



ISHI Workshop on New Loci and Kits
October 2, 2014 (Phoenix, AZ)
New Autosomal and Y-STR Loci and Kits:
Making Data Driven Decisions

STRBase Resources & Additional Information

John M. Butler

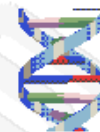
NIST Office of Special Programs

NIST STRBase Website

Serving the Forensic DNA Community for >15 Years



Short Tandem Repeat DNA Internet Database



NIST [Standard Reference Database](#) SRD 130

[\[Recent Updates\]](#)

Serving the forensic DNA and human identity testing communities for over 10 years... These data are intended to benefit research and application of short tandem repeat DNA markers to human identity testing. The authors are solely responsible for the information herein.

Please Rate Our Products and Services: <http://tsapps.nist.gov/MSDSurvey/default.aspx?ID=5&DB=130>

This database has been accessed **458551** times since 10/02/97. (Counter courtesy www.digits.com - see [disclaimer](#).)

Created by [John M. Butler](#)
and [Dennis J. Reeder](#) (*NIST Biochemical Science Division*),
with invaluable help from Jan Redman, Christian Ruitberg and Michael Tung
Site creators' curriculum vitae available using links above.

Partial support for the design and maintenance of this website is being provided by [The National Institute of Justice](#) through the [NIST Office of Law Enforcement Standards](#).

General Information

- [Purpose of STRBase/NAR 2001 Paper describing STRBase/Overview Presentation](#)
- [Publications and Presentations from NIST Human Identity Project Team](#) ◆
- [NIJ-Funded Projects](#) ◆
- [Training Materials](#) ◆
- [Links to other web sites](#) ◆
- [Glossary of commonly used terms](#)

<http://www.cstl.nist.gov/strbase/>

A Brief History of the STRBase Website

- Initial information was collected on STR markers while working on my PhD dissertation in 1993-1995
- Started a review article in 1996 while a NIST postdoc but wanted to create a dynamic rather than an out-of-date resource
- Created hundreds of individual web pages that were hyperlinked together
- **Website launched in July 1997** (discussed at ISHI 1997)
- Became a NIST Standard Reference Database (SRD 130) because of its high visibility
- **I continue to update the website** (via an HTML editor)...
- **I have more information than I have had time to upload** (i.e., there is additional information in development)

Core STRBase Information

Forensic STR Information

- [STRs101: Brief Introduction to STRs](#)
- [Core Loci: FBI CODIS Core STR Loci and European Core Loci](#)
- [STR Fact Sheets \(observed alleles and PCR product sizes\)](#)
- ➔ ○ [Multiplex STR kits](#)
- [Sequence Information \(annotated\)](#)
- [Variant Allele Reports](#) ◆
- [Tri-Allelic Patterns](#) ◆
- [Mutation Rates for Common Loci](#)
- [Published PCR primers](#)
- [Y-chromosome STRs](#) ◆
- [Low-template DNA Information](#)
- [Mixture Interpretation](#)
- [Kinship Analysis](#)
- [miniSTRs \(short amplicons\)](#) ◆
- [Null Alleles](#) - discordance observed between STR kits ◆
- [STR Reference List](#) - *now 3687 references* ◆

