16 Going on 24: Investigator 24plex QS Kit and the New STR Loci

Becky Steffen

National Institute of Standards and Technology

Qiagen Leadership Summit

Washington, DC

June 23, 2015

Presentation Outline

- Meet the new loci
- Investigator 24plex QS Kit
- Concordance study results
- Summary

Meet the New STR Loci

Expanded U.S. Core Loci

Forensic Science International: Genetics 6 (2012) e52–e54

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Letter to the Editor

Expanding the CODIS core loci in the United States

Dear Editor:

After over a decade of operation, the National DNA Index System (NDIS) continues to grow in importance and size [1]. While the STR DNA technology has remained relatively consistent, other key aspects of the NDIS program have been reevaluated and revisions implemented. For example, based upon recommendations of the Scientific Working Group on DNA Analysis Methods, the Director of the Federal Bureau of Investigation (FBI) issued revised Quality Assurance Standards (QAS) for Forensic DNA

major reasons for expanding the CODIS core loci in the United States:

(1) To reduce the likelihood of adventitious matches [7] as the number of profiles stored at NDIS continues to increase each year (expected to total over 10 million profiles by the time of this publication). There are no signs that this trend will slow down as States expand the coverage of their DNA database programs and increase laboratory efficiency and capacity.
(2) To increase international compatibility to assist law enforcement data sharing efforts.
(3) To increase discrimination power to aid missing persons cases.

Hares, D.R. (2012a) Expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 6(1), e52-4.

Forensic Science International: Genetics 6 (2012) e135

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Letter to the Editor

Addendum to expanding the CODIS core loci in the United States

Dear Editor,

An important objective in proposing new CODIS core loci is to ensure that all loci would be available for all potential manufacturers. During the evaluation process, appropriate steps were taken to document access to all proposed core loci. Since publication of the proposed list of core loci, additional information has come to our attention indicating that there may be outstanding issues with respect to some of the proposed loci. Consequently, to ensure the availability for all interested manufacturers in accordance with our stated objective, we are withdrawing Penta D and Penta E as proposed CODIS core loci and recommending the revised listing of core loci in Table 1. Manufacturers are still encouraged to attempt loci in Section B, in ranked order of preference, for inclusion in potential kits provided the impact on the kit's sensitivity and overall performance is negligible. Please

Hares, D.R. (2012b) Addendum to expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 6(5), e135.

New CODIS 20 Core Loci

Table 2
New 20 CODIS Core Loci. (For interpretation of the references to colour in this Table legend, the reader is referred to the web version of this article.)

Locus
CSF1PO
D3S1358
D5S818
D7S820
D8S1179
D13S317
D16S539
D18S51
D21S11
FGA
TH01
TPOX
vWA
D1S1656
D2S441
D2S1338
D10S1248
D12S391
D19S433
D22S1045

Red is for original CODIS Core 13 Loci.
Blue is for new additional CODIS Core Loci.

Forensic Science International: Genetics 17 (2015) 33–34

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Letter to the Editor

Selection and implementation of expanded CODIS core loci in the United States

System that contain the current 13 CODIS Core Loci, thus compelling the continued use of these loci in the core loci set. Review of the validation data did not result in exclusion of any of the proposed additional core loci. Based upon the validation data, it

Implemented by January 1, 2017

The 10 STR Loci Beyond the CODIS 13

STR Locus	Location	Repeat Motif	Allele Range*	# Alleles*
D2S1338	2q35	TGCC/TTCC	10 to 31	40
D19S433	19q12	AAGG/TAGG	5.2 to 20	36
Penta D	21q22.3	AAAGA	1.1 to 19	50
Penta E	15q26.2	AAAGA	5 to 32	53
D1S1656	1q42	TAGA	8 to 20.3	25
D12S391	12p13.2	AGAT/AGAC	13 to 27.2	52
D2S441	2p14	TCTA/TCAA	8 to 17	22
D10S1248	10q26.3	GGAA	7 to 19	13
D22S1045	22q12.3	ATT	7 to 20	14
SE33	6q14	AAAG‡	3 to 49	**178**

5 new European loci

**Allele range and number of observed alleles from Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology; ‡SE33 alleles have complex repeat structure*

NIST 1036 U.S. Population Samples

- 1032 males + 4 females
  - 361 Caucasians (2 female)
  - 342 African Americans (1 female)
  - 236 Hispanics
  - 97 Asians (1 female)

***Unrelated* samples**
All known or potential related individuals (based on autosomal & lineage marker testing) have been removed from the 1036 data set (e.g., only sons were used from father-son samples)

- Anonymous donors with self-identified ancestry
  - Interstate Blood Bank (Memphis, TN) – obtained in 2002
  - Millennium Biotech, Inc. (Ft. Lauderdale, FL) – obtained in 2001
  - DNA Diagnostics Center (Fairfield, OH) – obtained in 2007

- **Complete profiles with 29 autosomal STRs + PowerPlex Y23**
  - **Examined with multiple kits** and in-house primer sets enabling concordance

- Additional DNA results available on subsets of these samples
  - mtDNA control region/whole genome (AFDIL)
  - >100 SNPs (AIMs), 68 InDel markers, X-STRs (AFDIL)
  - NIST assays: miniSTRs, 26plex, >100 Y-STRs, 50 Y-SNPs

