Updating SRM 2391c: PCR-Based DNA Profiling Standard Why and When?

Becky Hill, Margaret Kline, and Peter Vallone

National Institute of Standards and Technology

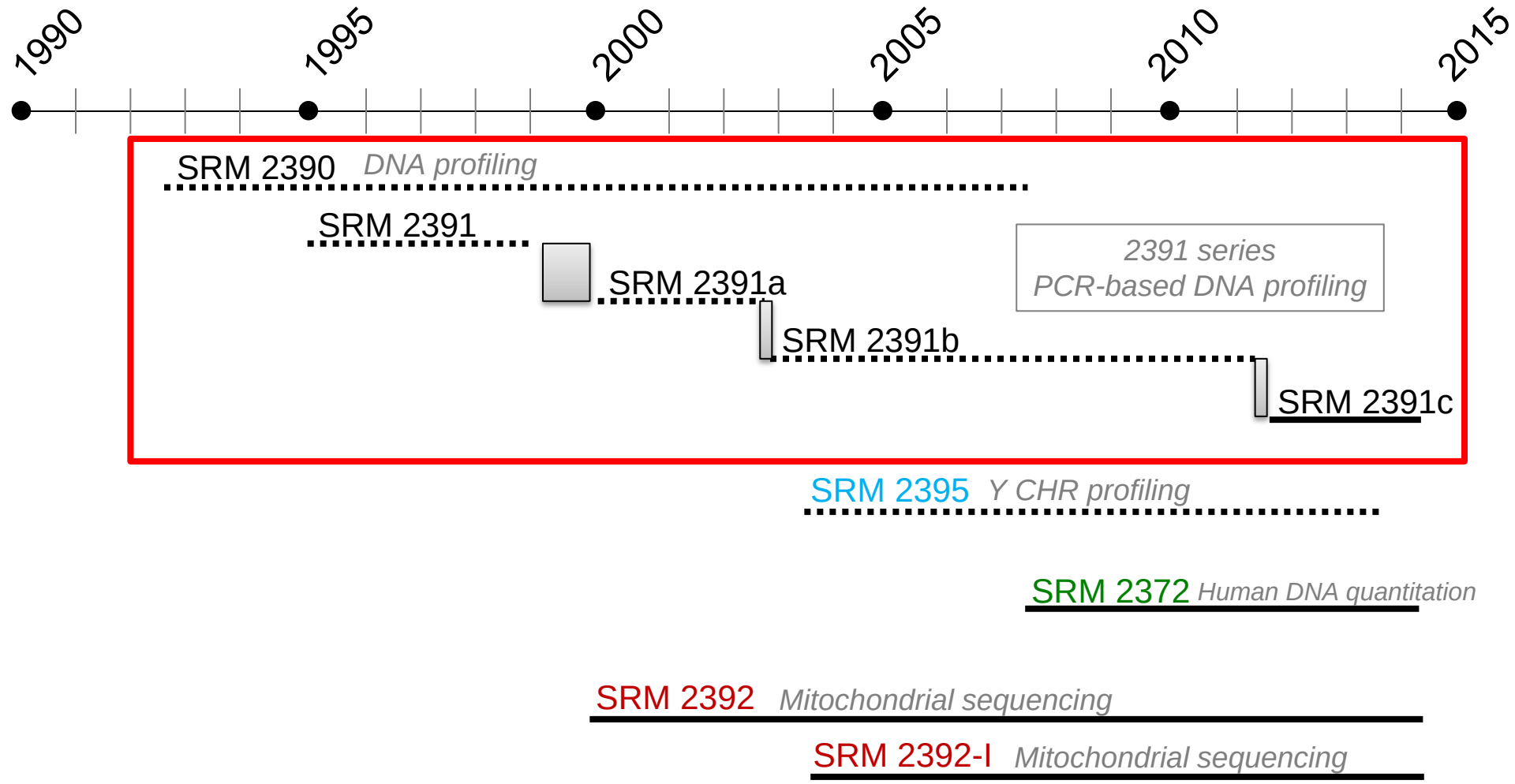
Mid-Atlantic Association of Forensic Scientists

Annual Conference

State College, PA

May 22, 2014

NIST Forensic SRM Timeline



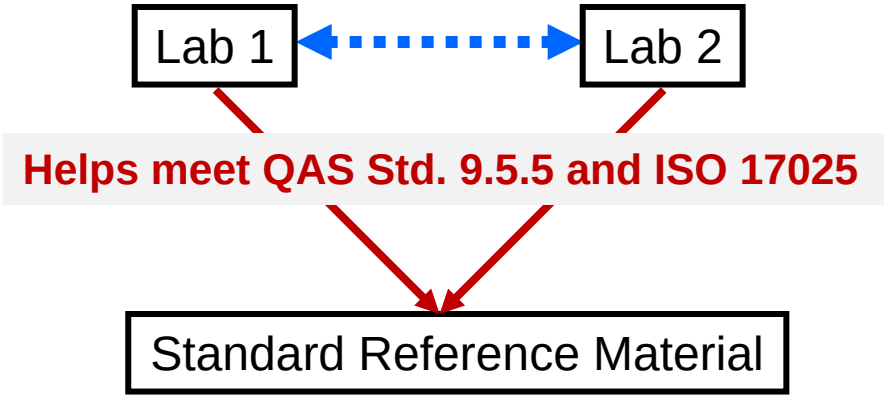
SRM 2391c: PCR-Based DNA Profiling Standard

- **S**tandard **R**eference **M**aterial 2391c is intended primarily for use:
  - Standardization of forensic and paternity quality assurance procedures for PCR-based genetic testing
  - Instructional law enforcement or non-clinical research purposes
  - Quality assurance when assigning values to in-house control materials

SRM 2391c:
PCR-Based DNA Profiling Standard

- Components A through D are DNA extracts in liquid form
- Components E and F are DNA spotted on 903 paper or FTA paper
- Certified values are for STR alleles based on length polymorphisms observed using capillary electrophoresis

Genomic DNAs characterized for the expanded CODIS core loci and Y-STRs

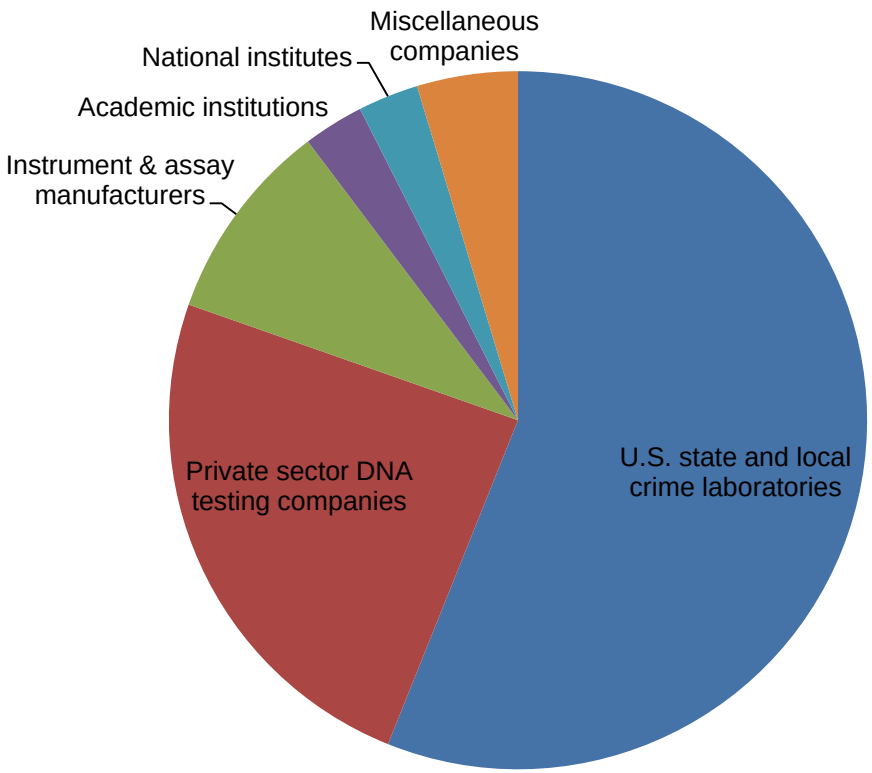


Calibration with SRMs enables confidence in comparisons of results between laboratories

