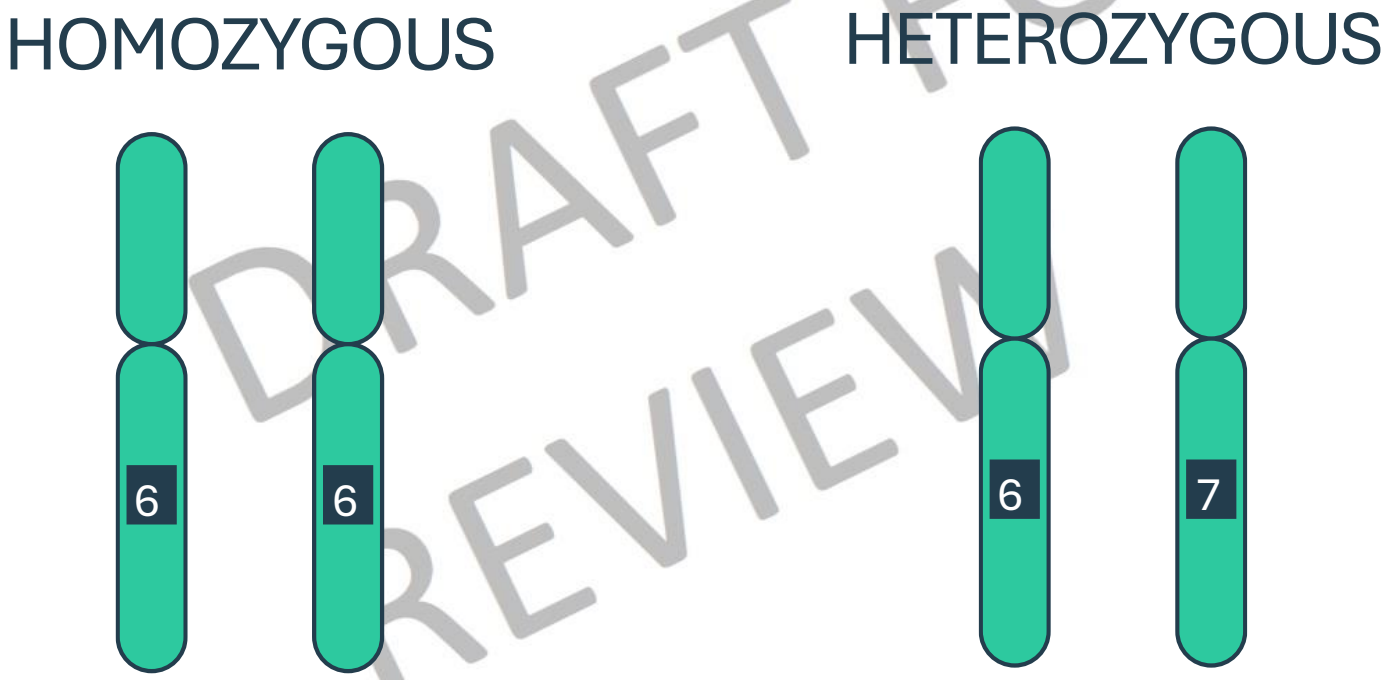
Allele reported as number of repeats

STRs selected on autosomes (1-22) and X and Y chromosomes

D16S539

NCBI GenBank Accession G07925 for D16S539 has 11 repeats

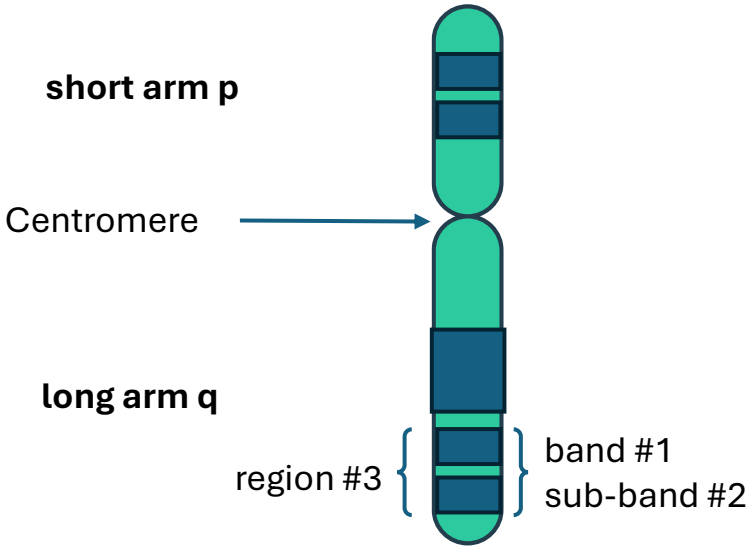
Structure and Nomenclature of STR Repeats for an Autosomal Locus



Structure and Nomenclature

- Nomenclature
  - Gene name, if previously named due to its adjacency to a gene or protein product
    - **TH01** (HUMTH01): located on chromosome 11 at p15.5, **hum**an **t**yrosine **h**ydroxylase gene intron **1**
    - **TPOX**: located on chromosome 2 at p25.3, human **t**hyroid **p**er**ox**idase gene intron 10
    - **vWA (VWF)**: located on chromosome 12 at p13.31; **v**on **W**illebrand Factor 40th intron
  - Number indicating site on chromosome, *s* indicating single copy
    - D5S818: located on chromosome 5 at q23.2
    - D7S820: located on chromosome 7 at q21.11
    - D18S51: located on chromosome 18 in the intron between exons 2 and 3 of the B cell lymphoma 2 (BCL2) gene at 18q21.33