Data available on STRBase: **http://www.cstl.nist.gov/strbase/NISTpop.htm**

25 Alleles Reported in the Literature for **D1S1656**

15 NIST observed alleles circled in red

Allele (Repeat #)	Promega ESX 17	Promega ESI 17	ABI NGM	Repeat Structure $[TAGA]_4[TGA]_{0-1}[TAGA]_nTAGG[TG]_5$	Reference
8	133 bp	222 bp	171 bp	$[TAGA]_8[TG]_5$	Phillips *et al.* (2010)
9	137 bp	226 bp	175 bp	$[TAGA]_9[TG]_5$	Phillips *et al.* (2010)
10 (a)	141 bp	230 bp	179 bp	$[TAGA]_{10}[TG]_5$	Lareu *et al.* (1998)
10 (b)	141 bp	230 bp	179 bp	$[TAGA]_{10}TAGG[TG]_5$	Phillips *et al.* (2010)
11	145 bp	234 bp	183 bp	$[TAGA]_{11}[TG]_5$	Lareu *et al.* (1998)
12 (a)	149 bp	238 bp	187 bp	$[TAGA]_{12}[TG]_5$	Lareu *et al.* (1998)
12 (b)	149 bp	238 bp	187 bp	$[TAGA]_{11}TAGG[TG]_5$	Lareu *et al.* (1998)
13 (a)	153 bp	242 bp	191 bp	$[TAGA]_{12}TAGG[TG]_5$	Lareu *et al.* (1998)
13 (b)	153 bp	242 bp	191 bp	$[TAGA]_{13}[TG]_5$	Phillips *et al.* (2010)
13.3	156 bp	245 bp	194 bp	$[TAGA]_1TGA[TAGA]_{11}TAGG[TG]_5$	Phillips *et al.* (2010)
14 (a)	157 bp	246 bp	195 bp	$[TAGA]_{13}TAGG[TG]_5$	Lareu *et al.* (1998)
14 (b)	157 bp	246 bp	195 bp	$[TAGA]_{14}[TG]_5$	Phillips *et al.* (2010)
14.3	160 bp	249 bp	198 bp	$[TAGA]_4TGA[TAGA]_9TAGG[TG]_5$	Phillips *et al.* (2010)
15	161 bp	250 bp	199 bp	$[TAGA]_{14}TAGG[TG]_5$	Lareu *et al.* (1998)
15.3	164 bp	253 bp	202 bp	$[TAGA]_4TGA[TAGA]_{10}TAGG[TG]_5$	Lareu *et al.* (1998)
16	165 bp	254 bp	203 bp	$[TAGA]_{15}TAGG[TG]_5$	Lareu *et al.* (1998)
16.3	168 bp	257 bp	206 bp	$[TAGA]_4TGA[TAGA]_{11}TAGG[TG]_5$	Lareu *et al.* (1998)
17	169 bp	258 bp	207 bp	$[TAGA]_{16}TAGG[TG]_5$	Lareu *et al.* (1998)
17.1	170 bp	259 bp	208 bp	Not published	Schröer *et al.* (2000)
17.3	172 bp	261 bp	210 bp	$[TAGA]_4TGA[TAGA]_{12}TAGG[TG]_5$	Lareu *et al.* (1998)
18	173 bp	262 bp	211 bp	$[TAGA]_{17}TAGG[TG]_5$	Phillips *et al.* (2010)
18.3	176 bp	265 bp	214 bp	$[TAGA]_4TGA[TAGA]_{13}TAGG[TG]_5$	Lareu *et al.* (1998)
19	177 bp	266 bp	215 bp	Not published	Asamura *et al.* (2008)
19.3	180 bp	269 bp	218 bp	$[TAGA]_4TGA[TAGA]_{14}TAGG[TG]_5$	Lareu *et al.* (1998)
20.3	184 bp	273 bp	222 bp	Not published	Gamero *et al.* (2000)

from Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology

NIST **U.S. Population Allele Frequencies**

D1S1656 (15 different alleles)

15 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
10	0.0146	0.0000	0.0028	0.0064
11	0.0453	0.0309	0.0776	0.0275
12	0.0643	0.0464	0.1163	0.0890
13	0.1009	0.1340	0.0665	0.1144
14	**0.2573**	0.0619	0.0789	0.1165
14.3	0.0073	0.0000	0.0028	0.0042
15	0.1579	**0.2784**	**0.1496**	0.1377
15.3	0.0292	0.0000	0.0582	0.0508
16	0.1096	0.2010	0.1357	**0.1758**
16.3	0.1023	0.0155	0.0609	0.0508
17	0.0278	0.0722	0.0471	0.0424
17.3	0.0497	0.0876	0.1330	0.1483
18	0.0029	0.0155	0.0055	0.0064
18.3	0.0234	0.0515	0.0499	0.0254
19.3	0.0073	0.0052	0.0152	0.0042

N=1036

(only unrelated samples used; fathers removed from this sample set)

D1S1656 Characteristics

- **15 alleles** observed
- **93 genotypes** observed
- **>89% heterozygotes** (heterozygosity = 0.8890)
- **0.0224 Probability of Identity** (P_I)

$$P_I = \sum (genotype\ \ frequencies)^2$$

These values have been calculated for all 29 STR loci across the U.S. population samples examined