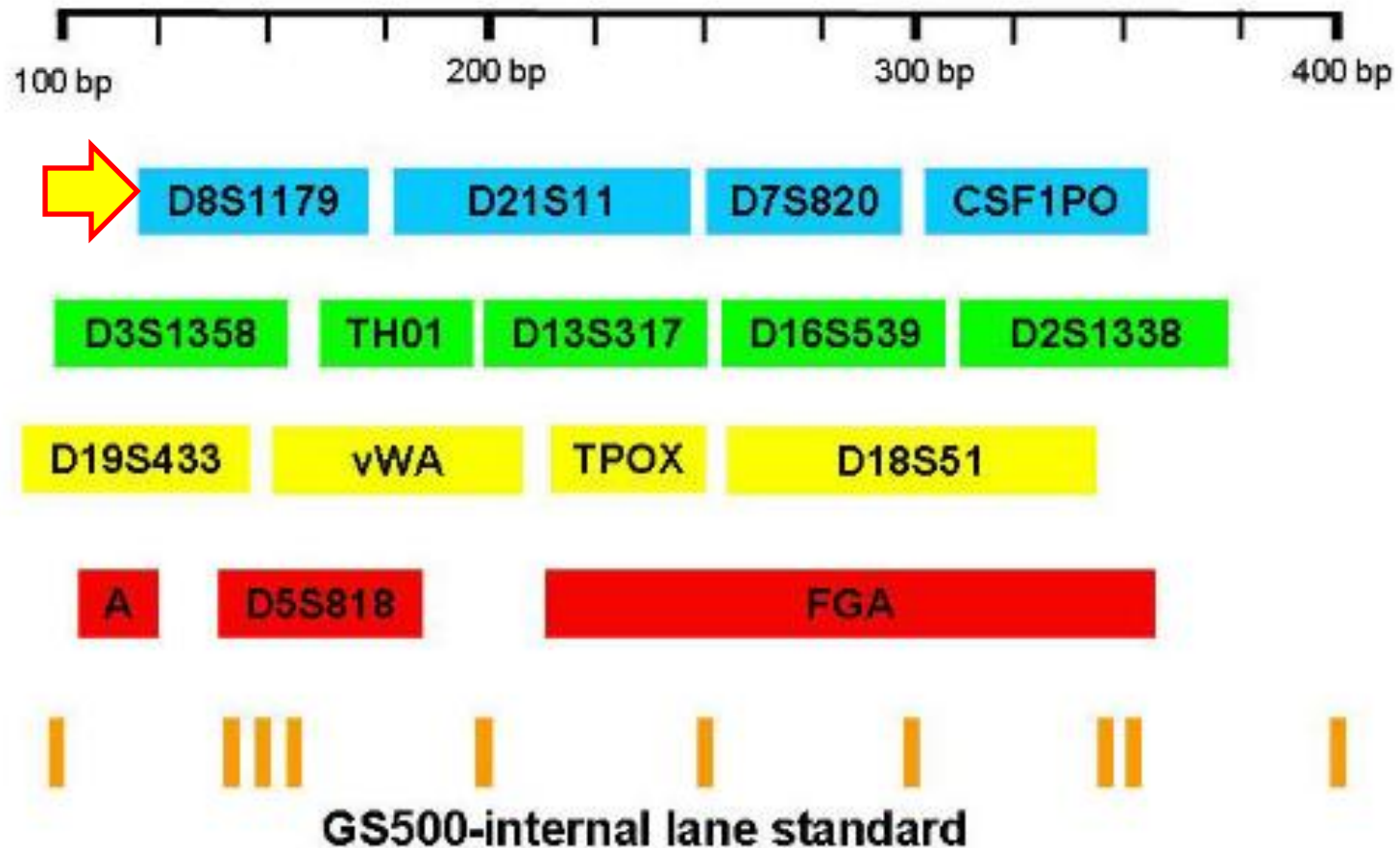
Multiplex STR Kit Information

STR Kits from [Applied Biosystems](#) (Foster City, CA)

- AmpF ℓ STR **Identifiler**: [D8S1179](#), [D21S11](#), [D7S820](#), [CSF1PO](#), [D3S1358](#), [TH01](#), [D13S317](#), [D16S539](#), [D2S1338](#), [D19S433](#), [VWA](#), [TPOX](#), [D1](#)
- AmpF ℓ STR **Identifiler Direct**: [D8S1179](#), [D21S11](#), [D7S820](#), [CSF1PO](#), [D3S1358](#), [TH01](#), [D13S317](#), [D16S539](#), [D2S1338](#), [D19S433](#), [VWA](#), [T](#)
- AmpF ℓ STR **Identifiler Plus**: [D8S1179](#), [D21S11](#), [D7S820](#), [CSF1PO](#), [D3S1358](#), [TH01](#), [D13S317](#), [D16S539](#), [D2S1338](#), [D19S433](#), [VWA](#), [TPC](#)
- AmpF ℓ STR **NGM**: [D10S1248](#), [VWA](#), [D16S539](#), [D2S1338](#), [Amelogenin](#), [D8S1179](#), [D21S11](#), [D18S51](#), [D22S1045](#), [D19S433](#), [TH01](#), [FGA](#), [D](#)
- AmpF ℓ STR **NGM Select**: [D10S1248](#), [VWA](#), [D16S539](#), [D2S1338](#), [Amelogenin](#), [D8S1179](#), [D21S11](#), [D18S51](#), [D22S1045](#), [D19S433](#), [TH01](#)
- AmpF ℓ STR **GlobalFiler**: [D3S1358](#), [VWA](#), [D16S539](#), [CSF1PO](#), [TPOX](#), [Yindel](#), [Amelogenin](#), [D8S1179](#), [D21S11](#), [D18S51](#), [DYS391](#), [D2S441](#), [D2S1338](#)
- AmpF ℓ STR **VeriFiler**: [D10S1248](#), [D1S1656](#), [Amelogenin](#), [D2S1338](#), [D22S1045](#), [D19S433](#), [TH01](#), [D2S441](#), [D6S1043](#), [D12S391](#)
- AmpF ℓ STR **MiniFiler**: [D13S317](#), [D7S820](#), [Amelogenin](#), [D2S1338](#), [D21S11](#), [D16S539](#), [D18S51](#), [CSF1PO](#), [FGA](#)
- AmpF ℓ STR **Yfiler**: [DYS456](#), [DYS389I](#), [DYS390](#), [DYS389II](#), [DYS458](#), [DYS19](#), [DYS385a/b](#), [DYS393](#), [DYS391](#), [DYS439](#), [DYS635](#), [DYS392](#), [Y](#)
- AmpF ℓ STR **SGM Plus**: [D3S1358](#), [VWA](#), [D16S539](#), [D2S1338](#), [Amelogenin](#), [D8S1179](#), [D21S11](#), [D18S51](#), [D19S433](#), [TH01](#), [FGA](#)
- AmpF ℓ STR **Profiler Plus**: [D3S1358](#), [VWA](#), [FGA](#), [Amelogenin](#), [D8S1179](#), [D21S11](#), [D18S51](#), [D5S818](#), [D13S317](#), [D7S820](#)
- AmpF ℓ STR **Profiler Plus ID**: [D3S1358](#), [VWA](#), [FGA](#), [Amelogenin](#), [D8S1179](#), [D21S11](#), [D18S51](#), [D5S818](#), [D13S317](#), [D7S820](#)
- AmpF ℓ STR **COfiler** : [D3S1358](#), [D16S539](#), [Amelogenin](#), [TH01](#), [TPOX](#), [CSF1PO](#), [D7S820](#)
- AmpF ℓ STR **Sinofiler** (*available only in China*): [D8S1179](#), [D21S11](#), [D7S820](#), [CSF1PO](#), [D3S1358](#), [D5S818](#), [D13S317](#), [D16S539](#), [D2S1](#)
- AmpF ℓ STR **Profiler**: [D3S1358](#), [VWA](#), [FGA](#), [Amelogenin](#), [TH01](#), [TPOX](#), [CSF1PO](#), [D5S818](#), [D13S317](#), [D7S820](#)
- AmpF ℓ STR **SEfiler**: [D3S1358](#), [VWA](#), [D16S539](#), [D2S1338](#), [Amelogenin](#), [D8S1179](#), [SE33](#), [D19S433](#), [TH01](#), [FGA](#), [D21S11](#), [D18S51](#)
- AmpF ℓ STR **SEfiler Plus**: [D3S1358](#), [VWA](#), [D16S539](#), [D2S1338](#), [Amelogenin](#), [D8S1179](#), [SE33](#), [D19S433](#), [TH01](#), [FGA](#), [D21S11](#), [D18S51](#)
- AmpF ℓ STR **Green I** : [Amelogenin](#), [TH01](#), [TPOX](#), [CSF1PO](#)
- AmpF ℓ STR **Blue**: [D3S1358](#), [VWA](#), [FGA](#)