Example: Chromosome 5
Chromosome location 5q31.2



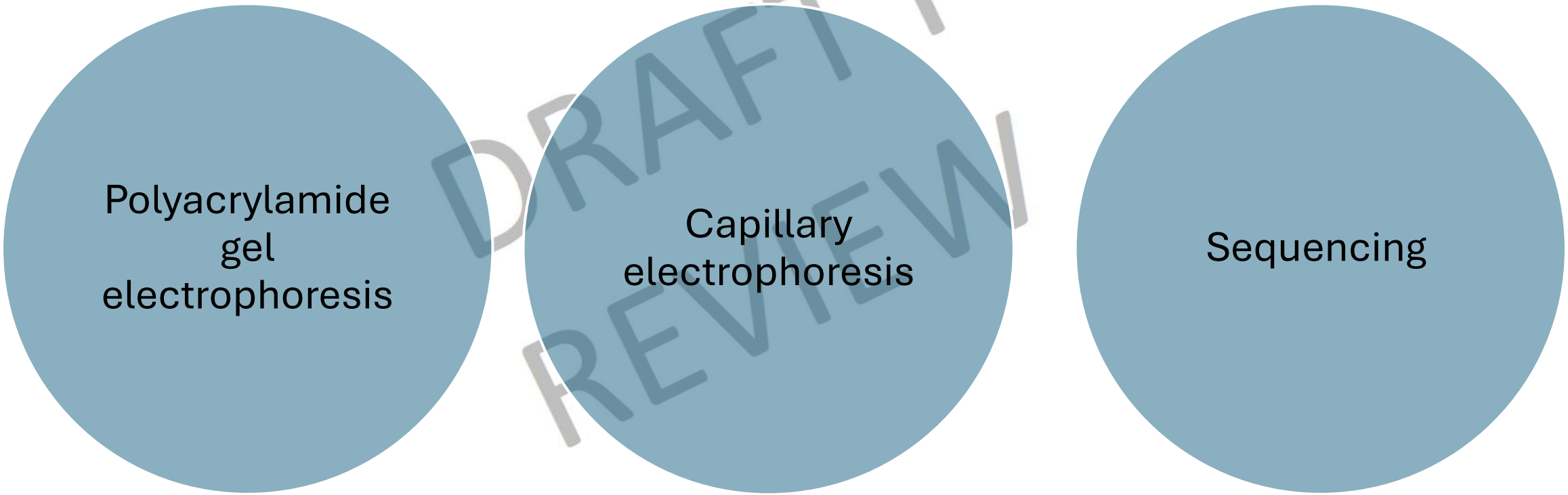
STR Structure and Nomenclature

- Simple sequence repeat (SSR) elements of identical length (true repeats) (TPOX, CSF1PO, D5S818, D13S317, D16S539)
  - TPOX –AATG–AATG–AATG–AATG–AATG–AATG–AATG–AATG–AATG–AATG–AATG– (GenBank M68651, 11 repeats)
- Simple repeats with non-consensus alleles (TH01, D18S51, D7S820)
  - Incomplete repeats or microvariants
    - TH01 9.3 –AATG–AATG–AATG–AATG–AATG–AATG–ATG–AATG–AATG–AATG–
- Compound repeat elements containing two or more SSRs in a string (vWA, FGA, D3S1358, D8S1179)
  - vWA $\mathrm{TCTA\ [TCTG]_4\ [TCTA]_{13}}$
- Complex repeat elements of variable length and sequence (D21S11)
  - D21S11 $\mathrm{[TCTA]_4\ [TCTG]_6\ [TCTA]_3\ TA\ [TCTA]_3\ TCA\ [TCTA]_2\ TCCATA\ [TCTA]_{11}}$

STR Structure and Nomenclature

Locus	Chromosome	GenBank Accession ID	Repeat Sequence in NCBI
D1S1656	1q42	NC_000001.9	$[TAGA]_{16}$ $[TGA][TAGA][TAGG]_1[TG]_5$
D2S441	2p14	AL079112	$[TCTA]_{12}$
TPOX	2p25.3	M68651	$[AATG]_{11}$
D2S1338	2q35	G08202	$[TGCC]_6$ $[TTCC]_{11}$
D3S1358	3p25.3	11449919	TCTA $[TCTG]_2$ $[TCTA]_{15}$
FGA (FIBRA)	4q28.2	M64982	$[TTTC]_3$ TTTTTTCT $[CTTT]_{13}$ CTCC $[TTCC]_2$
D5S818	5q23.2	G08446	$[AGAT]_{11}$
CSF1PO	5q33.1	X14720	$[AGAT]_{12}$
D7S820	7q21.11	G08616	$[GATA]_{12}$
D8S1179	8q23.1-23.2	G08710	$[TCTA]_{12}$
D10S1248	10q26.3	AL391869	$[GGAA]_{13}$
TH01	11p15.5	D000269	$[TCAT]_9$
D12S391	12p13.2	G08921	$[AGAT]_5$ GAT $[AGAT]_7$ $[AGAC]_6$ AGAT
vWA	12p13.31	M25858	TCTA $[TCTG]_4$ $[TCTA]_{13}$
D13S317	13q31.1	G09017	$[TATC]_{13}$
D16S539	16q24.1	G07925	$[GATA]_{11}$
D18S51	18q21.33	L18333	$[AGAA]_{13}$
D19S433	19q12	G08036	AAGG [AAAG] AAGG TAGG $[AAGG]_{11}$
D21S11	21q21.1	AP000433	$[TCTA]_4$ $[TCTG]_6$ $[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCATA $[TCTA]_{11}$
D22S1045	22q12.3	AL022314	$[ATT]_{14}$ ACT $[ATT]_2$

Methods of Analysis



STR Typing Systems: Commercially Produced Kits

- STRs selected for
  - Separate or distant chromosome locations to avoid linkage
  - High discriminating power (generally >0.9)
  - High heterozygosity (>70%)
  - Low mutation rate
  - Robust results when used by various labs
  - Reproducible results when markers multiplexed

- Tetranucleotide repeats selected because
  - Lower stutter (30% or more with di- and tri-nucleotide repeats)
  - Narrow size range of STR length
  - Can be copied by PCR
  - Reduced incidence of preferential amplification of shorter alleles
  - Reduced loss of amplification of degraded DNA
  - Easier to interpret than dinucleotide repeats

