NIST Applied Genetics

American Academy of Forensic Sciences
Washington, DC
February 21, 2013

NIST STRBase Resources
to Aid Work with New STR Kits and Loci

John M. Butler, Ph.D.
NIST Applied Genetics Group

NIST Applied Genetics

Product Disclaimer for AAFS

- **I will mention commercial STR kit names and information, but I am in no way attempting to endorse any specific products.**

- SRM 2391c is a Standard Reference Material sold by NIST for measurement calibration purposes.

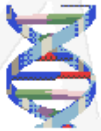
- **NIST Disclaimer**: Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

- **Points of view are mine** and do not necessarily represent the official position of the National Institute of Standards and Technology or the U.S. Department of Justice. **Our group receives or has received funding from the FBI Laboratory and the National Institute of Justice.**

NIST
Applied
Genetics

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 vitaes 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)

NIST Applied Genetics

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***

NIST Applied Genetics

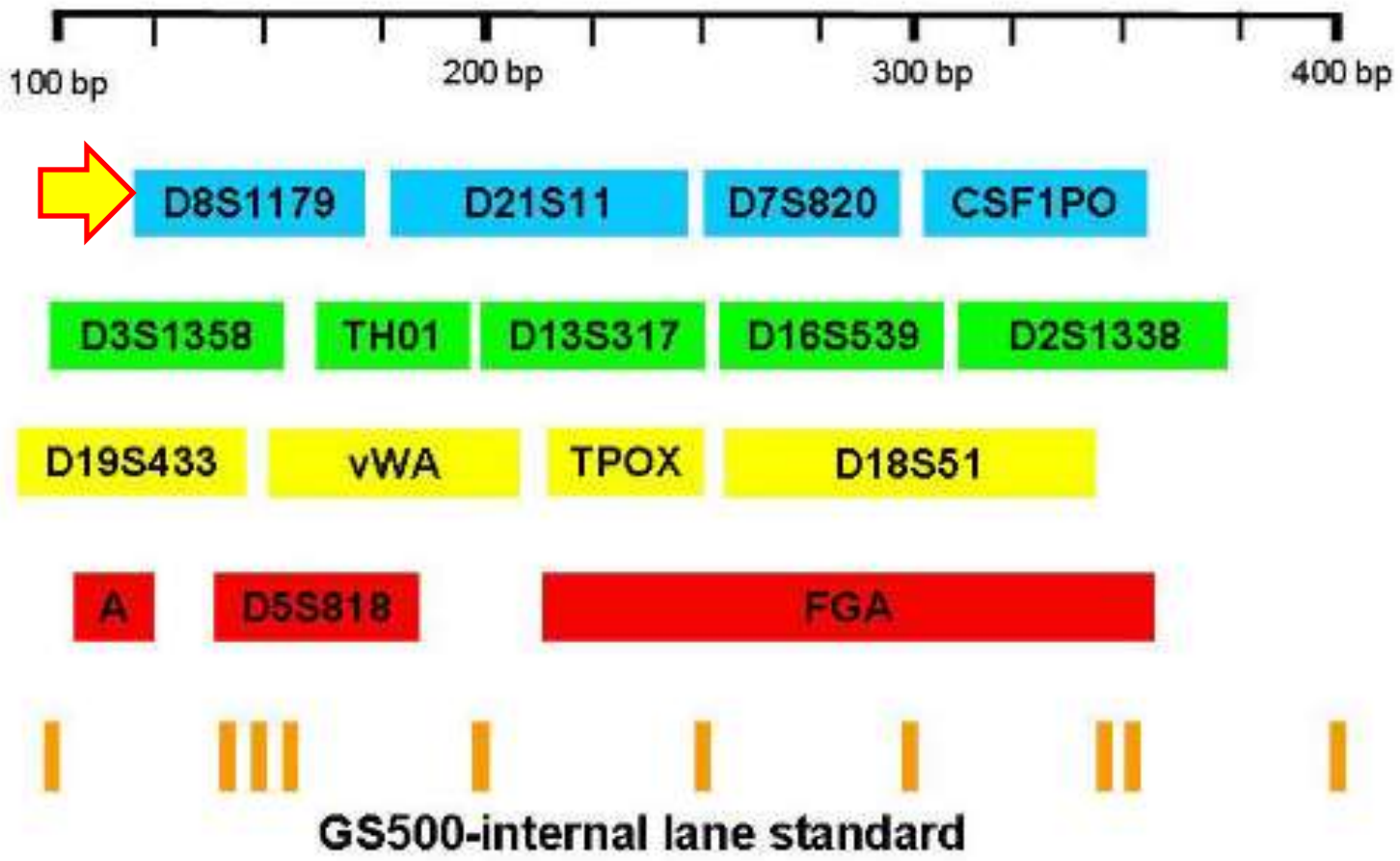
Multiplex STR Kit Information

STR Kits from Applied Biosystems (Foster City, CA)