SRM 2391c Sales

Unit sales			
FY11	FY12	FY13	YTD
26	143	112	62

SRM 2391c Customers



	Crime Laboratori
1	ALABAMA DEPT
2	ALAMEDA COUN
3	ANNE ARUNDEL
4	ANOKA COUNTY
5	ARMED FORCES
6	AZ DPS CRIME LA
7	BALTIMORE COU
8	BROWARD SHERI
9	BUREAU OF ATF
10	CALIFORNIA DEP
11	CHARLOTTE-MEC
12	CITY OF AUSTIN
13	CITY OF BOSTON
14	CITY OF SCOTTSD
15	CO BUREAU OF I
16	CT FORENSIC LAE
17	CUYAHOGA COU
18	DALLAS COUNTY
19	DENVER POLICE
20	DPS- DNA CRIME
21	ERIE COUNTY
22	FLORIDA DEPT O
23	FRESNO COUNTY
24	GLENDALE POLIC
25	GREENVILLE COU
26	ILLINOIS STATE P
27	KANSAS BUREAU
28	KENTUCKY STATE
29	KERN COUNTY CF
30	LAKE COUNTY RE

utes
LLEGHENY
VERSITY
INSTITUTE
VERNMENT
CED
C
DIES
RADING LIMI

Establishing Traceability to NIST SRM 2391c

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

Lot of DNA material

Run against SRM 2391c

Everyday use calibrant

Updating NIST SRM 2391c: Topics for Discussion

- Purpose of updating SRM 2391c
- Which loci are currently included in the Certificate of Analysis
- Coverage of the new loci is under development
- Full STR allele sequence coverage to aid future next-generation sequencing efforts
- When will the update be complete?

Purpose of Sequencing SRM 2391c

- To further characterize SRM 2391c to determine interesting genomic characteristics within STR fragments (SNPs, insertions/deletions, etc.)
- Initial progress presented as a poster (ISFG 2013)

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Hill-ISFG2013-SRM2391c.pdf

SRM 2391c: PCR-Based DNA Profiling Standard

- Includes 6 components:

Table 1. Description of Components in SRM 2391c

Component		Description	Amount	Concentration[a]
A	Single-source	Anonymous single-source female genomic DNA in TE^{-4} buffer	50 µL	1.1 – 2.1 ng/µL
B		Anonymous single-source male genomic DNA in TE^{-4} buffer	50 µL	1.1 – 2.1 ng/µL
C		Anonymous single-source male genomic DNA in TE^{-4} buffer	50 µL	1.1 – 2.1 ng/µL
D	Mixture	Mixed-source (Components A and C) genomic DNA in TE^{-4} buffer	50 µL	1.1 – 2.1 ng/µL
E	Stain	Anonymous single-source female cells spotted on 903 paper	Two 6 mm punches	7.5×10^4 cells per punch
F	Stain	Anonymous single-source male cells spotted on FTA paper	Two 6 mm punches	7.5×10^4 cells per punch

[a]DNA concentrations and cell counts are nominal values and are not intended for use as quantitative standards.

https://www-s.nist.gov/srmors/view_cert.cfm?srm=2391C

Certified Genotypes

Concordance with STR Kits

Autosomal STR Loci	Y-STR Loci
D2S1338	DYS19
D2S441	DYS385a
D3S1358	DYS385b
D5S818	DYS389I
D7S820	DYS389II
D8S1179	DYS390
D10S1248	DYS391
D13S317	DYS392
D16S539	DYS393
D18S51	DYS437
D19S433	DYS438
D21S11	DYS439
D22S1045	*Amelogenin
CSF1PO	
FGA	
TH01	
TPOX	
vWA	

DNA Sequencing of Alleles

Autosomal STR Loci	Y-STR Loci
D1S1656	DYS448
D8S1115	DYS456
D12S391	DYS458
Penta D	DYS635
Penta E	DY-GATA-H4
SE33	

41 STR Markers + Amelogenin are certified
26% have been Sanger Sequenced

***A NIST certified value is a value for which NIST has the highest confidence in its accuracy in that all known or suspected sources of bias have been investigated or taken into account.**

>2 STR Kits were tested for concordance

Assigning NIST Certified Values

- Goal: Sanger sequence all autosomal and Y-STR markers in commercial multiplex kits
- The remaining 30 markers are currently being sequenced for Components A-C, E & F to further characterize the repeat structure and flanking sequence.
- New markers in recently released commercial kits (PP Fusion, PP21, PPY23, GlobalFiler, & Yfiler Plus) are also included

Methods for Sanger Sequencing

- NIST DNA sequencing procedures and all sequencing primers were published in 2011 (see S1)
- Note: alternative primers were designed for D19S433

Forensic Science International: Genetics 5 (2011) 329–332

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

STR sequence analysis for characterizing normal, variant, and null alleles

Margaret C. Kline *, Carolyn R. Hill, Amy E. Decker [1], John M. Butler

National Institute of Standards and Technology, 100 Bureau Drive, M/S 8312, Gaithersburg, MD 20899, USA