Loci sorted on Probability of Identity (P_I) values

Locus	Alleles Observed	Genotypes Observed	Het (obs)	P_I Value n=1036
SE33	52	304	0.9353	0.0066
Penta E	23	138	0.8996	0.0147
D2S1338	13	68	0.8793	0.0220
D1S1656	15	93	0.8890	0.0224
D18S51	22	93	0.8687	0.0258
D12S391	24	113	0.8813	0.0271
FGA	27	96	0.8745	0.0308
D6S1043	27	109	0.8494	0.0321
Penta D	16	74	0.8552	0.0382
D21S11	27	86	0.8330	0.0403
D8S1179	11	46	0.7992	0.0558
D19S433	16	78	0.8118	0.0559
vWA	11	39	0.8060	0.0611
F13A01	16	56	0.7809	0.0678
D7S820	11	32	0.7944	0.0726
D16S539	9	28	0.7761	0.0749
D13S317	8	29	0.7674	0.0765
TH01	8	24	0.7471	0.0766
Penta C	12	49	0.7732	0.0769
D2S441	15	43	0.7828	0.0841
D10S1248	12	39	0.7819	0.0845
D3S1358	11	30	0.7519	0.0915
D22S1045	11	44	0.7606	0.0921
F13B	7	20	0.6911	0.0973
CSF1PO	9	31	0.7558	0.1054
D5S818	9	34	0.7297	0.1104
FESFPS	12	36	0.7230	0.1128
LPL	9	27	0.7027	0.1336
TPOX	9	28	0.6902	0.1358

Better for mixtures (more alleles seen)

29 STR Loci present in STR kits

rank ordered by their variability

N=1036

(only unrelated samples used)

There are several loci more polymorphic than the **current CODIS 13 STRs**

361 Caucasians
342 African Americans
236 Hispanics
97 Asians

Better for kinship (low mutation rate)

Probability of Identity Combinations
(assuming unrelated individuals)

STR Kit or Core Set of Loci	Total N=1036	Caucasians (n=361)	African Am. (n=342)	Hispanics (n=236)	Asians (n=97)
CODIS 13	5.02E-16	2.97E-15	1.14E-15	1.36E-15	1.71E-14
Identifiler	6.18E-19	6.87E-18	1.04E-18	2.73E-18	5.31E-17
PowerPlex 16	2.82E-19	4.24E-18	6.09E-19	1.26E-18	2.55E-17
PowerPlex 18D	3.47E-22	9.82E-21	5.60E-22	2.54E-21	7.92E-20
ESS 12	3.04E-16	9.66E-16	9.25E-16	2.60E-15	3.42E-14
ESI 16 / ESX 16 / NGM	2.80E-20	2.20E-19	6.23E-20	4.03E-19	9.83E-18
ESI 17 / ESX 17 / NGM SElect	1.85E-22	1.74E-21	6.71E-22	3.97E-21	1.87E-19
CODIS 20	9.35E-24	7.32E-23	6.12E-23	8.43E-23	4.22E-21
GlobalFiler & 24plex QS	7.73E-28	1.30E-26	3.20E-27	2.27E-26	1.81E-24
PowerPlex Fusion	6.58E-29	2.35E-27	1.59E-28	2.12E-27	1.42E-25
All 29 autosomal STRs	2.24E-37	7.36E-35	3.16E-37	2.93E-35	4.02E-32
29 autoSTRs + DYS391	1.07E-37	3.26E-35	1.77E-37	1.29E-35	2.81E-32

Qiagen 24plex QS gives ~12 orders of magnitude improvement using the NIST 1036 data set

NIST U.S. Population Data

- The data from our 1036 U.S. population samples is currently available on STRBase:

  http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm

- A summary of the NIST 1036 data set has been published in Profiles in DNA for autosomal and YSTR loci

- Population Data announcements have been published in FSI: Genetics for
  - 29 autosomal STR loci (*Hill et al*)
  - 23 Y-STR loci (*Coble et al*)


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

run and population statistics were confirmed using the Power-Marker v3.25 statistics program [10].

Investigator 24plex QS Kit

Qiagen Investigator 24plex QS

Launched in the U.S. on June 17, 2015

There's More to a Sample than a Profile — Global STR Analysis Including Quality Control

Print Bookmark Share

The FBI CODIS Core Loci Working Group have published the recommendation that the CODIS core loci should be expanded, and a combination of STR markers from the Combined DNA Index System (CODIS), the European Network of Forensic Science Institutes (ENFSI), and the European DNA Profiling Group (EDNAP) should be used to improve the accuaracy of forensic testing.

QIAGEN has developed new kits — Investigator 24plex Kits — that coamplify all 23 recommended markers. The kits use novel 6-dye technology to keep the amplicon length of markers short while avoiding overlapping of markers. Kits are available for purified DNA from casework and for reference samples.

This webinar will outline the benefits of the Investigator 24plex Kits and highlight the improved workflow, which saves unnecessary re-runs of a sample by use of an internal "Quality Sensor" control.

This unique Quality Sensor can distinguish:

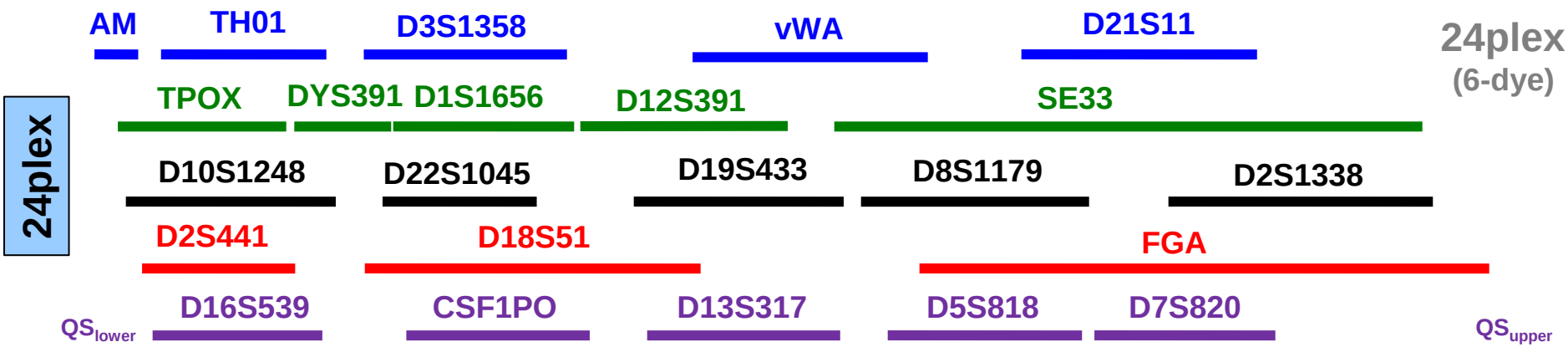
- Successful amplification
- Degraded DNA
- Inhibited DNA
- No DNA
- Failed PCR amplification