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

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

AmpFℓSTR® 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.*

NIST Applied Genetics

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' - TTTTTGTATTTCATGTGTACATTCG - 3' 5' - CGTAGCTATAATTAGTTCATTTTCA - 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'-**ATT**GCAACTTATATGTATTTTTGTATTTCATG-3' 5'-[TMR]-ACCAAATTGTGTTCATGAGTATAGTTTC-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		*variant allele*

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%**

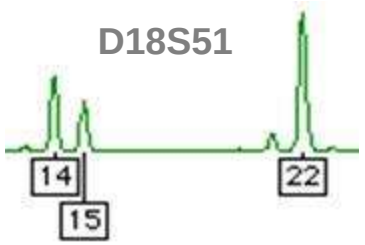
Information on Variant Alleles

- 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 Nov 30, 2012 update)

  661 variants at 41 loci

  329 tri-allelic patterns at 33 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

Additional U.S. Core CODIS Loci Are Coming…

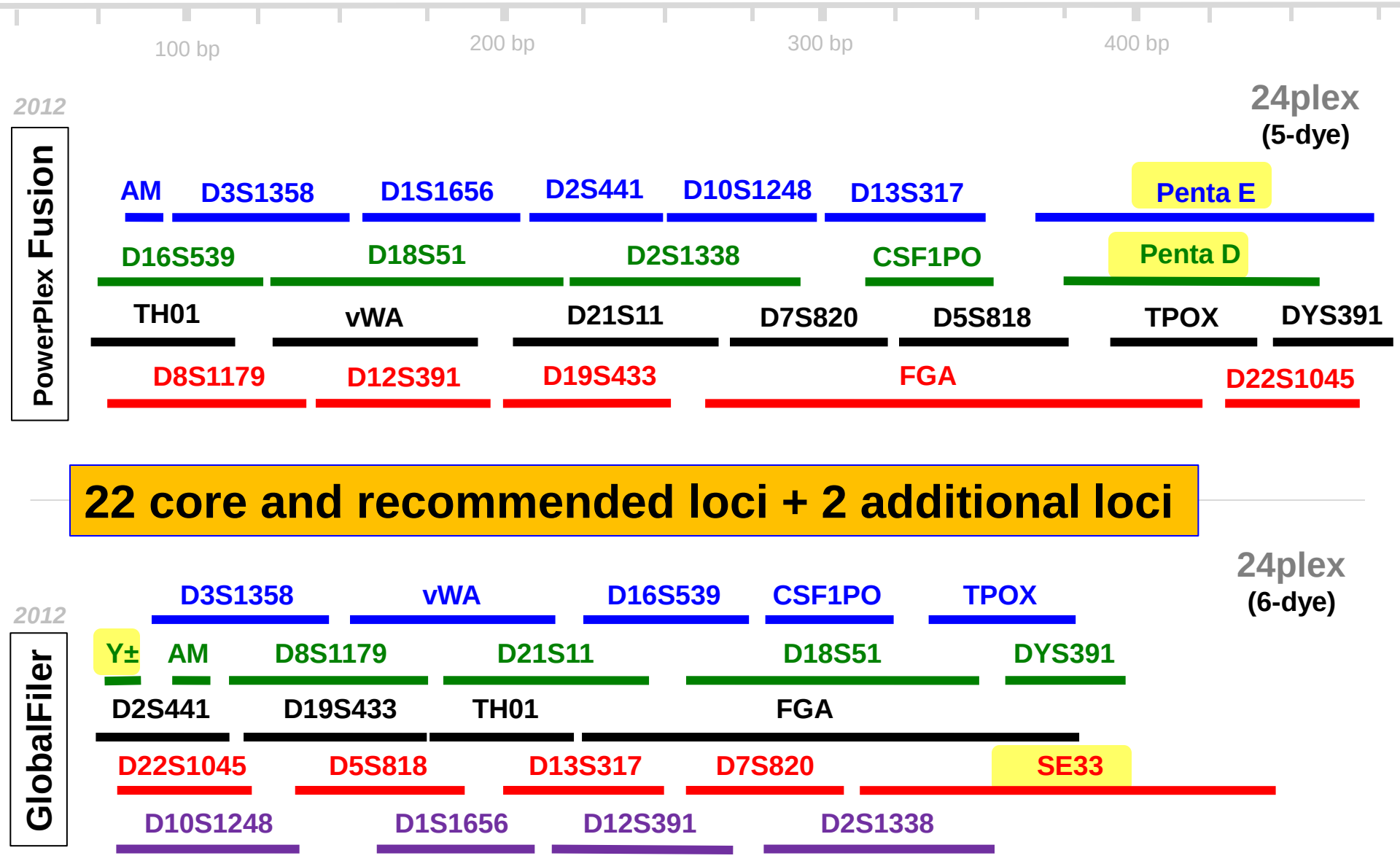
D.R. Hares (2012) Expanding the CODIS Core Loci in the United States. *Forensic Sci. Int. Genet.* 6(1): e52-e54
Addendum to expanding the CODIS core loci in the United States, *Forensic Sci. Int. Genet.* 6(5): e135