STR Typing Systems: Silver Stain Detection

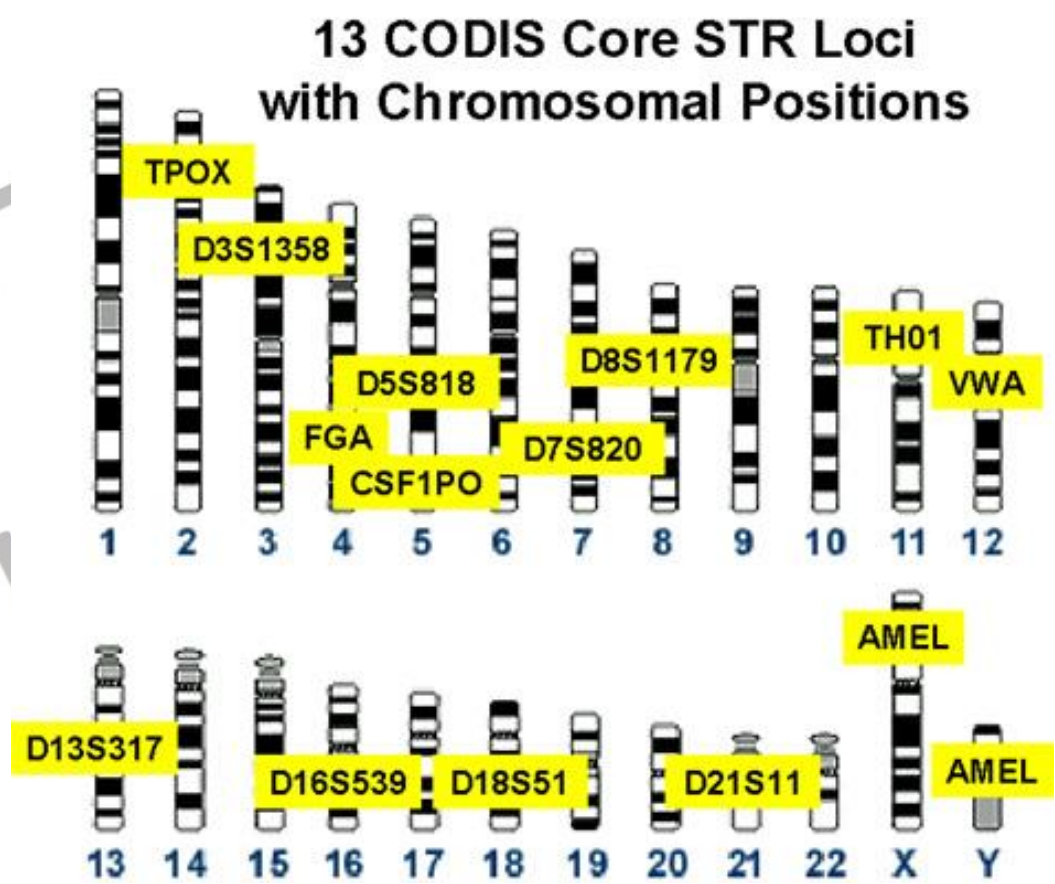
- Commercial kits were first offered in 1993 by Promega Corporation
- The earliest kits detected DNA by silver stain
- Targeted 1-4 loci
- FSS in house Quadruplex kit in 1994 targeted vWA, THO1, FES, and F13A1

Year	Kit Name	Manufacturer	Loci
1993	TH01	Promega	1
1994	CTT (CSF1PO, TPOX, TH01)	Promega	3
	CTTv	Promega	4
1994	FFv	Promega	3
	FFFL	Promega	4
	TH01, FES/FPS, vWA, F13A1	FSS	4
	SilverSTR™ III Multiplex	Promega	3
1996	GammaSTR™ Multiplex	Promega	4

Core STR Loci

Combined DNA Index System (CODIS)
U.S. database original **13 core** loci
October 1998 – December 31, 2016
- TPOX (chr. 2)
- D3S1358
- FGA (chr. 4)
- D5S818
- CSF1PO (chr. 5)
- D7S820
- D8S1179
- TH01 (chr. 11)
- VWA (chr. 12)
- D13S317
- D16S539
- D18S51
- D21S11

AMEL: amelogenin sex marker



Butler, J.M. 2005. Forensic DNA Typing, Elsevier.
https://strbase-archive.nist.gov/fbicore.htm

STR DNA Typing Systems: Fluorescent Dye Detection of Core STR Loci and Size Standard

Kits enabled with fluorescent dye detection were offered by Promega Corporation and Applied Biosystems beginning in 1996

Year	Kit Name	Manufacturer	Loci	Dyes
1996	AmpF*l*STR® SGM	Applied Biosystems	7	3
1996	PowerPlex™ System	Promega	8	3
1996	AmpF*l*STR® Blue	Applied Biosystems	3	2
1997	AmpF*l*STR® Green I	Applied Biosystems	4	2
1997	AmpF*l*STR® Profiler	Applied Biosystems	10	4
1997	AmpF*l*STR® Profiler Plus®	Applied Biosystems	9	4
1997	PowerPlex™ 1.1	Promega	13	3
1998	PowerPlex™ 1.2	Promega	8	3
1998	AmpF*l*STR® COfiler®	Applied Biosystems	7	4
1999	PowerPlex™ 2.1	Promega	9	3
1999	AmpF*l*STR® SGM Plus®	Applied Biosystems	11	4

STR DNA Typing Systems: 2000-2009

Year	Kit Name	Manufacturer	Loci	Dyes
2000	PowerPlex™ 16	Promega	16	4
2001	AmpF*l*STR™ Identifiler™	Applied Biosystems	16	5
2001	AmpF*l*STR™ Profiler Plus® ID	Applied Biosystems	10	4
2002	AmpF*l*STR™ SEfiler	Applied Biosystems	12	5
2002	PowerPlex™ ES	Promega	9	4
2007	AmpF*l*STR™ MiniFiler	Applied Biosystems	9	5
2007	AmpF*l*STR™ SEfiler Plus®	Applied Biosystems	12	5
2007	PowerPlex™ S5	Promega	5	3
2008	AmpF*l*STR™ Sinofiler	Applied Biosystems	16	5
2009	AmpF*l*STR™ Identifiler™ Direct	Applied Biosystems	16	5
2009	AmpF*l*STR™ NGM	Applied Biosystems	16	5
2009	PowerPlex™ 16 HS	Promega	16	5
2009	PowerPlex™ ESX 16	Promega	16	5
2009	PowerPlex™ ESX 17	Promega	17	5
2009	PowerPlex™ ESI 16	Promega	16	5
2009	PowerPlex™ ESI 17	Promega	17	5
2009	PowerPlex™ CS7	Promega	7	3

STR DNA Typing Systems: 2010-2012

Year	Kit Name	Manufacturer	Loci	Dyes
2010	AmpFlSTR™ Identifiler™ Plus	Applied Biosystems	16	5
2010	AmpFlSTR™ NGM SElect	Applied Biosystems	17	5
2010	Investigator ESSplex Plus	Qiagen	16	5
2010	Investigator ESSplex SE	Qiagen	17	5
2010	Investigator ESSplex SE QS	Qiagen	17	5
2010	Investigator IDplex Plus!	Qiagen	16	5
2010	Investigator IDplex Go!	Qiagen	16	5
2010	Investigator Nonaplex ESS	Qiagen	9	4
2010	Investigator Hexaplex ESS	Qiagen	7	4
2010	Investigator HDplex	Qiagen	13	4
2010	Investigator Triplex AFS QS	Qiagen	3	4
2010	Investigator Triplex DSF	Qiagen	3	3
2010	Investigator Argus X-12 QS	Qiagen	14	5
2011	PowerPlex™ 18D	Promega	18	5
2012	PowerPlex™ 21	Promega	21	6