Internal "Quality Sensor"

Join us to learn how to improve results and streamline your forensic STR analysis with the new Investigator 24plex Kits.

http://www.qiagen.com/resources/e-learning/webinars/human-identity-and-forensics/multiplex-pcr-global-str/

Qiagen Investigator **24plex QS**



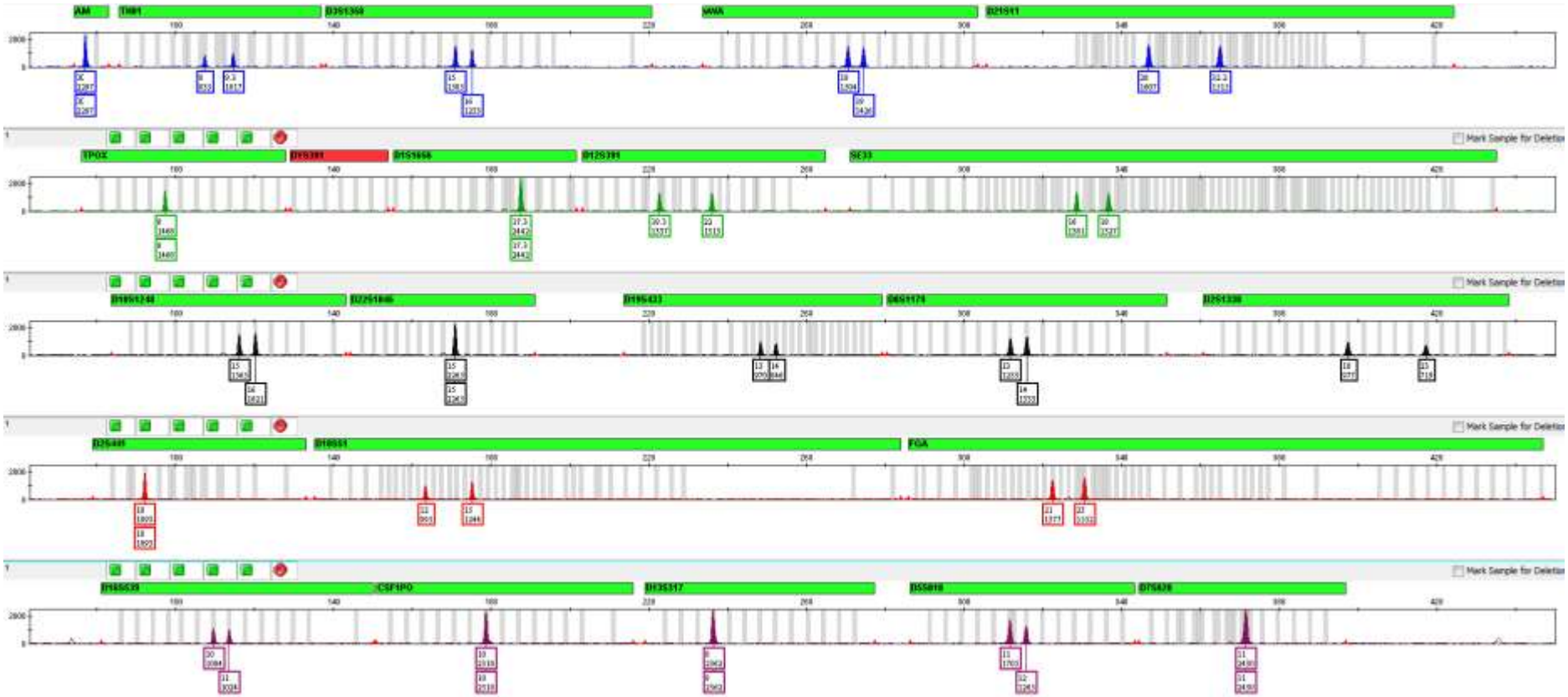
- 22 STR loci plus Amelogenin in 6 dyes (3500 use or 3130 upgrade required)
  - Includes SE33
- Includes Quality Sensors for detecting degraded and/or inhibited DNA
- Uses the BT6 matrix: **6-FAM**, **BTG**, **BTY**, **BTR2**, **BTP** & **BTO** (size standard)
- Size Standard is in **BTO**, 60-550 bp
- Investigator 24plex GO!: direct amplification capabilities
  - Single source samples: 45 min amplification
- Investigator 24plex QS
  - Casework samples: 60 min amplification

Two separate kits

- Qiagen 24plex gives ~12 orders of magnitude improvement using the NIST 1036 data set