What	Why	Who/How	When
Form a Working Group (WG) to discuss initial selection	Establishes target goals	CODIS Core Loci Working Group with FBI Chair and 5 members; Web meetings	May 2010 - present
Announce proposed additional CODIS core loci	Sets desired target goals and informs manufacturers	WG Chair; Publish proposed listing of CODIS core loci	April 2011 online (published Jan 2012)
Ongoing Progress Reports	Provides updates for DNA community	WG Chair; Present updates on status of CODIS Core Loci project at meetings	2010-2012
Implementation Considerations & Strategy	Identify issues for implementation and timeline	WG	June 2011 - present
Manufacturers develop prototype kits	Creates tools to meet target goals	Manufacturers; Provide status reports to WG for timeline	2011-2012
Test and validate prototype kits	Examines if target goals can be met	Validation Laboratories; Follow QAS compliant validation plan	Beginning in 2012
Review and evaluate data from validation	Evaluates if desired performance is obtained	NIST, SWGDAM and FBI; Provide feedback, if any, to Manufacturers	In conjunction with and at the conclusion of validation
Selection of new CODIS core loci	Allows protocols to be established	FBI; seek input from DNA community and stakeholders; Notify Congress	After evaluation of validation data and kit production factors
Implementation of new CODIS core loci at the National DNA Index System	Enables target goals to be met	All NDIS-participating labs	~ 24 months after selection of new CODIS core loci