STR DNA Typing Systems: Y-plex Systems

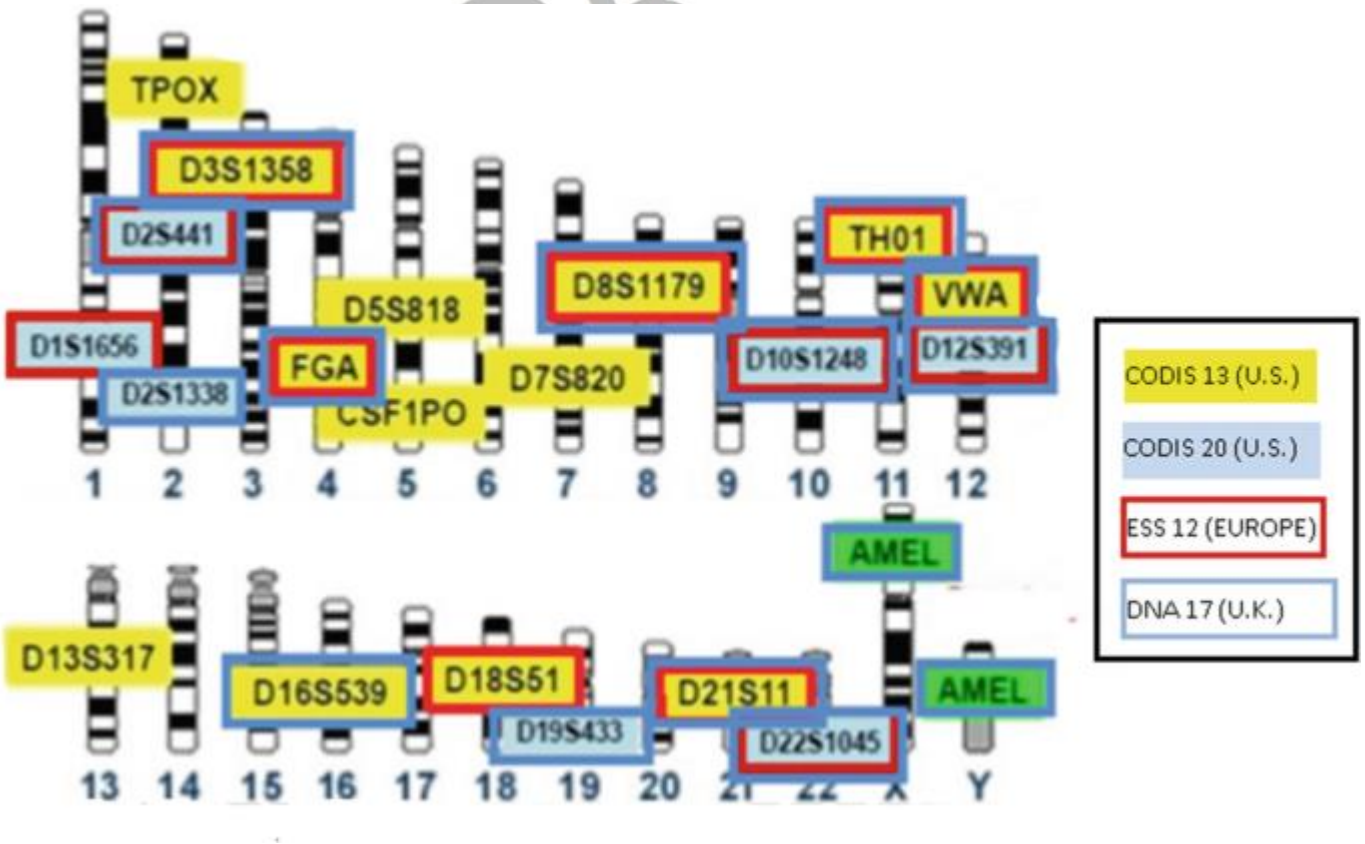
Kits enabled focused on determining male contributors have been offered since 2003

Year	Kit Name	Manufacturer	Loci	Dyes
2003	PowerPlex™ Y	Promega	12	4
2003	Y-PLEX™5	ReliaGene	5	4
2003	Y-PLEX™6	ReliaGene	6	3
2004	Y-PLEX™11 with amelogenin	ReliaGene	12	4
2004	AmpF*l*STR™ Yfiler	Applied Biosystems	16	5
2010	Investigator Argus Y-12 QS	Qiagen	11	5
2012	PowerPlex™ Y23	Promega	23	5
2021	Investigator Argus Y-28 QS	Qiagen	27	6

Butler JM. Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. Forensic Sci Rev 15:91; 2003.

Expanded Core STR Loci

- **Core 20** CODIS loci as of January 1, 2017
- Original 13 loci plus
  - D1S1656
  - D2S441
  - D2S1338
  - D10S1248
  - D12S391
  - D19S433
  - D22S1045



Dixit, S., Rana, M., Kushwaha, P., Kumar, A., Dev, K., Shrivastava, P. (2023). DNA and Its Significance in Forensic Science. In: Shrivastava, P., Lorente, J.A., Srivastava, A., Badiye, A., Kapoor, N. (eds) Textbook of Forensic Science . Springer, Singapore. https://doi.org/10.1007/978-981-99-1377-0_19
https://media.springernature.com/lw685/springer-static/image/chp%3A10.1007%2F978-981-99-1377-0_19/MediaObjects/505341_1_En_19_Fig3_HTML.png

STR DNA Typing Systems: 24plex+

Year	Kit Name	Manufacturer	Loci	Dyes
2012	GlobalFiler™	Applied Biosystems	24	6
2012	PowerPlex® Fusion 5C	Promega	24	5
2016	PowerPlex® Fusion 6C	Promega	27	6
2017	Investigator 24plex GO!	Qiagen	24	6
2017	Investigator 24plex DS	Qiagen	24	6
2018	AmpF*l*STR™ VeriFiler	Applied Biosystems	27	6
2021	Investigator 26plex QS	Qiagen	26	6
2022	PowerPlex® 35GY System	Promega	32	8