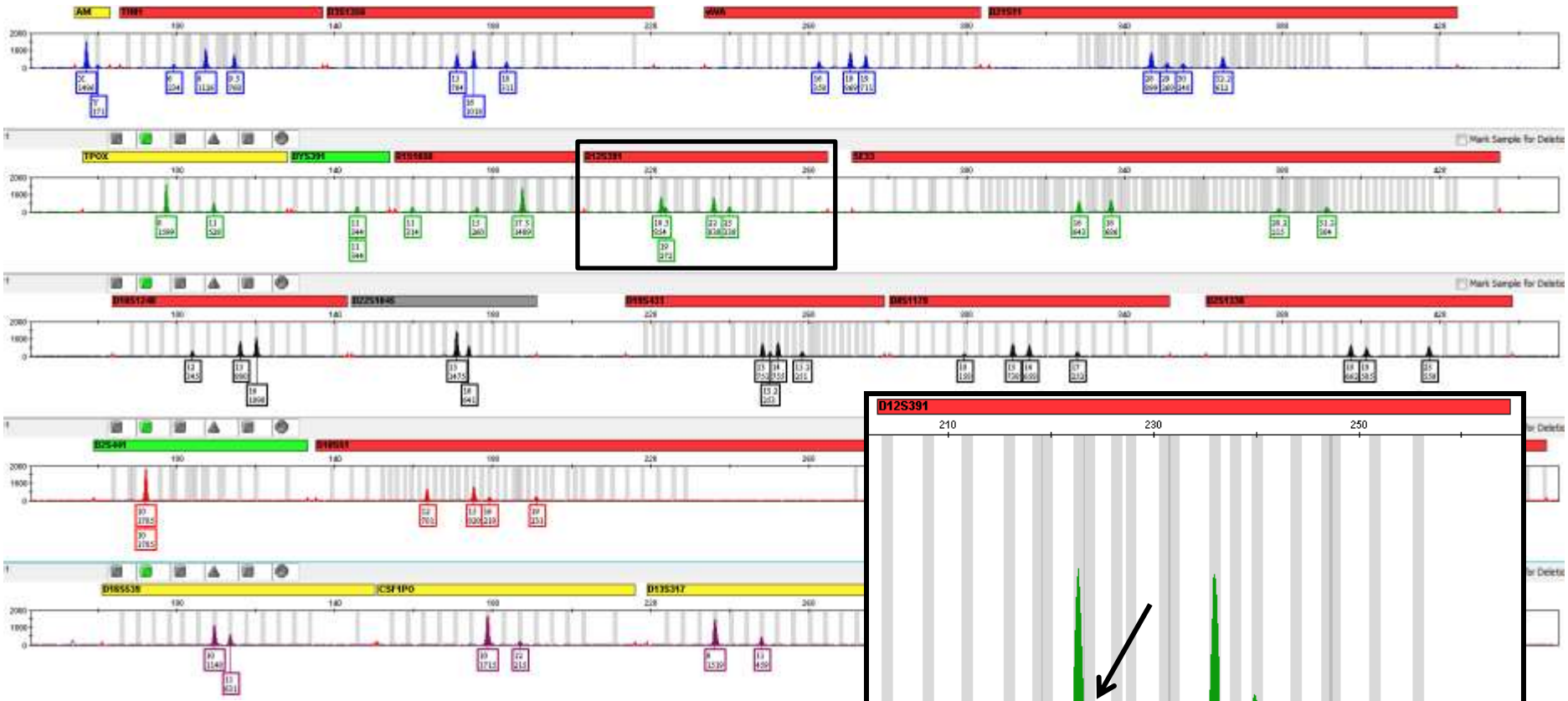
http://www.qiagen.com/resources/e-learning/webinars/human-identity-and-forensics/multiplex-pcr-global-str/

SRM 2391c is **fully concordant** at all loci for 24plex kit – **Component A Profile**