Sequencing Flow Chart

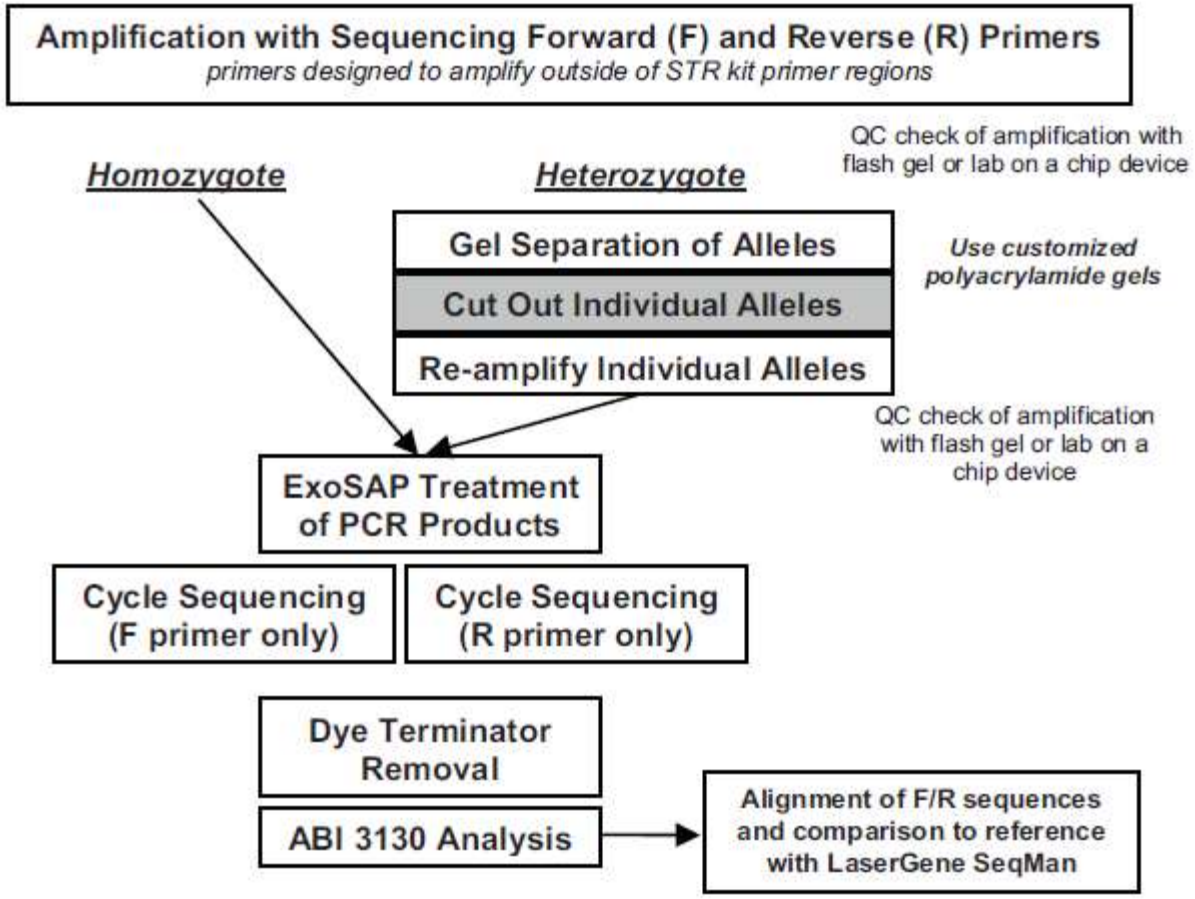
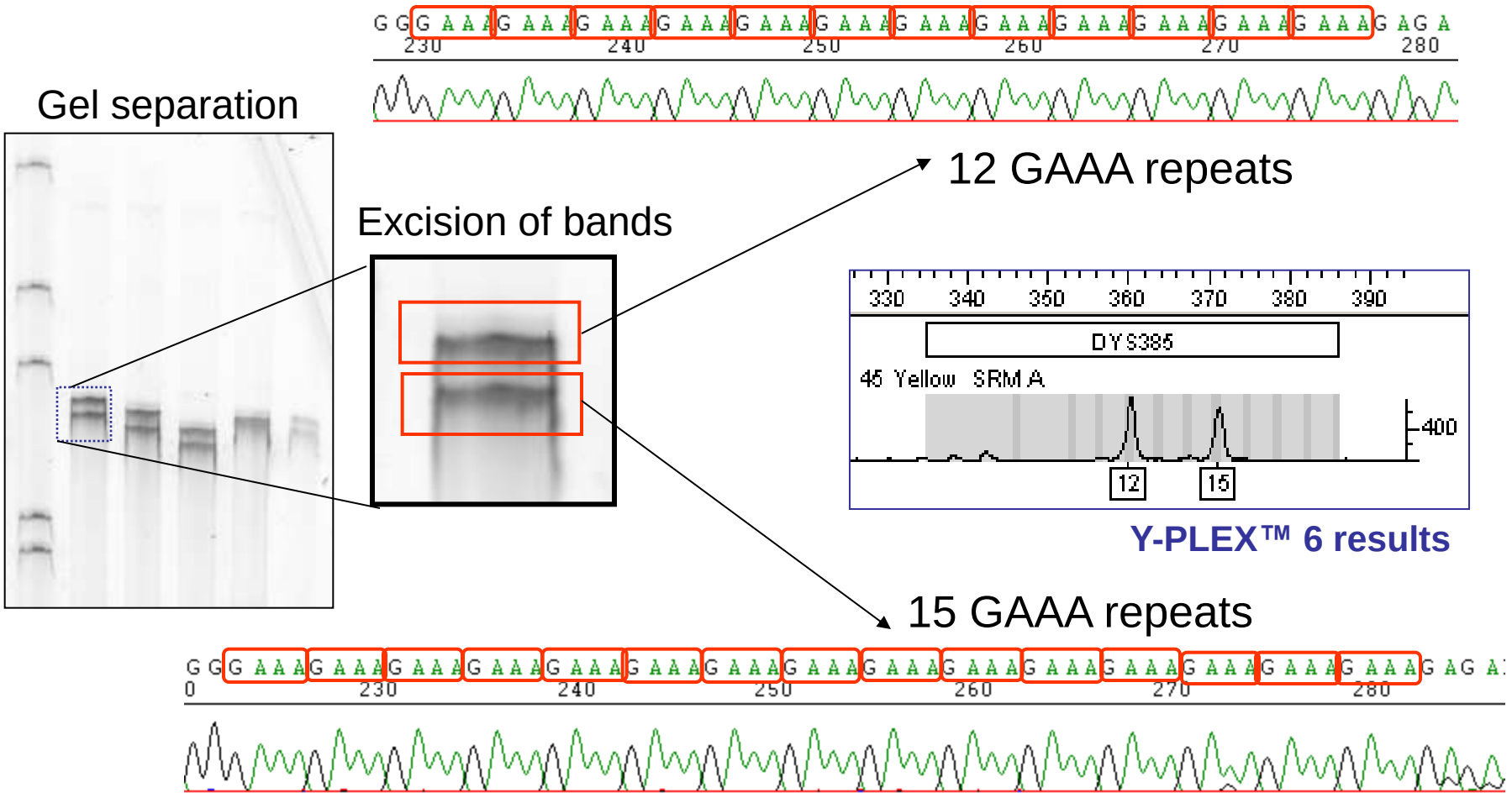


Fig. 1. Summary of the steps used in sequencing STR alleles.

Sequencing Individual Heterozygous (DYS385) Alleles



Kline, M.C., Hill, C.R., Decker, A.E., Butler, J.M. (2011) STR sequence analysis for characterizing normal, variant, and null alleles. Forensic Sci. Int. Genet. 5(4): 329-332

GenBank Reference Sequences

- The GenBank Accession numbers and reference alleles were obtained were based on the May 2004 assembly of the human genome, build 35.
- Sequences were aligned de novo using LaserGene SeqMan software and compared to SeqBuilder maps based on the listed GenBank reference sequences.

Marker	GenBank Accession Number	Marker	GenBank Accession Number	Marker	GenBank Accession Number	Marker	GenBank Accession Number
D1S1656	G07820	D13S317	AL353628.2	SE33	V00481	DYS393	AC006152
D2S1338	AC010136	D16S539	AC024591.3	TH01	D00269	DYS437	AC002992
D2S441	AC079112	D18S51	AP001534	TPOX	M68651	DYS438	AC002992
D3S1358	AC099539	D19S433	AC008507.6	vWA	M25858	DYS439	AC002992
D5S818	AC008512	D21S11	AP000433	DYS19	AC017019	DYS448	AC025227
D7S820	AC004848	D22S1045	AL022314	DYS385	AC022486	DYS456	AC010106.2
D8S1179	AF216671	CSF1PO	X14720	DYS389	AF140635	DYS458	AC010902
D8S1115	AC090739	FGA	M64982	DYS390	AC011289	DYS635	AC004772
D10S1248	AL391869	Penta D	AP001752	DYS391	AC011302	Y GATA H4	AC011751
D12S391	G08921	Penta E	AC027004	DYS392	AC06152		

Sequencing Results

- All sequencing results of Components A-C for 41 STR markers, including repeat structures of individual alleles, can be found on the following poster:

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Hill-ISFG2013-SRM2391c.pdf

Marker	Component	Allele	Allele Repeat Structure
D8S1179	C	17	$[TCTA]_2$ TCTG $[TCTA]_{14}$
D12S391	A	22	$[AGAT]_{13}$ $[AGAC]_8$ AGAT
D12S391	C	19	$[AGAT]_{13}$ $[AGAC]_5$ AGAT
D12S391	C	23	$[AGAT]_{12}$ $[AGAC]_{10}$ AGAT
D21S11	B	32	$[TCTA]_4$ $[TCTG]_6$ {$[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCATA} $[TCTA]_{14}$
SE33	C	31.2	$[AAAG]_2$ AG $[AAAG]_3$ AG $[AAAG]_9$ AAAAAG $[AAAG]_{21}$ G AAGG$[AAAG]_2$ AG
DYS389II	B	31	$[TCTG]_6$ $[TCTA]_{12}$ $[TCTG]_3$ $[TCTA]_{10}$
DYS458	B	17.2	$[GAAA]_{15}$ AA $[GAAA]_2$
DYS635	B	20	$[TCTA]_4$ $[TGTA]_2$ $[TCTA]_2$ $[TGTA]_2$ $[TCTA]_{10}$
DYS635	C	21	$[TCTA]_4$ $[TGTA]_2$ $[TCTA]_2$ $[TGTA]_2$ $[TCTA]_{11}$

Novel repeat motifs that were not listed in Butler J.M. (2012) or STRBase fact sheets

SNPs Found in Repeat Flanking Regions

- Multiple SNPs were found in the DNA sequence in the repeat flanking regions. Primers that bind on SNPs can result in null alleles when STR typing.
- Note that the variants characterized in this work are constrained by the size of the original PCR amplicon generated (Kline et al. 2011).