<http://www.cstl.nist.gov/strbase/multiplex.htm>

AmpFSTR® Identifiler™



The schematic diagram illustrates the fluorescent dye label color and relative PCR product size ranges for the various STR loci present in this particular kit. [Click on the locus name to learn more about the STR marker of interest.](http://www.cstl.nist.gov/strbase/kits/Identifiler.htm)

<http://www.cstl.nist.gov/strbase/kits/Identifiler.htm>

STR Fact Sheet for D8S1179

D8S1179

Other Names	Chromosomal Location	GenBank Accession
D6S502 UniSTS: 83408	8q24.13 Chr 8; 125.976 Mb (May 2004, NCBI build 35)	G08710 ; has 12 repeat units AF216671 ; has 13 repeat units

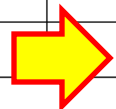
Repeat: [TATC] = GenBank top strand (called [TCTA] by FSS {375})

Reported Primers	Ref.	PCR Primer Sequences
Set 1	369	5' - TTTTGTATTTCATGTGTACATTCG - 3' 5' - CGTAGCTATAATTAGTTCATTTCA - 3'
Set 2	PE ABI	Profiler Plus (JOE labeled), SGM Plus (JOE labeled), Identifiler (6-FAM labeled)
Set 3	Promega	PowerPlex 2.1 (TMR labeled), PowerPlex 16 (TMR labeled) primer sequences 5'-ATTGCAACTTATAATGTAATTTTGTATTTCATG-3' 5'-[TMR]-ACCAAATTGTGTTTCATGATAGTTC-3'

PCR Product Sizes of Observed Alleles

Allele (Repeat #)	Set 1	Set 2	Set 3	Repeat Structure	Ref.
7	157 bp	123 bp	203 bp	[TCTA] ₇	716

19	205 bp	171 bp	251 bp	[TCTA] ₂ [TCTG] ₂ [TCTA] ₁₅	716
20	209 bp	175 bp	255 bp		<i>variant allele</i>



Allelic Ladders: Commercially available from [Promega](#) and [Applied Biosystems](#)

Common Multiplexes: [PowerPlex 2.1](#), [PowerPlex 16](#), [Profiler Plus](#), [SGM Plus](#), [Identifiler](#)

Mutation Rate: 0.14%

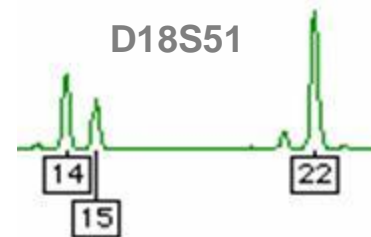
Information on Variant Alleles

- We collect contributions from all over the world where unusual results have been observed with STR data
- Enables laboratories to check if others have seen a specific variant allele or tri-allelic pattern
- Currently (as of July 29, 2014 update)

714 variants at 43 loci

353 tri-allelic patterns at 36 loci

Type 1 tri-allelic pattern



From **D2S1338 Variants Table** (http://www.cstl.nist.gov/strbase/var_D2S1338.htm)