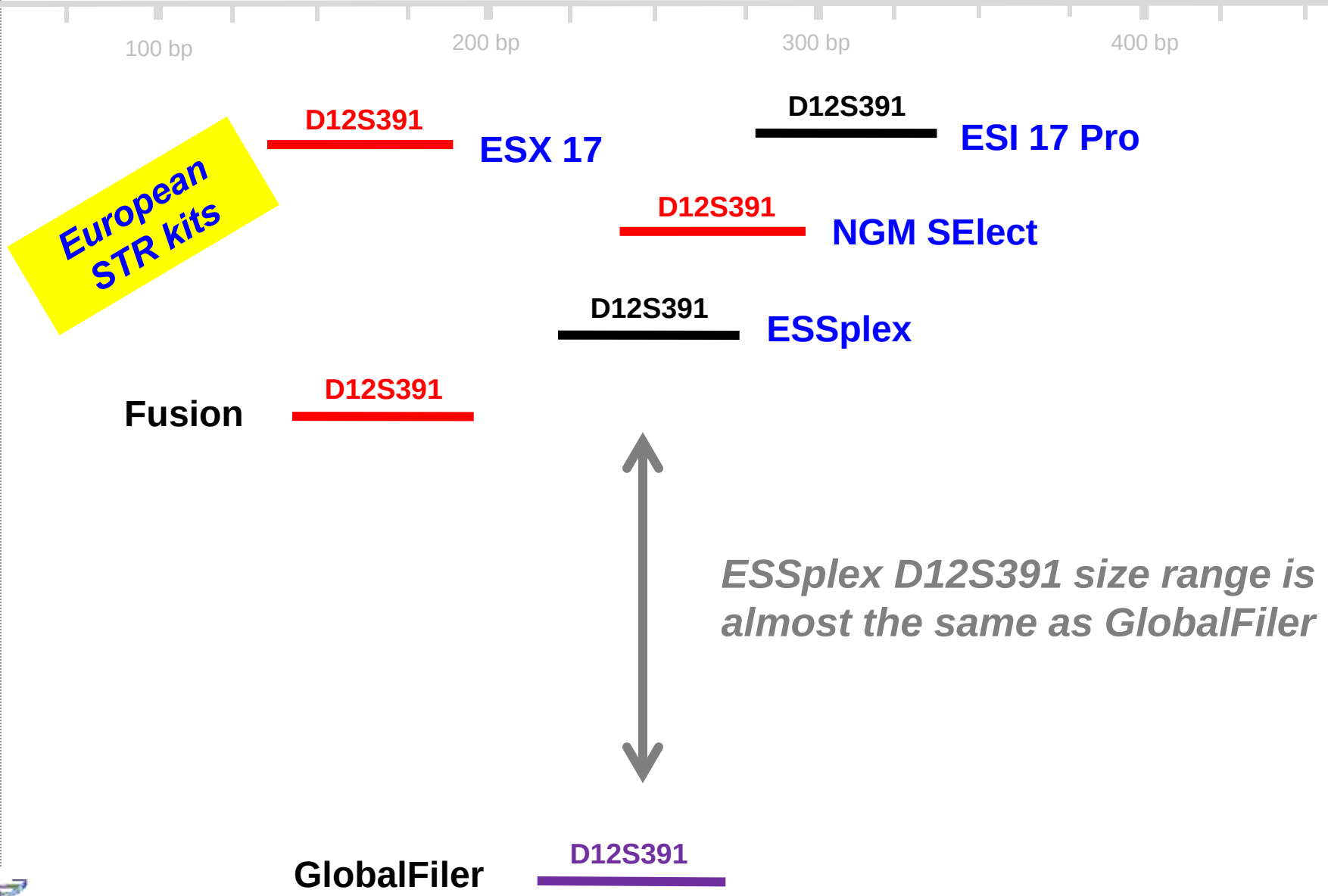
http://www.fbi.gov/about-us/lab/codis/planned-process-and-timeline-for-implementation-of-additional-codis-core-loci

NIST Applied Genetics

STR Marker Layouts for New U.S. Kits



Relative Positions of D12S391



NIST Applied Genetics

Information on "New" STR Loci

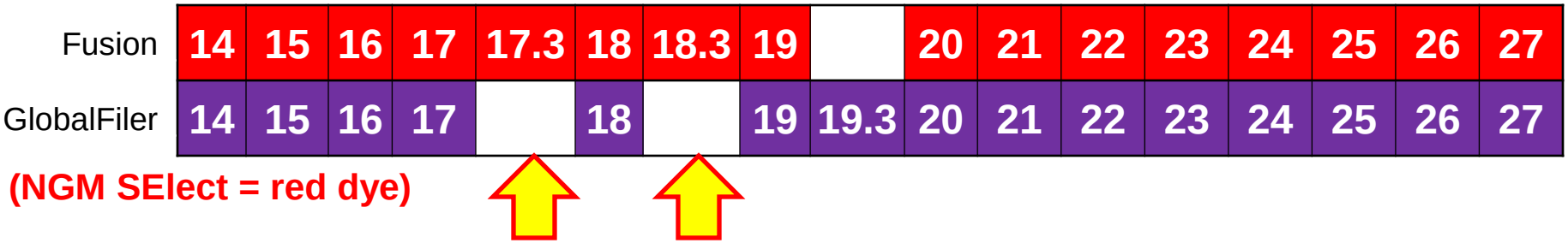
STR Locus	Chromosomal Location	Physical Position	Repeat Motif	Ladder Range
D1S1656	1q42	Chr 1 230.905 Mb	compound TAGA	10 to 19.3
D2S441	2p14	Chr 2 68.239 Mb	compound TCTA/TCAA	8 to 17
SE33	6q14 beta-actin related pseudogene	Chr 6 88.987 Mb	complex AAAG	6.3 to 36
D10S1248	10q26.3	Chr 10 131.093 Mb	simple 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

Butler, J.M., & Hill, C.R. (2013) *Topics on Forensic DNA Analysis: Current Practices and Emerging Technologies* (CRC Press). Chapter 9. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis (in press)