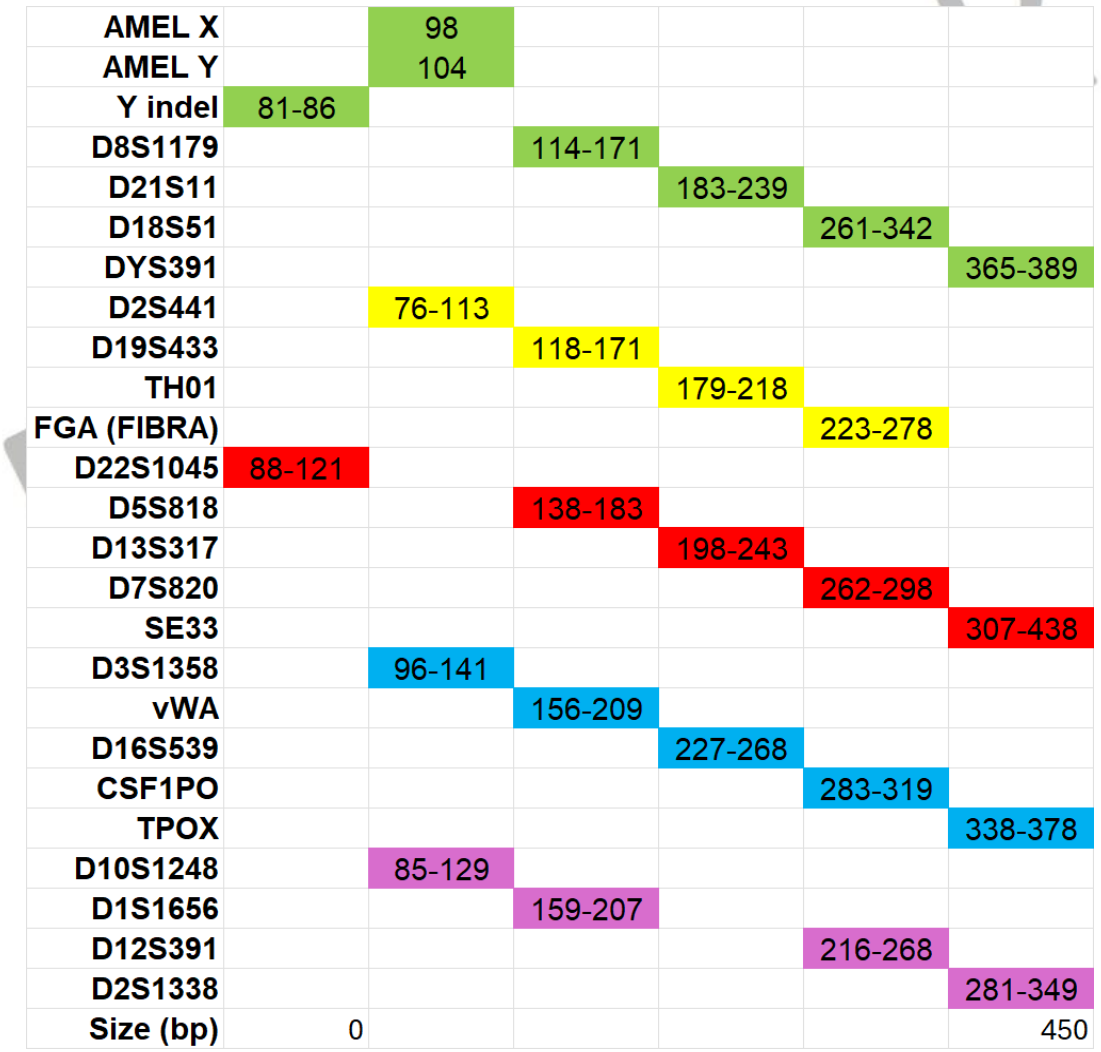
Investigator 24plex also includes QS1 and QS2 diagnostic features

Commercial STR Kit Comparison (24plex+):

Loci, dye color, and size vary so kit selection will vary

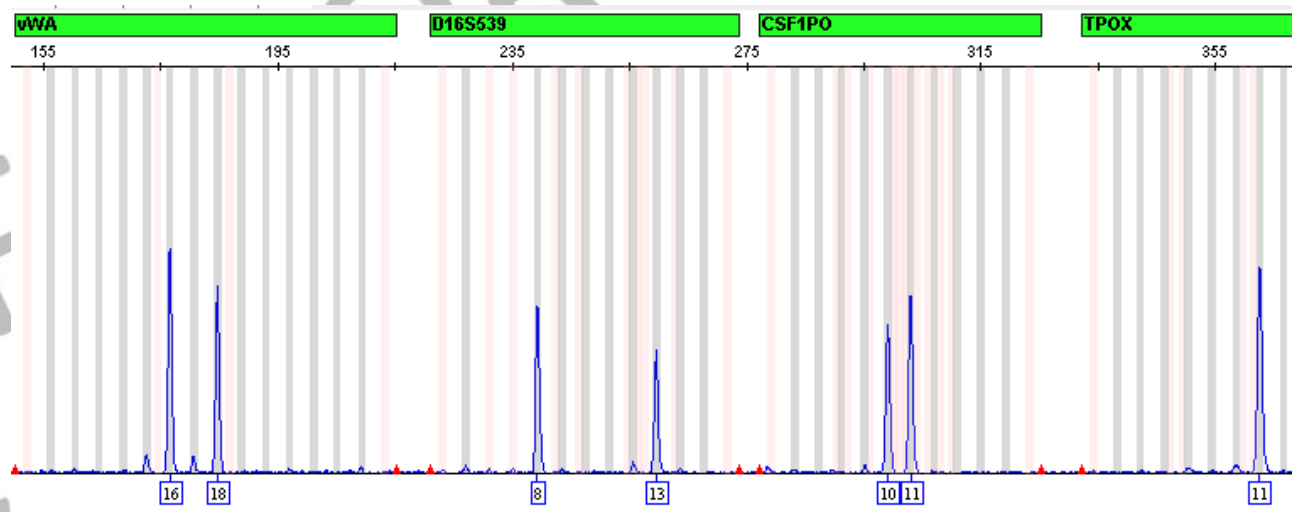
	Investigator 24plex (bp)	Investigator 26plex (bp)	GlobalFiler Expless (bp)	PowerPlex Fusion 5C (bp)	PowerPlex Fusion 6C (bp)
D1S1656	155.5-201	155.5-201	159-207	161-208	161-208
D2S441	79-133	79-133	76-113	214-250	216-252
TPOX	76-127	76-127	338-378	393-441	393-441
D2S1338	360-450	360-450	281-349	224-296	224-296
D3S1358	138-220	138-196	96-141	103-147	103-147
FGA (FIBRA)	287-445.5	296-460	223-278	265-411	143-289
D5S818	287-343	287-343	138-183	321-369	321-369
CSF1PO	151.5-216	151.5-216	283-319	318-362	318-362
SE33	272-434		307-438		270-408
D7S820	345-397	345-397	262-298	267-313	269-313
D8S1179	281-351	281-351	114-171	76-124	76-124
D10S1248	83-143	83-143	85-129	255-299	259-303
TH01	85-136	85-136	179-218	72-115	72-115
D12S391	203-265	203-265	216-268	133-185	133-185
vWA	234-304	237-295	156-209	127-183	127-183
D13S317	219-277	219-277	198-243	302-350	308-358
D16S539	81-150.5	81-150.5	227-268	84-132	84-132
D18S51	134-285	135-235	261-342	134-214	134-214
D19S433	213-279	213-279	118-171	193-245	193-245
D21S11	306-424	371-440	183-239	203-259	203-259
D22S1045	144-191	144-191	88-121	425-464	431-470
PENTA D		197-280		377-450	377-450
PENTA E		293-415		371-466	371-471
AMEL X	74	74	98	89	89
AMEL Y	83	83	104	95	95
DYS391	129-154.5	129-154.5	365-389	442-486	86-130
DYS570					393-453
DYS576					308-356
Y indel			81-86		