Allele Designation	Allele Size	Instrument	Amp Kit*	Contributor	Verification/Conformation Method(s)	Notes	Frequency
11	290.61	ABI 310	ID	Cintia alves, IPATIMUP, Porto, Portugal	Re-extracted and re-amplified	Portuguese Caucasian sample	1 in 780

http://www.cstl.nist.gov/strbase/var_tab.htm

STR Marker Layouts for New U.S. Kits

100 bp

200 bp

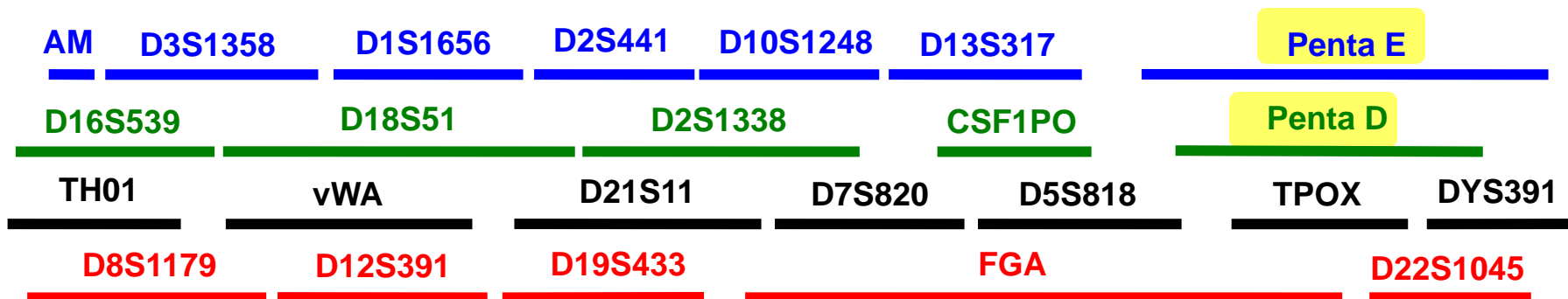
300 bp

400 bp

24plex
(5-dye)

2012

PowerPlex Fusion

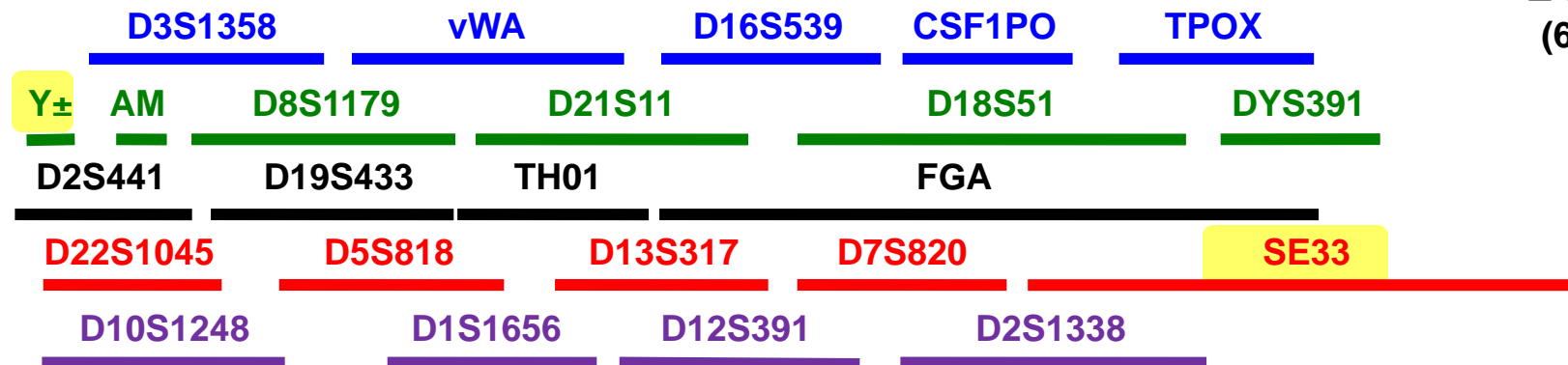


22 core and recommended loci + 2 additional loci

24plex
(6-dye)