1 ng DNA, 30 cycles, 3500xl

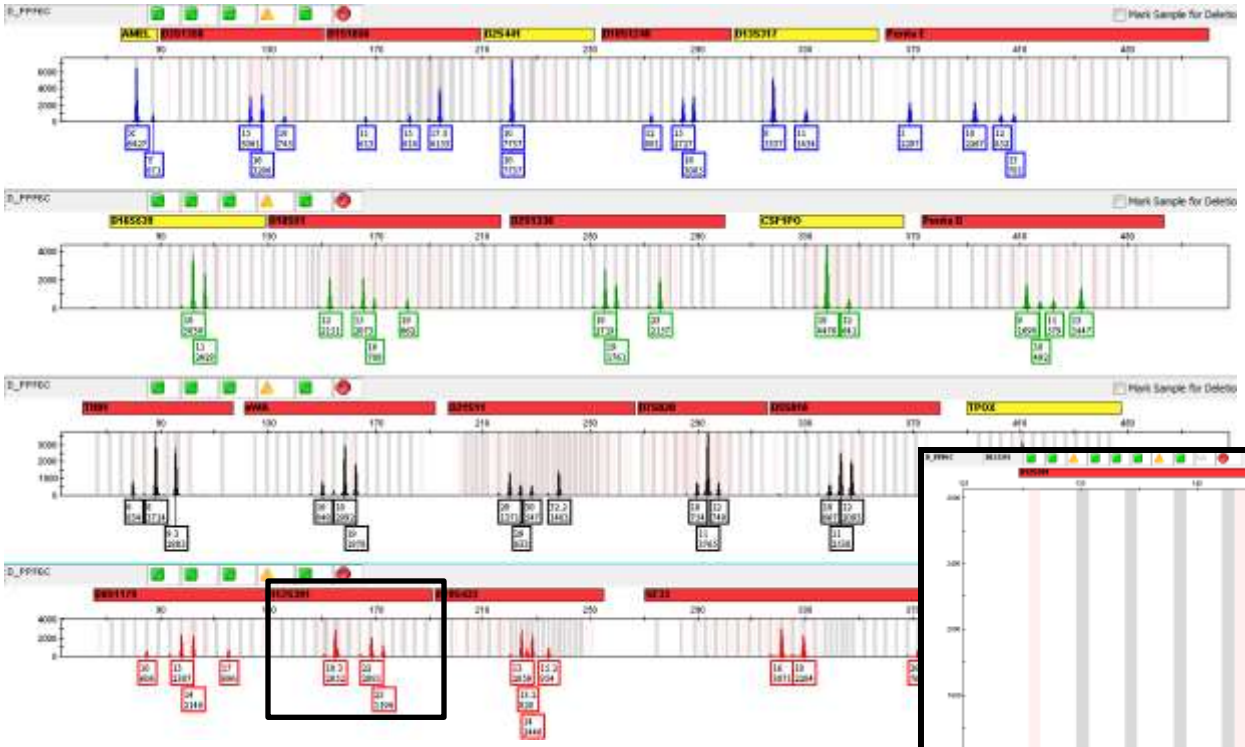
SRM 2391c Mixture Component D

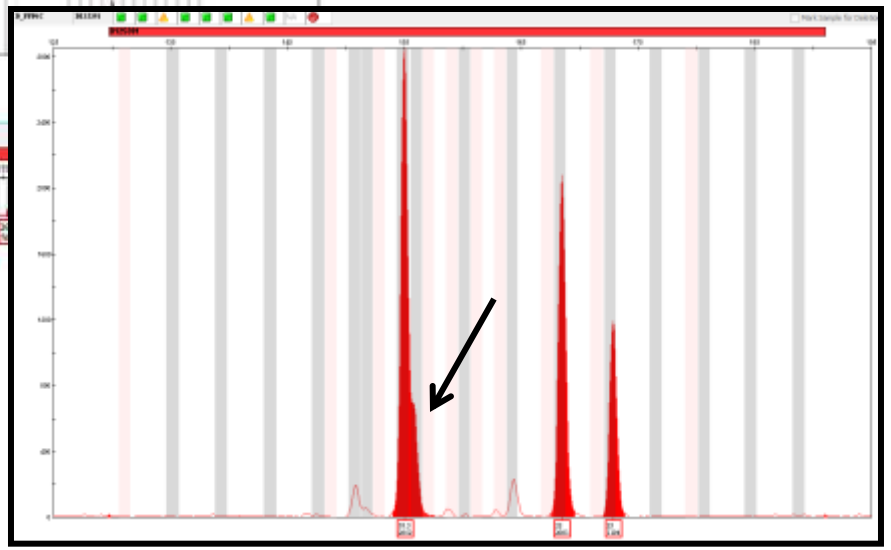


At D12S319, the 18.3 & 19 peaks are fully resolved

1 ng DNA, 30 cycles, 3500xl

Component D Resolution Issue at 12S391 PowerPlex Fusion





At D12S319, the 18.3 & 19 peaks are not fully resolved so the 19 peak is not called

1 ng DNA, 30 cycles

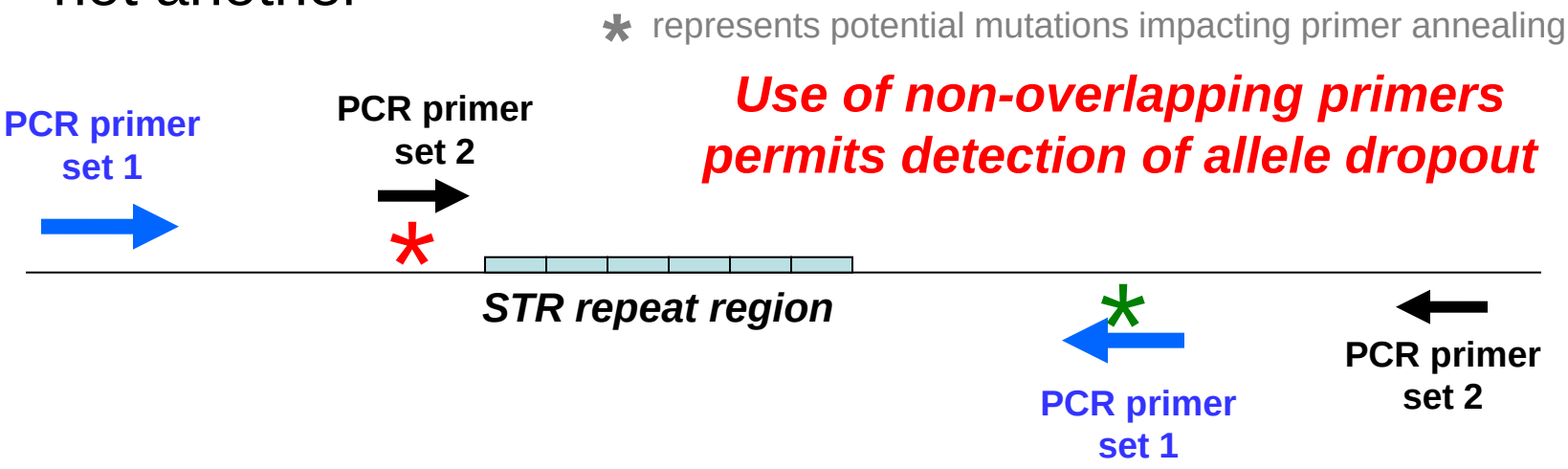
Concordance Testing at NIST

STR Kit Concordance Testing

- Many of these STR kits have different primer sequences for amplifying the same STR locus
- Need to analyze the same DNA samples with different STR typing kits looking for differences
- In some rare cases, allele dropout may occur due to mutations in primer binding regions