Marker	Component	Allele	Flanking Region Variants
D5S818	A	12	T→C 13 bp us of the repeat
D5S818	B	13	T→C 13 bp us of the repeat
D5S818	B	13	G→T 4 bp ds of the repeat
D5S818	C	10	T→C 13 bp us of the repeat
D5S818	C	11	T→C 13 bp us of the repeat
D7S820	C	10	T→G 65 bp ds of the repeat
D13S317	C	11	A→C 115 bp ds of the repeat
D16S539	A	10	A→C 16 bp ds of the repeat
D16S539	A	10	C→A 95 bp us of the repeat
D16S539	A	11	C→A 95 bp us of the repeat
D16S539	B	10	C→A 95 bp us of the repeat
D16S539	C	10	C→A 95 bp us of the repeat
Penta E	A	10	G→A 123 bp us of the repeat
Penta E	A	10	A→G 268 bp us of the repeat
Penta E	A	10	A→C 280 bp us of the repeat
Penta E	B	7	G→A 123 bp us of the repeat
Penta E	B	7	A→G 268 bp us of the repeat
Penta E	B	7	A→C 280 bp us of the repeat
Penta E	B	15	G→A 123 bp us of the repeat
Penta E	B	15	A→G 268 bp us of the repeat
Penta E	B	15	A→C 280 bp us of the repeat
TPOX	A	8	T→G 148 bp ds of the repeat
TPOX	B	8	T→G 148 bp ds of the repeat

Abbreviations: bp = base pairs, us = upstream, ds = downstream

Sequencing Progress

SRM 2391c - Autosomal STR Sequencing					
Marker	A	B	C	E	F
D1S1656	Yes	Yes	Yes	Yes	Yes
D2S1338	Yes	Yes	Yes	Yes	Yes
D2S441	Yes	Yes	Yes	Yes	Yes
D3S1358	Yes	Yes	Yes	Yes	Yes
D5S818	Yes	Yes	Yes	Yes	Yes
D6S1043	Yes	Yes	Yes	Yes	No
D7S820	Yes	Yes	Yes	Yes	Yes
D8S1179	Yes	Yes	Yes	Yes	Yes
D8S1115	Yes	Yes	Yes	Yes	Yes
D10S1248	Yes	Yes	Yes	Yes	No
D12S391	Yes	Yes	Yes	Yes	Yes
D13S317	Yes	Yes	Yes	Yes	Yes
D16S539	Yes	Yes	Yes	Yes	Yes
D18S51	Yes	Yes	Yes	Yes	Yes
D19S433	Yes	Yes	Yes	Yes	Yes
D21S11	Yes	Yes	Yes	Yes	Yes
D22S1045	Yes	Yes	Yes	Yes	Yes
CSF1PO	Yes	Yes	No	No	No
FGA	Yes	Yes	Yes	Yes	Yes
Penta D	Yes	Yes	Yes	Yes	Yes
Penta E	Yes	Yes	Yes	Yes	Yes
SE33	Yes	Yes	Yes	Yes	Yes
TH01	Yes	Yes	Yes	Yes	Yes
TPOX	Yes	Yes	Yes	Yes	Yes
vWA	Yes	Yes	Yes	Yes	Yes

SRM 2391c - Y-STR Sequencing			
Marker	B	C	F
DYS19	Yes	Yes	Yes
DYS385a	Yes	Yes	No
DYS385b	Yes	Yes	No
DYS389I	Yes	Yes	No
DYS389II	Yes	Yes	Yes
DYS390	Yes	Yes	Yes
DYS391	Yes	Yes	Yes
DYS392	Yes	Yes	Yes
DYS393	Yes	Yes	Yes
DYS437	No	No	No
DYS438	Yes	Yes	Yes
DYS439	Yes	Yes	Yes
DYS448	Yes	Yes	Yes
DYS456	Yes	Yes	No
DYS458	Yes	Yes	No
DYS635	Yes	Yes	Yes
Y GATA H4	Yes	Yes	Yes

SRM 2391c - New Y-STR Sequencing			
Marker	B	C	F
DYS449	No	No	No
DYS460	No	No	No
DYS481	Yes	Yes	Yes
DYS518	No	No	No
DYS533	No	No	No
DYS549	No	No	No
DYS570	Yes	Yes	Yes
DYS576	Yes	Yes	Yes
DYS627	No	No	No
DYS643	No	No	No
DYF387S1ab	No	No	No

Other Candidates for Sequencing

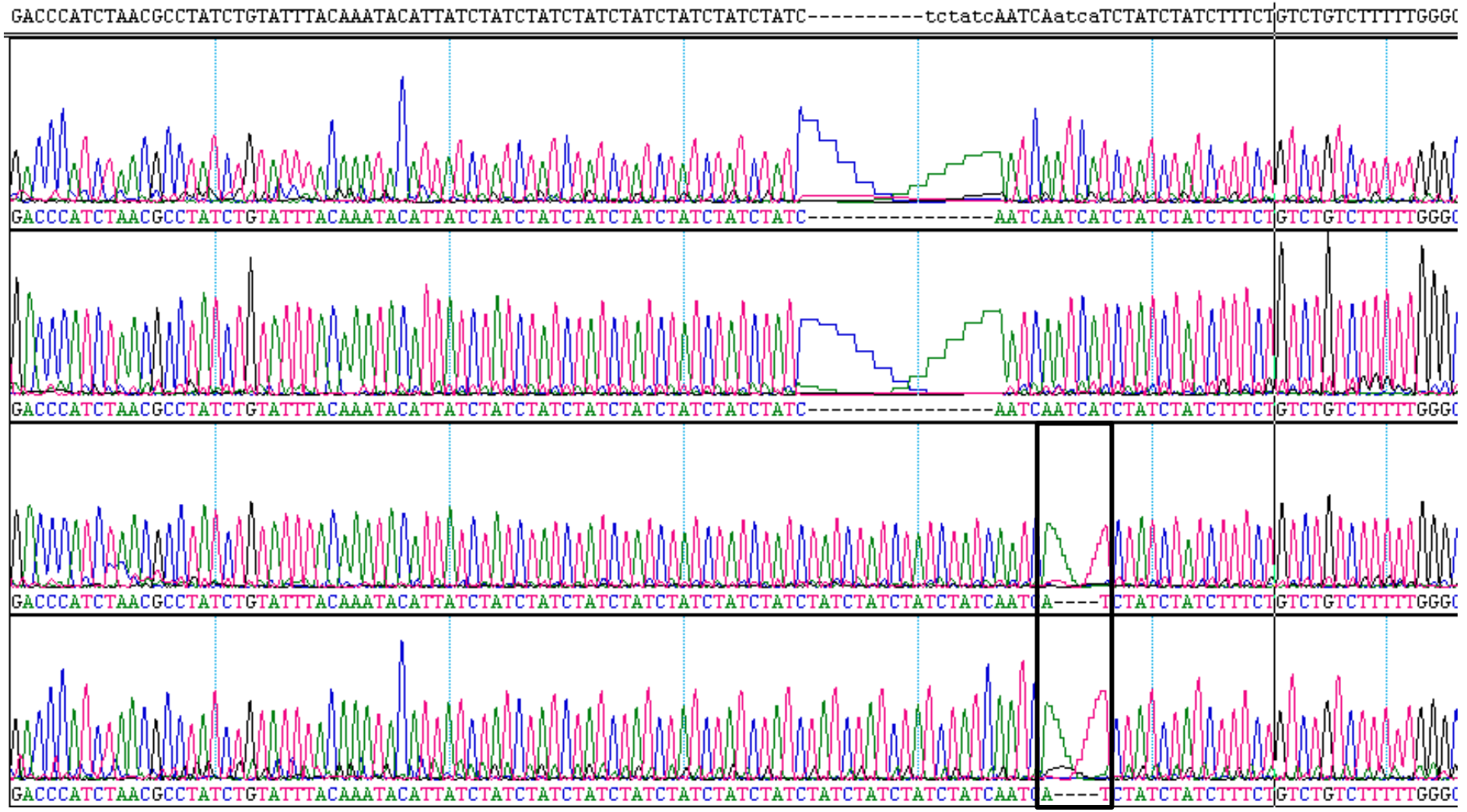
- Additional non-core autosomal STR markers
  - D6S1043 (Sinofiler, PowerPlex 21)
  - 22 miniSTR loci (not including D2S441, D10S1248, D22S1045, D8S1115)
  - Penta C
  - FFFL loci (F13A01, F13B, FESFPS, LPL)
- Y-STR markers to sequence
  - DYS460, DYS481, DYS533, DYS549, DYS643 (PowerPlex Y23, Yfiler Plus)
- Rapidly mutating (RM) Y-STRs
  - 13 total (**DYF387S1a/b**, DYF399S1, DYF403S1a/b, DYF404S1, **DYS449**, **DYS518**, DYS526a/b, DYS547, **DYS570**, **DYS576**, DYS612, DYS626, **DYS627**)