2012

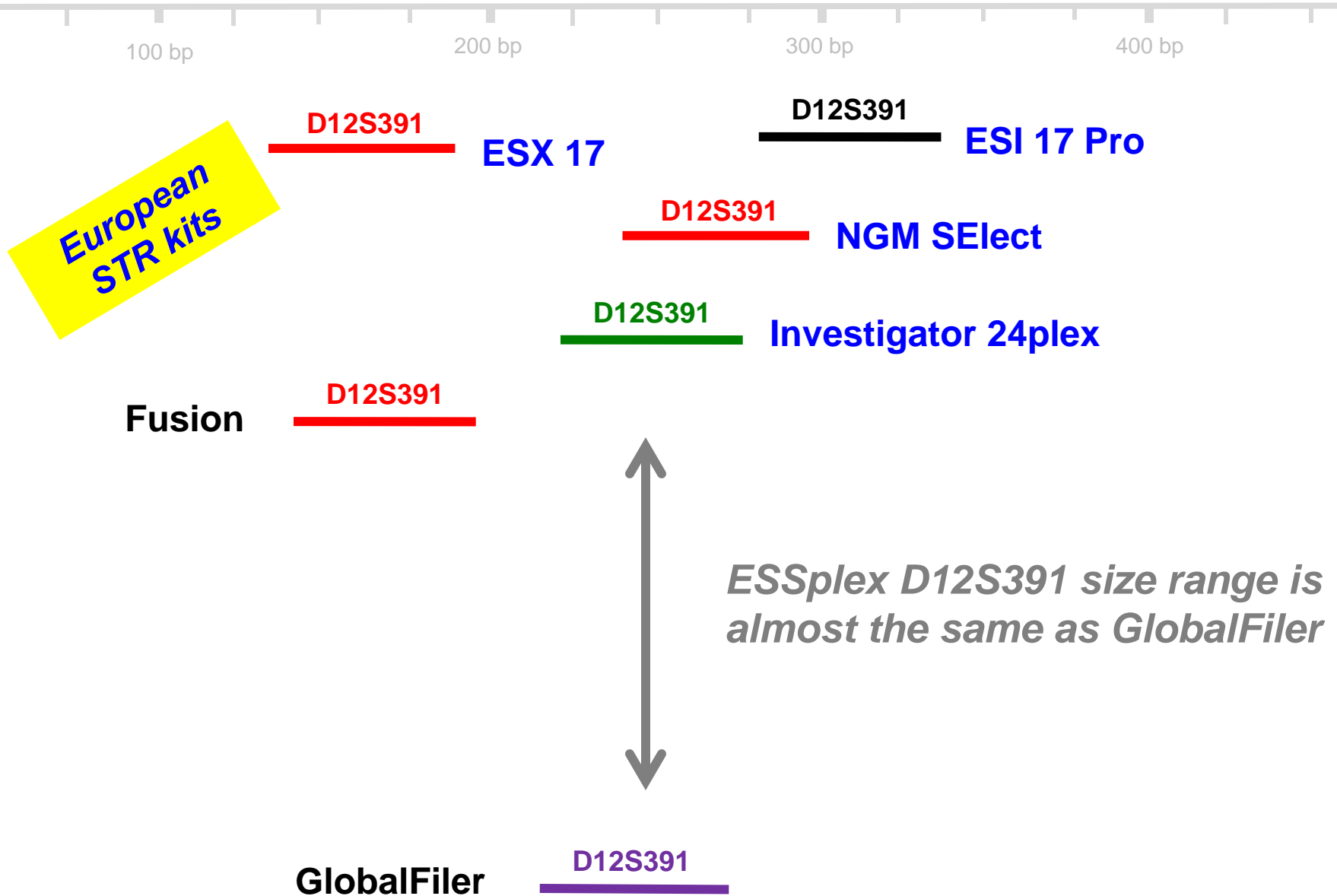
GlobalFiler



Information on “New” STR Loci

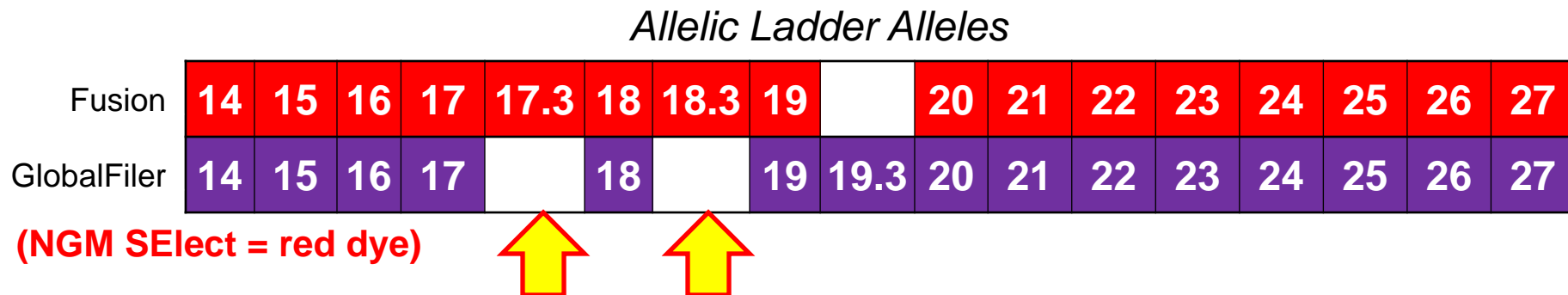
STR Locus	Chromosomal Location	Physical Position	Repeat Motif	Ladder Range
D1S1656	1q42	Chr 1 230.905 Mb	compound TAGA	10 to 19.3
D2S441	2p14	Chr 2 68.239 Mb	compound TCTA/TCAA	8 to 17
SE33	6q14 beta-actin related pseudogene	Chr 6 88.987 Mb	complex AAAG	6.3 to 36
D10S1248	10q26.3	Chr 10 131.093 Mb	simple GGAA	8 to 19
D12S391	12p13.2	Chr 12 12.450 Mb	compound AGAT/AGAC	14 to 27
D22S1045	22q12.3	Chr 22 37.536	simple ATT	8 to 19

Relative Positions of D12S391



STR Locus D12S391 “Variant” Alleles

What is defined as a variant (or off-ladder) allele by a laboratory is typically based on alleles present in STR kit allelic ladder