Purpose of Concordance Studies

When different primer sets are utilized, there is a concern that allele dropout may occur due to primer binding site mutations that impact one set of primers but not another

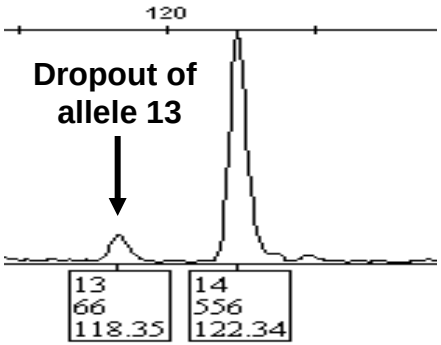


If no primer binding site mutations

If a primer binding site mutation exists

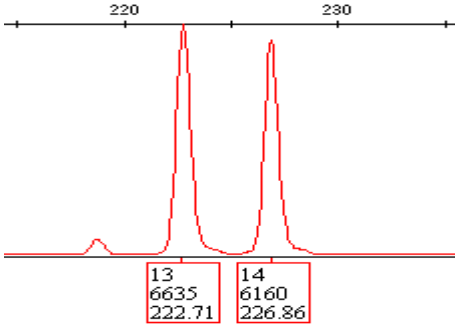
Example Primer Binding Site Mutation that Causes a Null Allele

Identifiler = 14,14



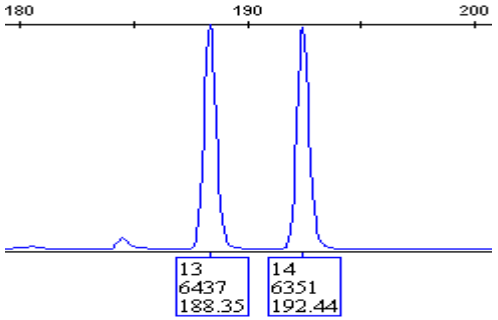
PHR = 11.9%

PP ESX 17 = 13,14



PHR = 92.8%

PP ESI 17 = 13,14



PHR = 98.7%

G → A SNP

D19S433 repeat region

ggaaagaaggtaggaaggaaggaaggaagtaagtaaggaaggaaggaaggaaggaaggaaggagagaggaagaaagagagaagatttttattcgggtaat

X

This region could potentially represent where the reverse primer is located to include the primer binding site mutation

*Applied Biosystems does not publish their primer sequences

Concordance Results for Investigator 24plex QS

Concordance Testing with 24plexQS

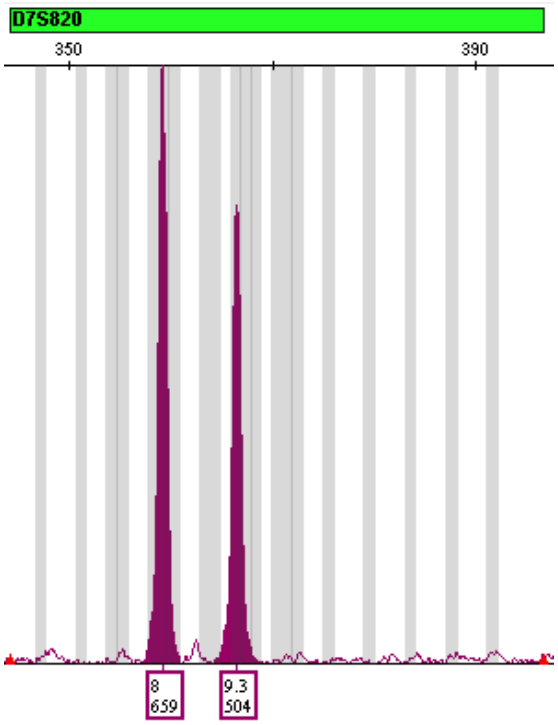
- 24pQS results compared to NIST final data set with **656 unrelated individuals** (NIST U.S. population set (650 samples) and SRM 2391c (6 samples))
- 24pQS is **fully concordant with NIST SRM 2391c** certified values
- 24pQS **null alleles**:
  - None
- 24pQS **discordance**:
  - D7S820 (1/656)

Out of 29,520 alleles compared, 1 discordant call was observed for a 0.003% discordance

99.997% concordance
between 24pQS and NIST final data set

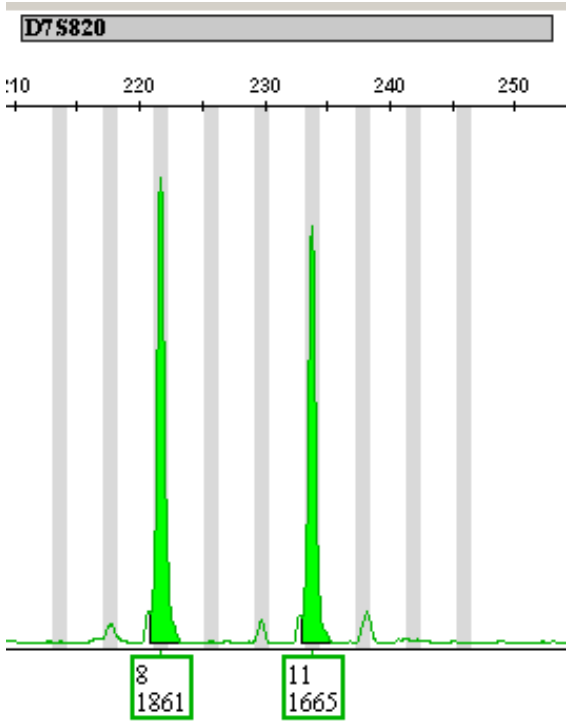
D7S820 Discordant Result

24plex QS



Identifiler, PowerPlex Fusion (8,9.3)