*Loci in red have already been sequenced

Interesting Examples

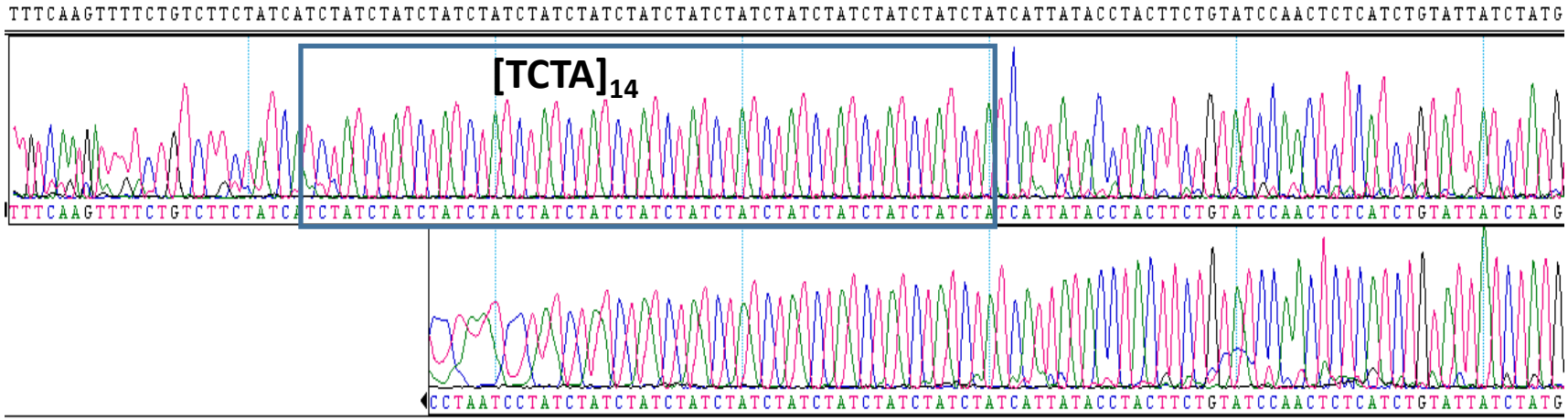
- Occasionally we find samples that sequence differently than what are genotyped
- Usually this is due to an insertion, deletion, or SNP inside the primer sequences that amplify the marker for that sample
- So far, we have seen three examples of this while sequencing the components of SRM 2391c

D13S317, Component C



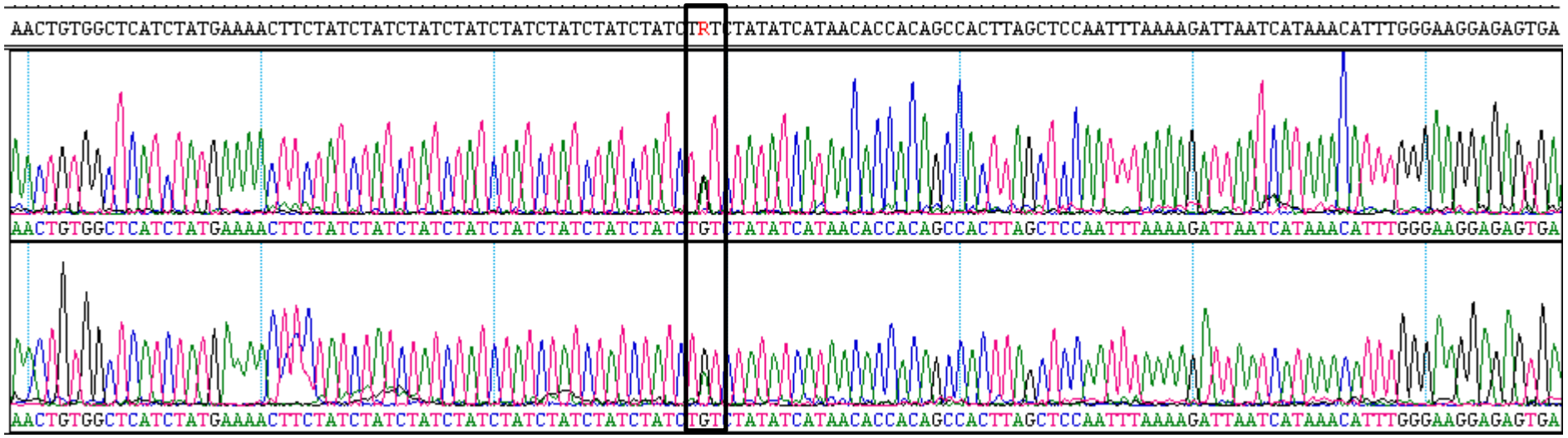
D13S317, Component C, is a homozygous sample with an (11,11) genotype. However, it is sequencing as a (12,12) with $[TATC]_{12}$ repeats. The cause is a 4 bp [ATCA] deletion (del) located 6 bp downstream (ds) of the repeat, that is within most, if not all, of the primer sequences used in commercial STR typing kits.

DYS389 I, Component F



DYS389 I, Component F, genotypes as a 13 haplotype. However, it is sequencing as a 14 with $[TCTA]_{14}$ repeats. The cause is most likely 4 bp deletion (del) that is within most, if not all, of the primer sequences used in commercial STR typing kits (PPY, PPY23, & Yfiler). A forward sequencing primer must be redesigned to capture the deletion.

D2S441, Component A



D2S441, Component A, is a homozygous sample with a (10,10) genotype. This is an unusual example of where each allele in a homozygote has a different repeat structure. One allele has a simple repeat of $[TATG]_{10}$ and the other has an A→G SNP, causing the repeat structure to be $[TCTA]_8$ TCTG TCTA. This is evident because of the mixture of the A and G bases within the repeat indicating that one allele has the A and the other has the G.

SRM 2391c STR Kit Testing

- With the expansion of the CODIS database, many new larger multiplexes have recently been released and new instruments are being run
  - PP Fusion and GlobalFiler
  - 3500, 3500xl

- First step at NIST is to test SRM 2391c with all new STR typing kits
  - Concordance with NIST reference materials is important

SRM 2391b/2391c PCR-Based Profiling Standard

- The first set of samples run with new STR multiplex kits is SRM 2391b/SRM 2391c

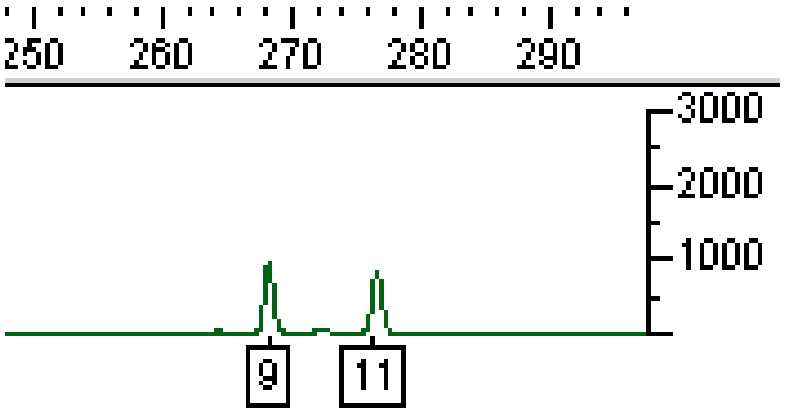
- All new kits tested have been completely concordant with the certified values of all markers for each component for SRM 2391b and 2391c

- One exception for SRM 2391b: MiniFiler
  - Genomic 8 with D16S539

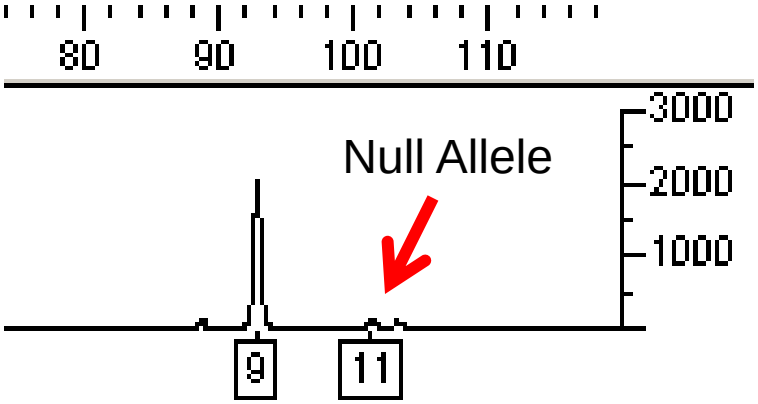
SRM 2391b Genomic 8 with D16S539

All allele calls with MiniFiler for CSF1PO, D7S820, D13S317, D18S51, D21S11, FGA, and D16S539 (with the exception noted below) ***match previously certified values.***