Commercial STR Kit GlobalFiler Express (24plex)

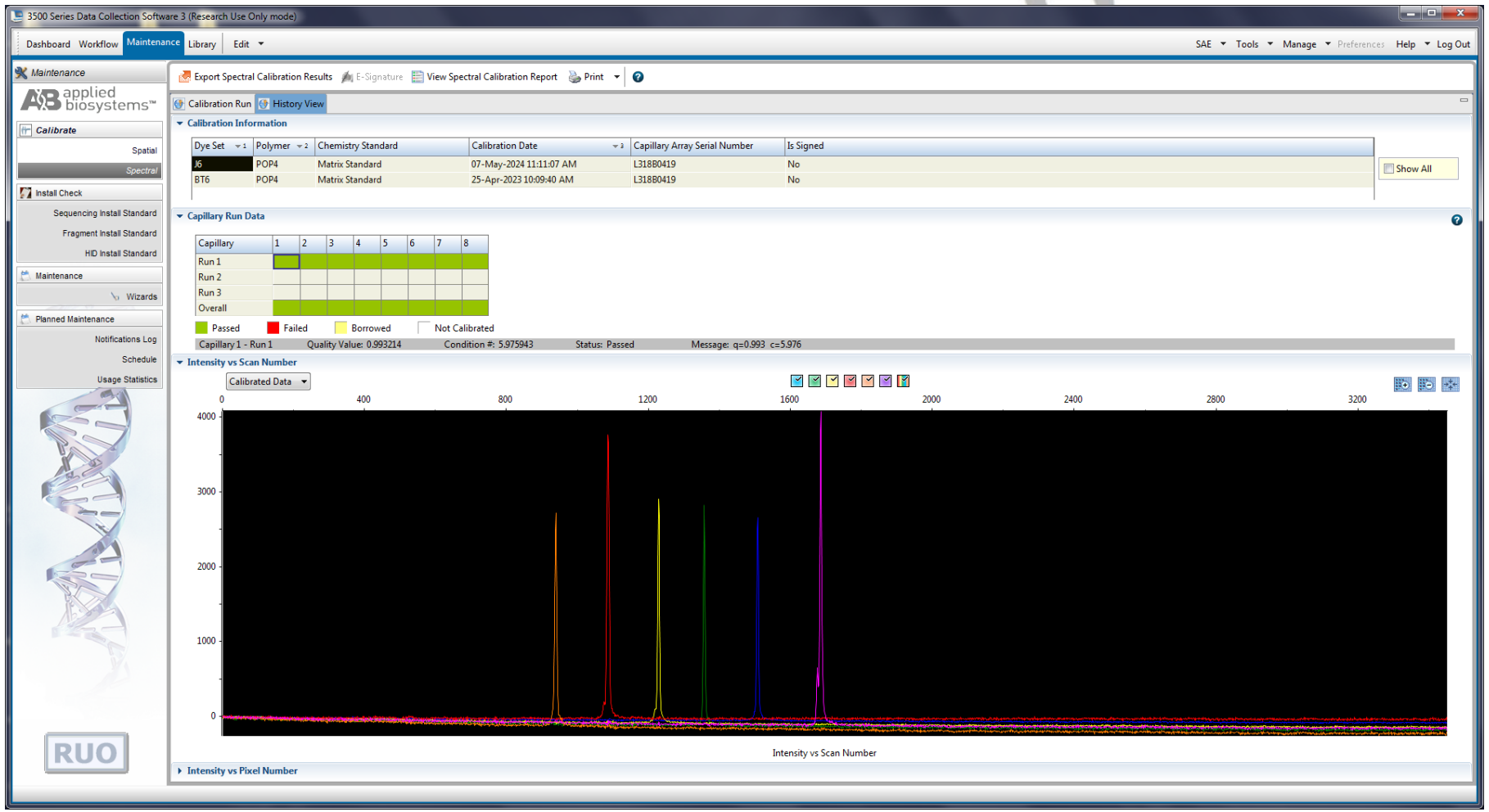


Interpretation: Software

- Peak identification
  - ± 0.5 bp bin around each allele
- Color separation using matrix file
- Sizing using internal size standard
- Correlation of peak retention time with ladder to assign allele