PP16

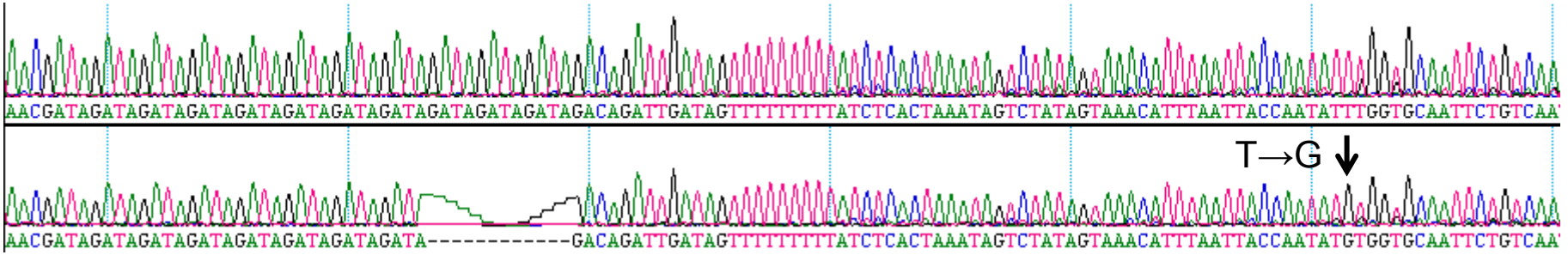


MiniFiler, IDplex (8,11)

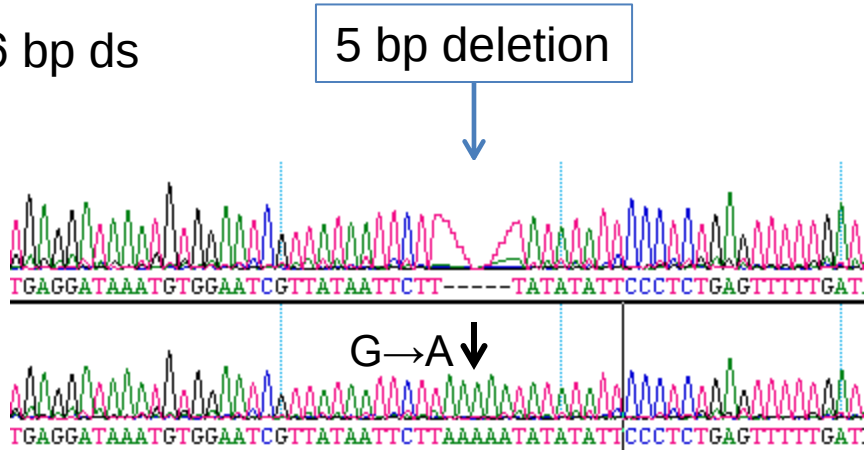
Sequenced: 5 bp deletion 114 bp downstream from the repeat

D7S820 (8,11) or (8,9.3)

Allele 11 or 9.3: $[GATA]_{11}$ 5 bp del 113 bp ds
Deletion occurs between the PP16 and the AB rev primer.



Allele 8: $[GATA]_8$ T→G 65 bp ds, G→A 116 bp ds



Summary

- NIST has a set of 1036 unrelated U.S. population samples that have been used to fully characterize 29 autosomal STR loci available in commercial STR multiplex kits, including all of the required new loci

- Investigator 24plex QS is a 6 dye STR multiplex kit that includes the required new loci for upload to CODIS
  - It gives ~12 orders of magnitude improvement using the NIST 1036 data set

- Investigator 24plex QS performed well in our concordance study with a **99.997%** concordance rate using the NIST population samples

- The CODIS 13 core loci is expanding to 20 core loci and must be implemented by all CODIS labs by January 1, 2017. Using the Investigator 24plex QS kit will satisfy this requirement.

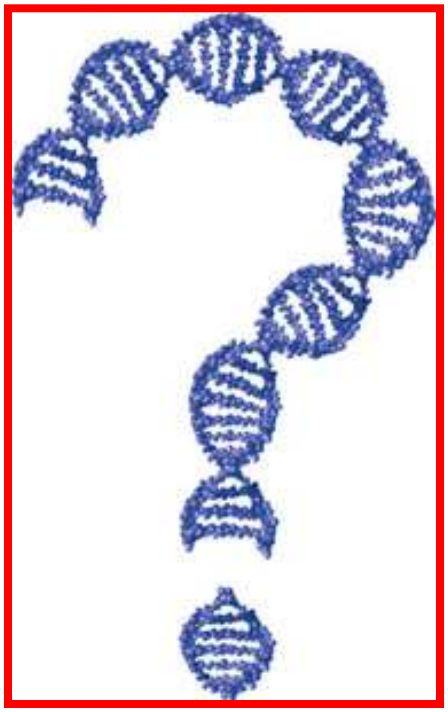
Thank you for your attention!

Our team publications and presentations are available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

http://www.cstl.nist.gov/biotech/strbase

Questions?

becky.steffen@nist.gov
301-975-4275



Acknowledgements

Thanks to **Daniel Mueller** and **Mario Scherer** for supplying all of the kits and for all of their guidance throughout this study.

Funding from the **National Institute of Justice (NIJ)** through NIST Office of Law Enforcement Standards