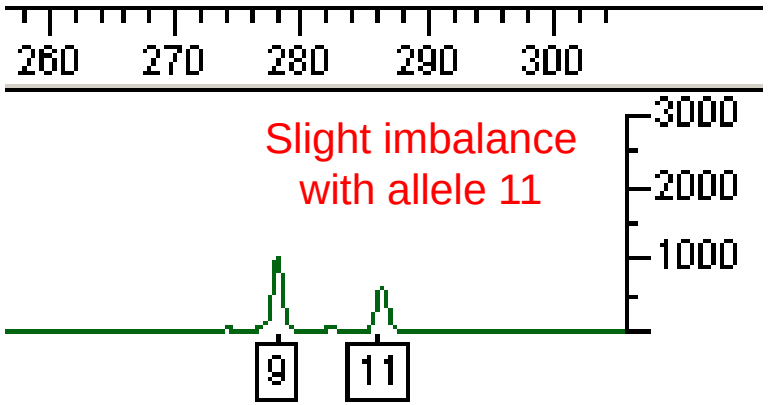
Identifiler



MiniFiler



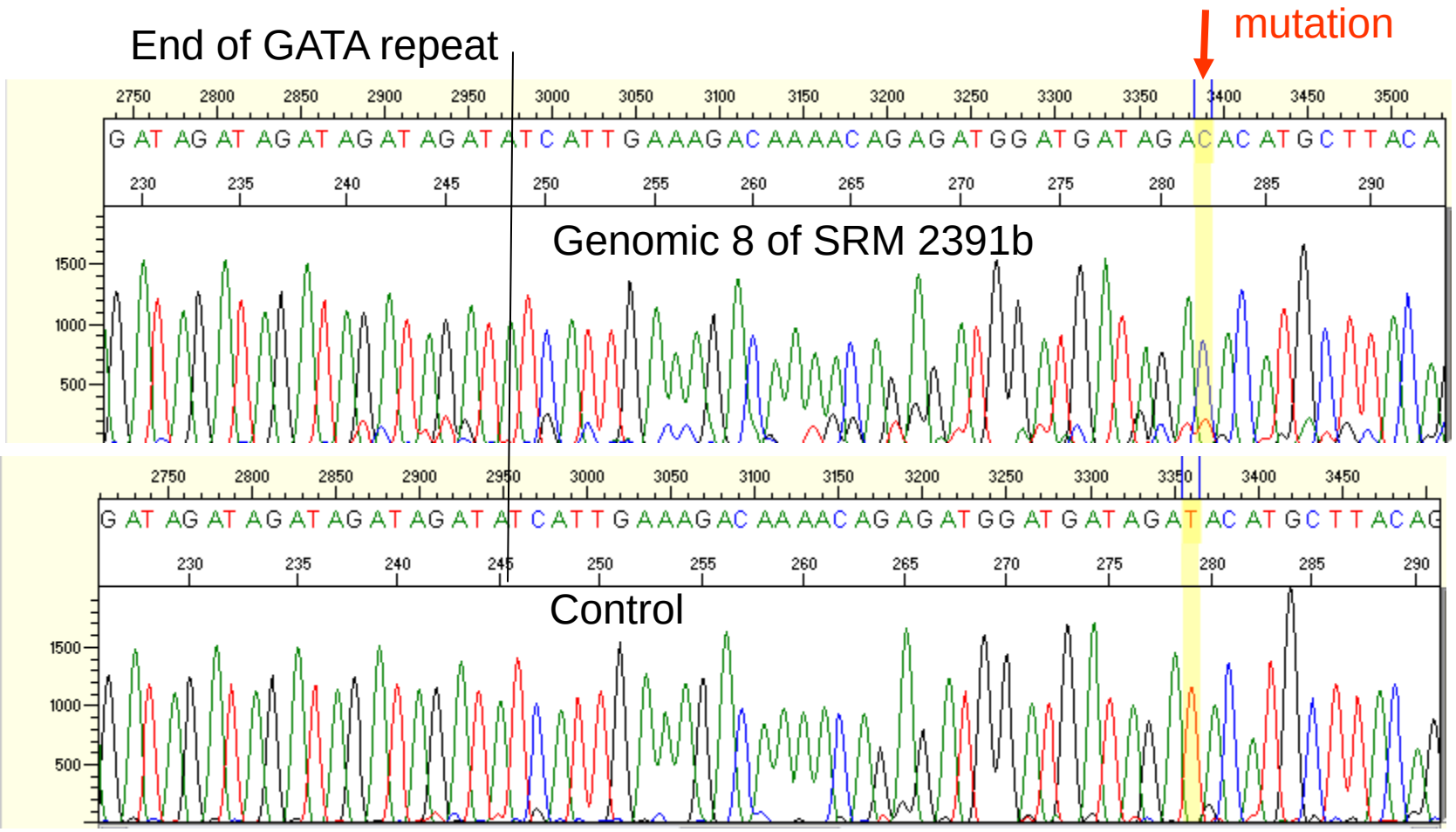
PowerPlex 16



***Due to primer binding site mutation**

D16S539 SRM 2391b Genomic 8

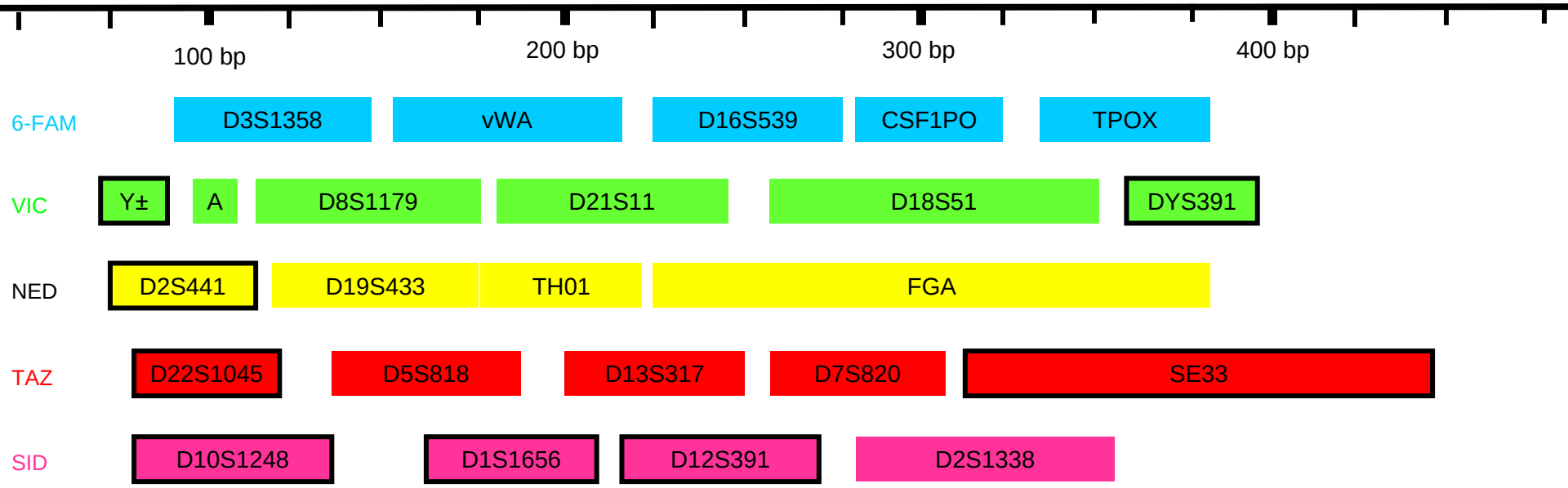
T→C mutation 34 bp downstream of the repeat



Position of the T→C probably affects the reverse primer of Minifiler and is the 3rd base found the 5'end of the Reverse PP16 primer. This could explain the imbalance of the allele seen when using PP16.