D12S391 variant alleles (126 total) reported so far in STRBase

(data provided based on 123 NGM SElect, 1 ESI16, 1 NGM, and 1 PP21)

As of Feb 2013	Variant	# times	Variant	# times
	16.1	1x	19.1	1x
	17.1	2x	20.1	2x
	17.3	43x	20.3	2x
	18.1	3x	21.3	1x
	18.3	66x	28	1x

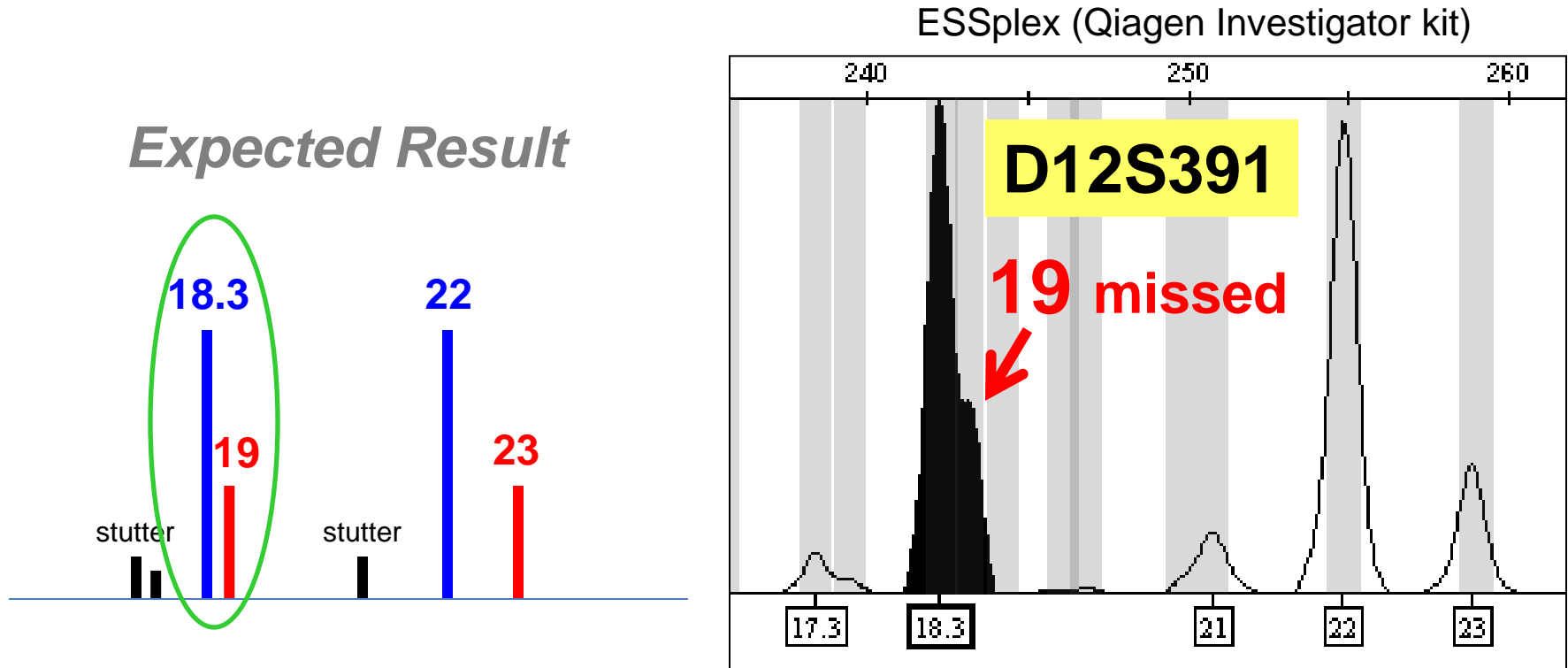
1 tri-allele reported

17,19,20

Sinofiler (China)

NIST SRM 2391c Component D

Provides a Single Base Resolution Challenge



Resolution challenges exist with D12S391 alleles 18.3 and 19, which differ by a single nucleotide; resolution can be impacted by the size of the PCR products in the specific STR kit and electrophoresis conditions (especially run voltage and polymer concentration)

NIST U.S. Population Data

DNA Data [\[Autosomal Markers\]](#) [\[Y-Chromosome Markers\]](#) [\[Mitochondrial DNA\]](#)

NIST 1036 U.S. Population Dataset - 29 autosomal STR loci and 23 Y-STR loci **NEW**

- covers all STR loci present in current commercially available STR kits from Life Technologies and Promega Corporation
- Butler, J.M., Hill, C.R., Coble, M.D. (2012) Variability of new STR loci and kits in U.S. population groups. *Profiles in DNA*. Available at <http://www.promega.com/resources/articles/profiles-in-dna/2012/variability-of-new-str-loci-and-kits-in-us-population-groups/>
- [Data as Excel file](#)

- We have analyzed **1036 unrelated samples** with **29 autosomal STRs and 23 Y-STRs** (all current STR and Y-STR kit loci)
- Becky Hill and Mike Coble have described this data set in their presentations