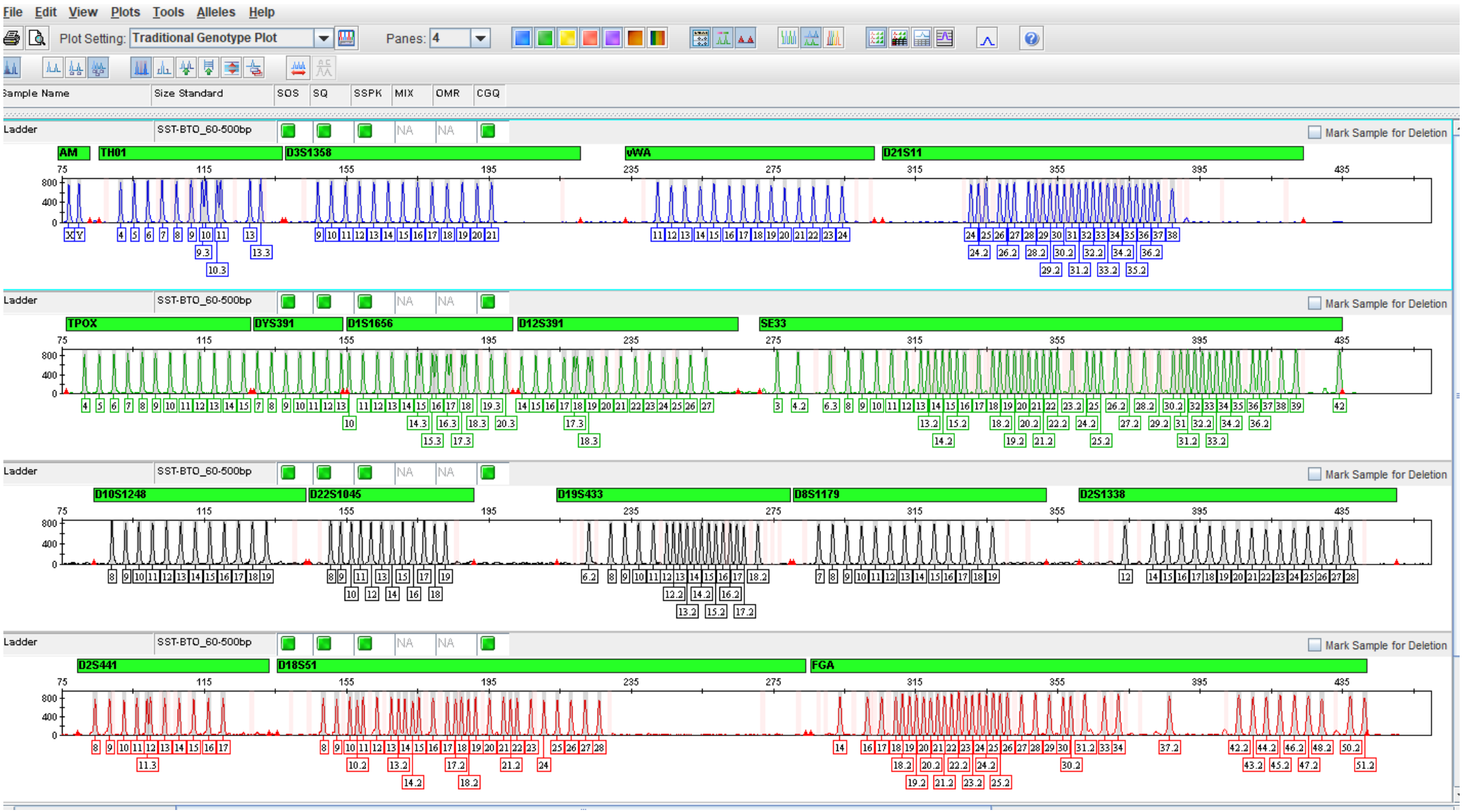


J6 Dye Set Matrix File on AB 3500 CE



Internal Sizing Standard Run on AB 3500 CE

GlobalFiler Ladder For Allele Calls: Run on AB 3500 CE



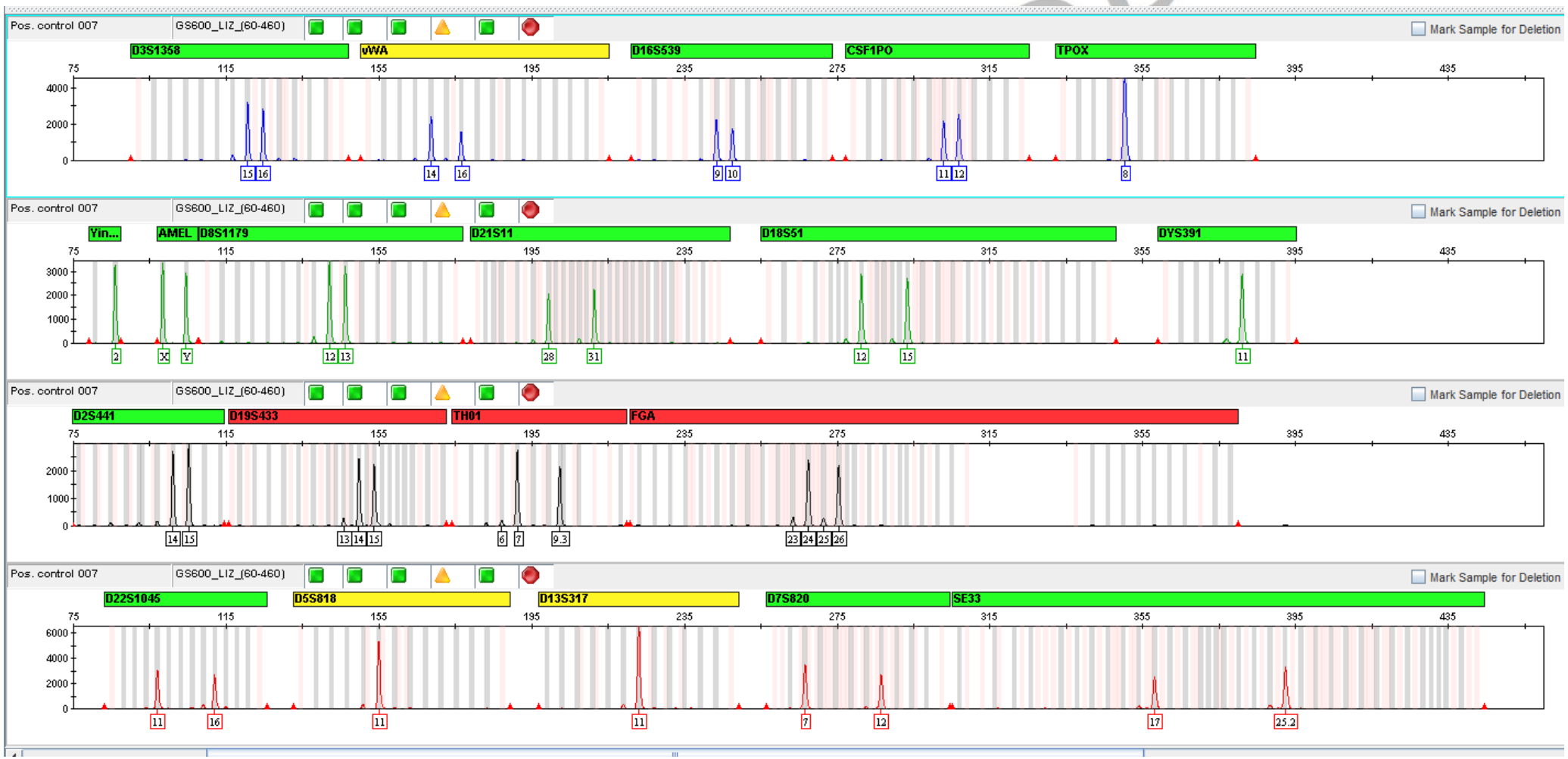
Interpretation: Analyst

- Review of "called" alleles
  - Compare repeat sequences in publications for the allele with the ladder
  - Note artifacts (e.g., pull-up, stutter, incomplete adenylation)
- Editing as need in accordance with lab SOP
  - Report any incomplete repeat motif(s) (i.e., off-ladder allele within the range of the alleles represented by the ladder) using the number of complete repeats with a decimal point and the number of base pairs in the incomplete repeat (e.g., FGA 18.2 allele)
  - For alleles that fall outside the range of the allele ladder, designate the allele as greater than (>) or less than (<) the respective ladder allele, or interpolate if within guidelines
- Compilation of genotype table (if within laboratory SOP guidelines)
- Technical review by another qualified analyst

Interpretation

- Assignment of alleles at each locus
  - Heterozygote
  - Homozygote
- Computing statistics
  - Allele frequency table
  - Minimum allele frequency
  - Random match probability
  - Likelihood ratio
- Single contributor
- Partial profile
- Mixture
- Biological relationships, where applicable
- CODIS
  - Six fully deconvoluted loci minimum for the state database and eight for mixtures and partials