Life Technologies **GlobalFiler**

24plex



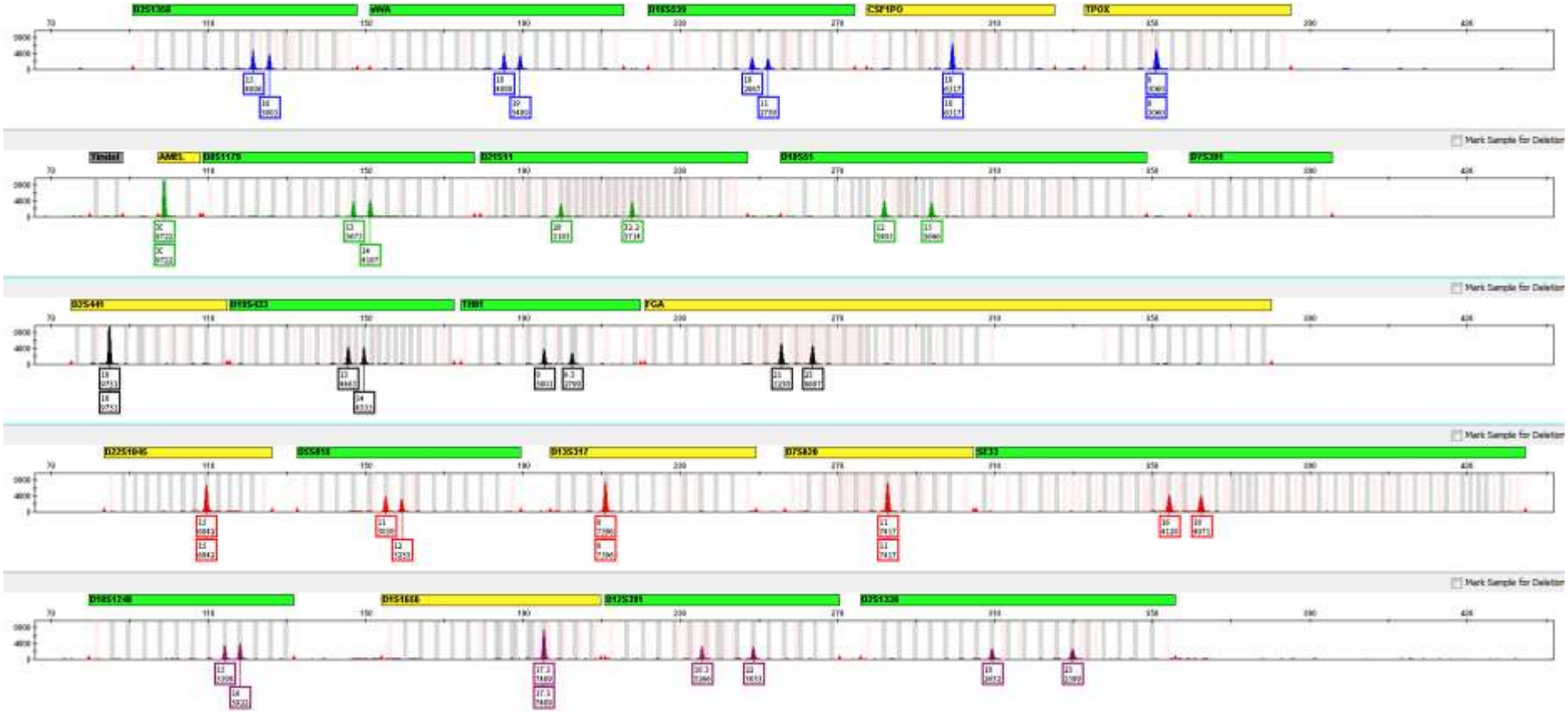
- 24 STR loci in 6 dyes (3500 use or 3130 upgrade required)
  - Includes SE33 and a Y-indel
- GlobalFiler Express: direct amplification capabilities
  - Single source samples: 40 min amplification
- GlobalFiler Casework (not yet available)
  - Casework samples: 80 min amplification

Two separate kits

- GlobalFiler gives ~12 orders of magnitude improvement using the NIST 1036 data set

http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Human-Identification/globalfiler_str_kit.html

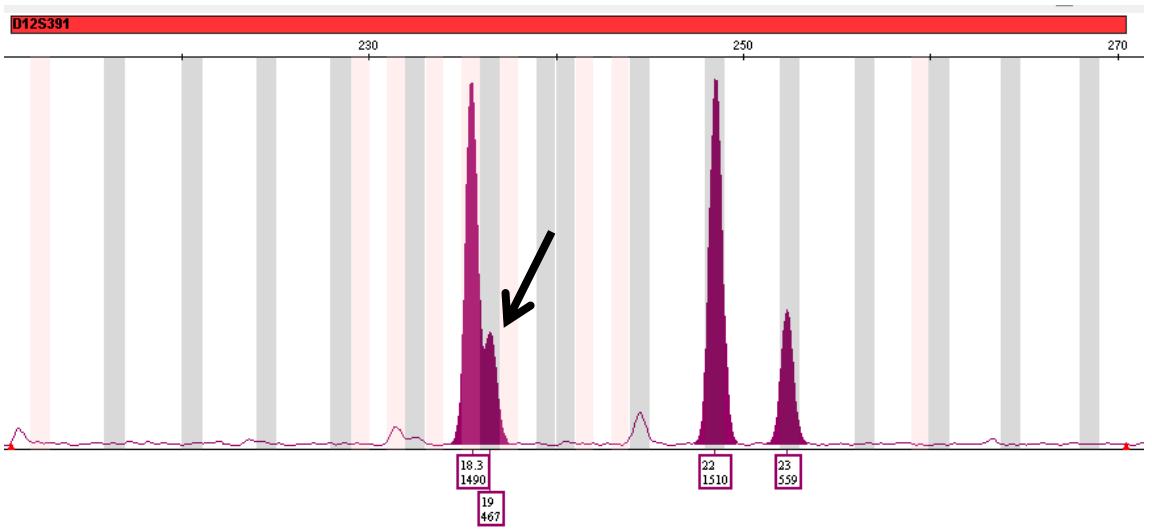
Component A


D3S1358
vWA
D16S539
CSF1PO
TPOX
Mark Sample for Deletion
Yindel
AMEL
D8S1179
D21S11
D18S51
D7S381
Mark Sample for Deletion
D3S449
D19S433
THO1
FGA
Mark Sample for Deletion
D22S1045
D5S818
D13S317
D7S820
SE33
Mark Sample for Deletion
D6S1248
DYS1608
D12S391
D2S1338

SRM 2391c Concordance: GlobalFiler

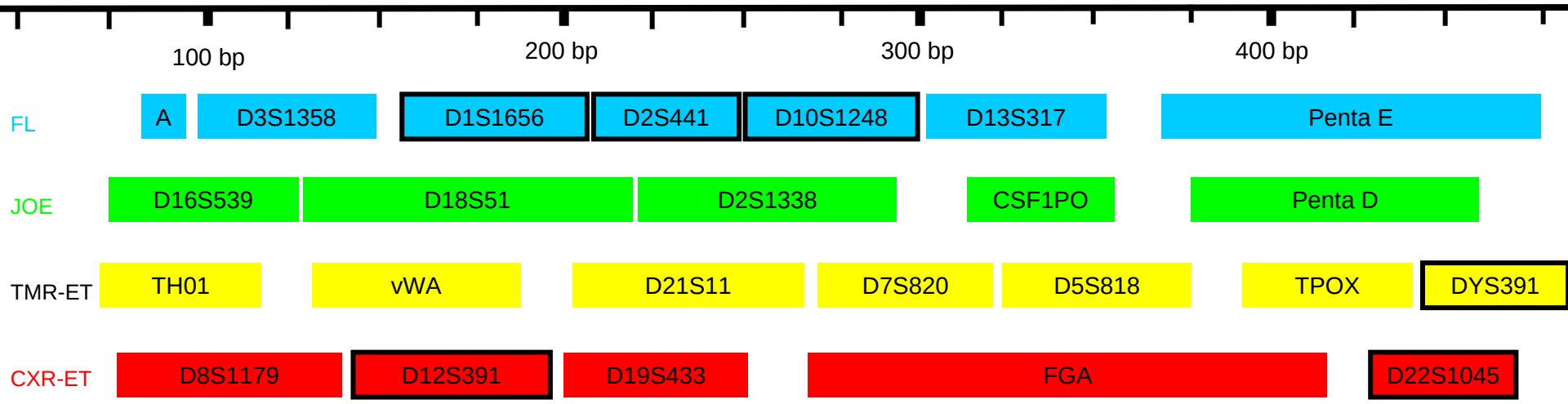
- All SRM 2391c components run with GlobalFiler Casework were **concordant** at all loci
  - Exception: Y indel was not included in the comparison because no other kits use this marker
- Component D at D12S391 shows 1 bp resolution (run on 3500xl):