From NIST 1036 data set (Butler et al. 2012 *Profiles in DNA*)

Allele	Total		Populations, %			
	#	%	AfAm	Asian	Cauc	Hisp
14	1	0.0	0.1			
15	105	5.1	7.7	4.1	3.2	4.4
16	84	4.1	6.7	1.0	2.2	4.2
17	258	12.5	16.7	8.2	12.7	7.6
17.1	3	0.1	0.4			
17.3	26	1.3	0.4		2.1	1.7
18	432	20.8	25.3	26.3	17.2	17.8
18.1	1	0.0	0.1			
18.3	27	1.3	0.4		2.5	1.3
19	314	15.2	14.8	17.5	12.5	18.9
19.1	7	0.3	0.9			0.2
19.3	10	0.5	0.4	0.5	0.4	0.6
20	262	12.6	10.4	19.6	11.1	15.5
20.1	2	0.1	0.3			
20.3	1	0.0				0.2
21	209	10.1	6.4	9.8	12.9	11.2
22	137	6.6	3.7	5.7	9.6	6.8
22.2	1	0.0				0.2
23	102	4.9	2.9	2.6	6.9	5.7
24	53	2.6	1.3	1.0	4.7	1.7
24.3	1	0.0		0.5		
25	24	1.2	0.9	1.5	1.7	0.6
26	7	0.3		1.0	0.3	0.6
27	5	0.2		0.5	0.1	0.6

D12S391

NIST U.S. Allele Frequencies

Theoretical heterozygotes (2pq)

$2 \times 0.013 \times 0.208 = \mathbf{0.54\%}$ (17.3,18)
 $2 \times 0.013 \times 0.152 = \mathbf{0.40\%}$ (18.3, 19)

Observed heterozygotes with a single nucleotide difference

9 out of 1036 = 0.87%

- 17, 17.1**
- 17.3, 18 (3x)**
- 18, 18.1**
- 18.3, 19 (2x)**
- 19, 19.1**
- 19.3, 20**

Comparison of Allelic Ladder Alleles

D1S1656

Fusion	9	10	11	12	13	14	14.3	15	15.3	16	16.3	17	17.3	18	18.3	19	19.3	20.3
GlobalFiler	9	10	11	12	13	14	14.3	15	15.3	16	16.3	17	17.3		18.3		19.3	20.3

D8S1179

Fusion			7	8	9	10	11	12	13	14	15	16	17	18	19
GlobalFiler	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

D16S539

Fusion	4	5	6	7	8	9	10	11	12	13	14	15	16
GlobalFiler		5			8	9	10	11	12	13	14	15	

DYS391

Fusion	5	6	7	8	9	10	11	12	13	14		16
GlobalFiler			7	8	9	10	11	12	13			

More Allelic Ladder Allele Comparisons

D2S441

Fusion	8	9	10	11	11.3	12	13	14	15	16	17
GlobalFiler	8	9	10	11	11.3	12	13	14	15	16	17