Interpretation



Limitations of the Technology

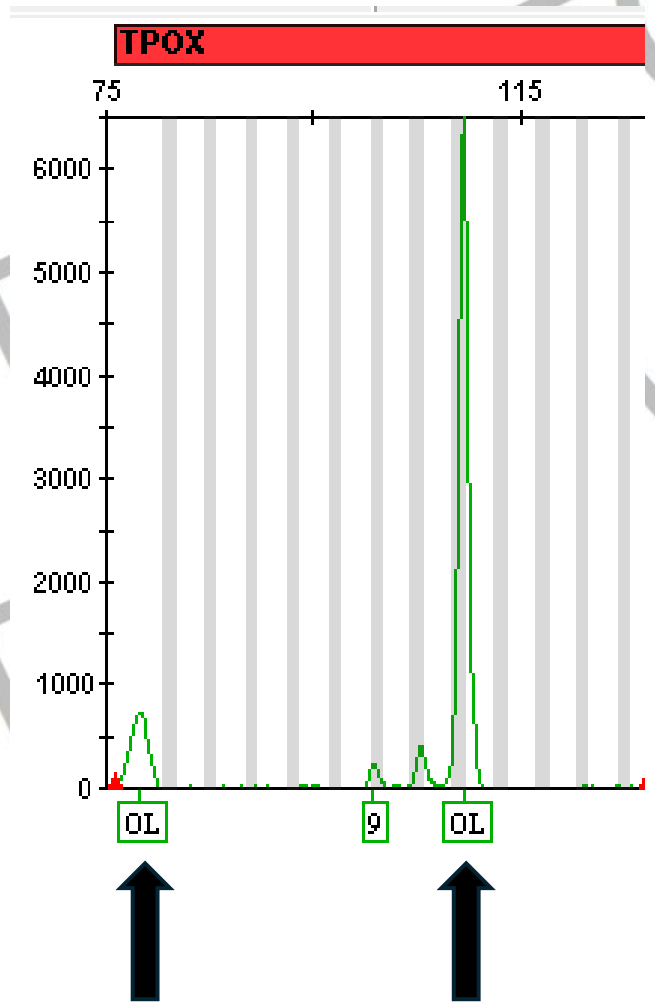
Ladders generally contain alleles with tetranucleotide repeats

- Alleles may not be included in the ladder if they have an incomplete repeat (or are off-ladder or rare)

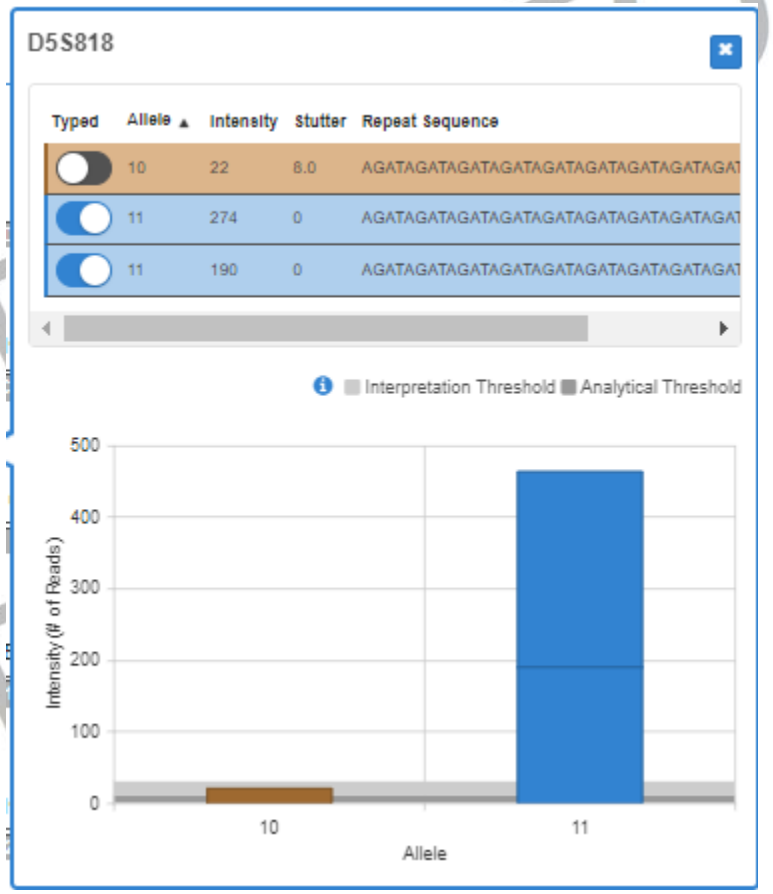
Target loci are not sequenced with a length-based analysis tool

- Variation in sequence is not captured
- Amplicons of the same size and detection label will be called as the same allele

Limitations of the Technology
Off Ladder (OL) Alleles



Limitations of the CE Technology
Fragment Size Interpreted Not Sequence



NGS data yields sequence

Study Questions

- Why are most of the STR loci used for forensic DNA typing tetrameric repeats (i.e., 4 bp) as opposed to dimeric repeats (i.e., 2 bp)? Are there STR loci used in the laboratory that are not tetrameric? If so, which ones?
- List the different classifications used to describe the complexity of the STR core repeat sequences. Give an example for each type.
- What is a non-consensus or microvariant allele? Give three examples for different loci and explain the allele nomenclature.
- What are the CODIS STRs and why were they selected?
- Explain why pentameric (5 bp) loci are more discriminatory. What is a possible limitation of these loci?
- Are the CODIS STR loci human specific?
- Are the STR loci currently used in the lab human specific?
- What is an off-ladder allele and how can it be interpreted?
- Explain how to report an STR profile.
- What is performed to ensure profile accuracy?

Suggested Readings

- Butler, J.M. Advanced Topics in Forensic DNA Typing: Methodology, Ch. 5: STR Loci and Kits, 2011.
- ANSI/ASB Standard 115, Standard for Training in Forensic Short Tandem Repeat Typing Methods using Amplification, DNA Separation, and Allele Detection. 2020. 1st Ed. https://www.aafs.org/sites/default/files/media/documents/115_Std_e1.pdf
- DNA Advisory Board. Quality assurance standards for convicted offender DNA databasing laboratories (approved April 1999), Forensic Science Communications(July 2000) 2. Available at www.fbi.gov/programs/lab/fsc/backissu/july2000/codispre.htm
- DNA Advisory Board. Quality assurance standards for forensic DNA testing laboratories (approved October 1998), Forensic Science Communications(July 2000) 2. Available at www.fbi.gov/programs/lab/fsc/backissu/july2000/codispre.htm
- DNA Commission, ISFH. DNA recommendations: 1994 report concerning further recommendations regarding PCR-based polymorphisms in STR (short tandem repeat) systems, Forensic Science International (1994) 69:103–104.
- Federal Bureau of Investigation. National DNA Index System (NDIS) Procedures Manual. U.S. Department of Justice, Washington, DC, February 1999 (revised).