Correct Type:
(18.3,19,22,23)



Promega PowerPlex **FUSION**

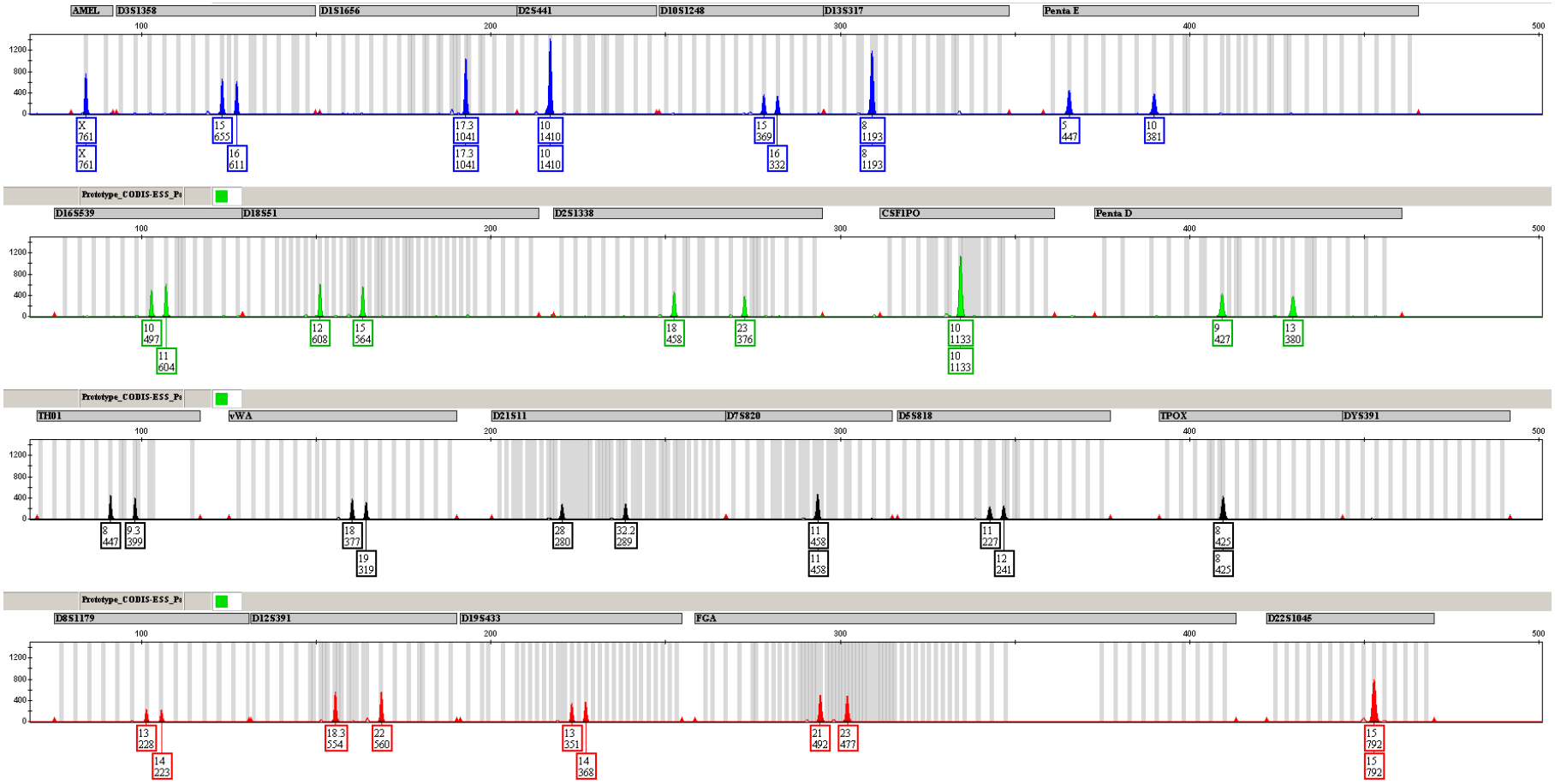
24plex



- 24 STR loci in 5 dyes (3130 and 3500 instrument use)
  - Includes Penta D and E
- Direct amplification and casework capabilities: 85 min amp for both (one kit)
- PowerPlex Fusion gives ~13 orders of magnitude improvement using the NIST 1036 data set

http://www.promega.com/products/pm/genetic-identity/powerplex-fusion/

Component A

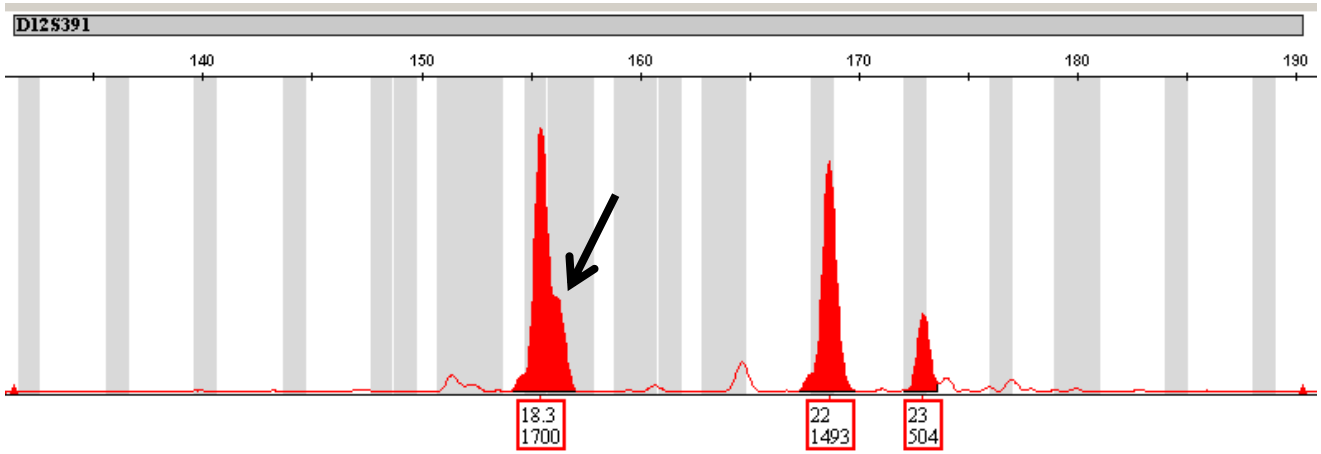

AMEL
D3S1358
D1S1656
D2S441
D10S1248
D13S317
Penta E
Prototype_CODIS-ESS_P
D16S539
D18S51
D2S1338
CSF1PO
Penta D
TH01
vWA
D21S11
D7S820
D5S818
TPOX
DYS391
D8S1179
D12S391
D19S433
FCA
D22S1045

SRM 2391c Concordance: PowerPlex Fusion

- All SRM 2391c components run with PowerPlex Fusion were **concordant** at all loci

- Component D at D12S391 shows lack of resolution between 18.3 and 19 (run on 3130xl):

Correct Type:
(18.3,19,22,23)



Future Directions

- Sequencing of Components A-C, E & F will be completed for all remaining autosomal and Y-STR loci, including non-core loci to raise all reference and informational genotypes to a certified level (Component D is a mixture of Components A and C).
- Once sequencing is complete, the SRM 2391c Certificate of Analysis will be updated with this new information.
- We anticipate the update to occur in the **Fall of 2014**

- This work also supports the high throughput next generation sequencing technologies at NIST for forensic typing applications.
- SRM 2391c has replaced SRM 2395 for Y-STR typing.

Acknowledgments

NIST Funding: Interagency Agreement 2008-DN-R-121 between the National Institute of Justice and NIST Office of Law Enforcement Standards

NIST Disclaimer: 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 the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

NIST Team for This Work

Pete Vallone

Margaret Kline

Contact Info: becky.hill@nist.gov, 301-975-4275