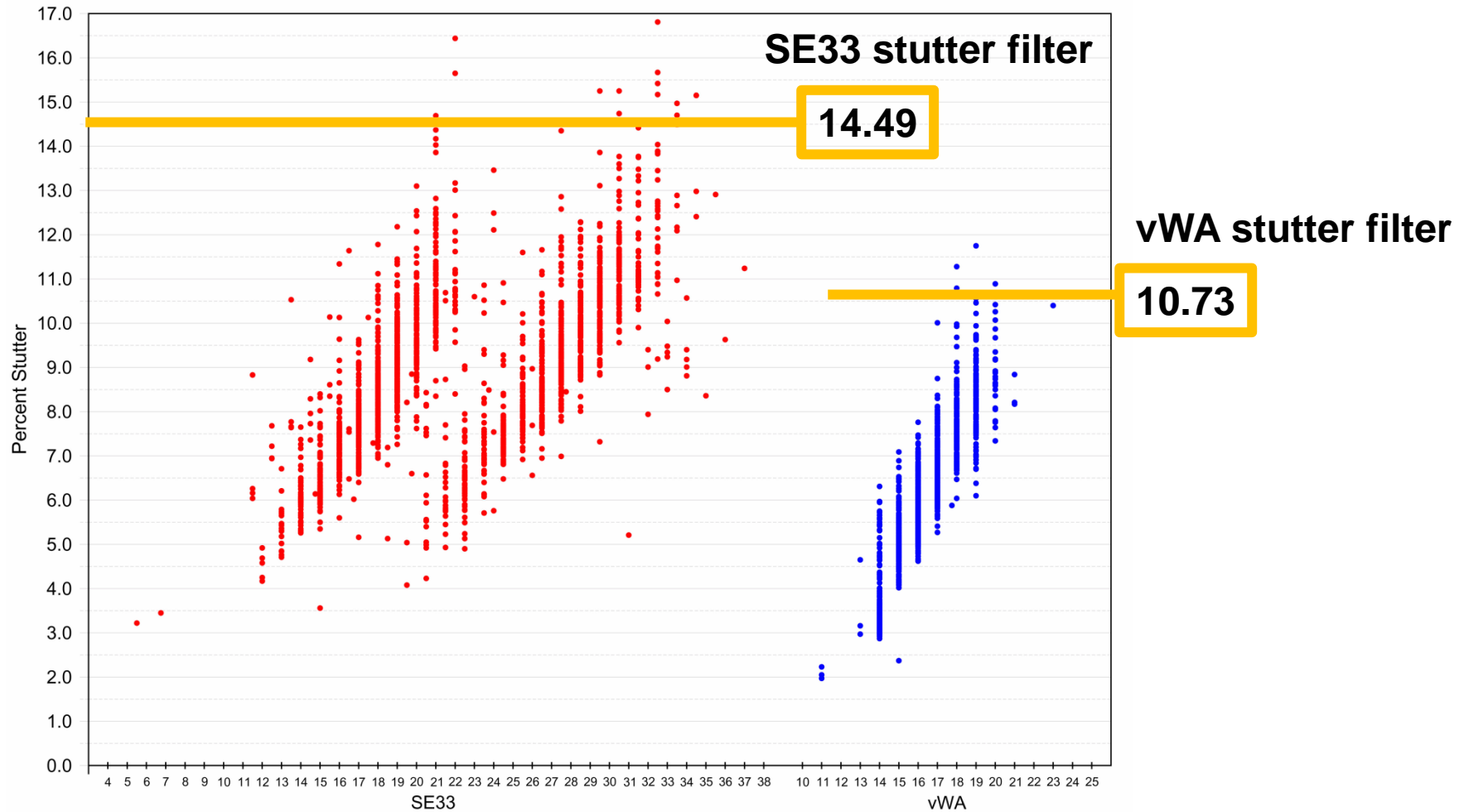
D10S1248

Fusion	8	9	10	11	12	13	14	15	16	17	18	19
GlobalFiler	8	9	10	11	12	13	14	15	16	17	18	19


D22S1045

Fusion	7	8	9	10	11	12	13	14	15	16	17	18	19	20
GlobalFiler		8	9	10	11	12	13	14	15	16	17	18	19	

Stutter Observations vs Stutter Filters



Recent NIST Publications on “New” STR Loci

1. Hill, C.R., Duewer, D.L., Kline, M.C., Sprecher, C.J., McLaren, R.S., Rabbach, D.R., Krenke, B.E., Ensenberger, M.G., Fulmer, P.M., Stort, D.R., Butler, J.M. (2011) Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Sci. Int. Genet.* 5(4): 269-275.
2. O'Connor, K.L., Hill, C.R., Vallone, P.M., Butler, J.M. (2011) Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples. *Forensic Sci. Int. Genet.* 5(5): 538-540.
3. Butler, J.M., Hill, C.R., Kline, M.C., Bastisch, I., Weirich, V., McLaren, R.S., Storts, D.R. (2011) SE33 variant alleles: sequences and implications. *Forensic Sci. Int. Genet.: Suppl. Ser.* 3: e502-e503.
4. Butler, J.M. and Hill, C.R. (2012) Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *Forensic Sci. Rev.* 24(1): 15-26.
5.  Butler, J.M., Hill, C.R., Coble, M.D. (2012) Variability of new STR loci and kits in U.S. population groups. *Profiles in DNA*. Available at <http://www.promega.com/resources/articles/profiles-in-dna/2012/variability-of-new-str-loci-and-kits-in-us-population-groups/>.
6. Coble, M.D., Hill, C.R., Butler J.M. (2013) Haplotype data for 23 Y-chromosome markers in four U.S. population groups. *Forensic Sci. Int. Genet.* 7: e66-e68.
7. Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. population data for 29 autosomal STR loci. *Forensic Sci. Int. Genet.* 7: e82-e83.
8. Butler, J.M. and Hill, C.R. (2013) Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Chapter 9 in Shewale, J. (ed.), *Forensic DNA Analysis: Current Practices and Emerging Technologies*. Taylor & Francis/CRC Press: Boca Raton. pp. 181-198.

Additional Information Needed/Planned

- **Mutation rate information** to aid kinship analysis
 - More father/son studies are needed with D12S391, D1S1656, D2S441, D10S1248, and D22S1045
- A complete summary of **flanking region variation** and null alleles produced from primer binding site mutations
- Future plans for STRBase: listing of **full sequences for detected STR alleles** (repeats and flanking regions) to aid next-generation sequencing efforts
 - Will enable nomenclature and classification of sub-allele variation for STR markers

Summary

- The U.S. forensic DNA community will soon expand to additional STR loci and new kits are now available to help with this effort
- New information is being developed at NIST, published, and added to STRBase to aid understanding of these additional STR loci
- STRBase can be a model for other forensic disciplines in sharing information with the forensic science community

Acknowledgments

\$ National Institute of Justice, FBI, and NIST

Becky Hill and Mike Coble (NIST Applied Genetics)
Jeff Nye (Michigan State Police)
Hope Olson (North Dakota Office of Attorney General)
Jason Kokoszka & Angelo Della Manna (Alabama DFS)



Contact info:

john.butler@nist.gov

+1-301-975-4049

Final version of this presentation available at:

<http://www.cstl.nist.gov/strbase/NISTpub.htm>