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

Allelic Ladder Alleles



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

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

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
Applied
Genetics

NIST SRM 2391c

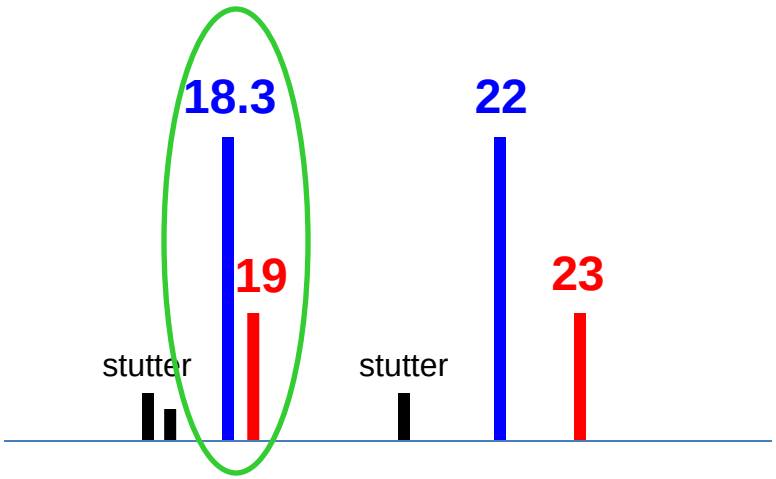
Component D is a mixture

- **Contains certified values** for **29 autosomal STR loci** and **17 Y-STR loci** available in commercial kits (plus some additional reference values for miniSTRs)
- In 2013, **we plan to add certified values for the six additional Y-STR loci in PowerPlex Y23** and any new loci included in Life Technologies Y-STR kit

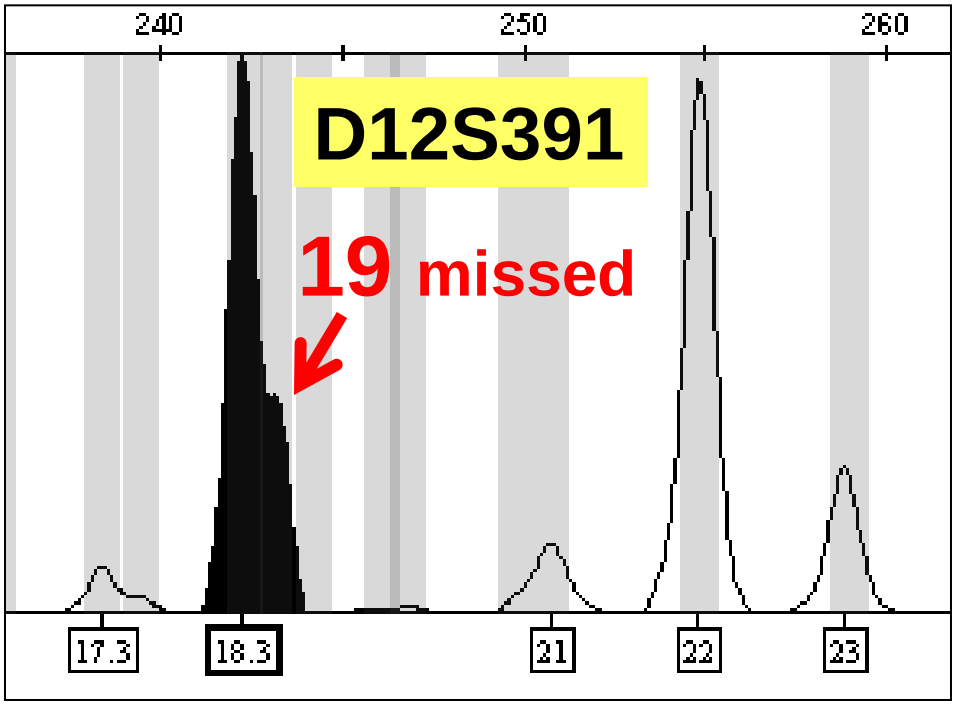
NIST SRM 2391c Component D

Provides a Single Base Resolution Challenge

Expected Result



ESSplex (Qiagen Investigator kit)



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 will describe this data set in more detail in their AAFS presentations

NIST Applied Genetics

D12S391

NIST U.S. Allele Frequencies

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

	Total		Populations, %			
Allele	#	%	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

Theoretical heterozygotes (2pq)

2 x 0.013 x 0.208 = **0.54%** (17.3,18)
2 x 0.013 x 0.152 = **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

Hill, C.R., et al. (2013) U.S. population data for 29 autosomal STR loci. *Forensic Sci. Int. Genet. (in press).*

NIST Applied Genetics

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. (2012) *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Academic Press: San Diego. [pp. 120-121]

5. 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.

6. 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/.

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. (in press)*.

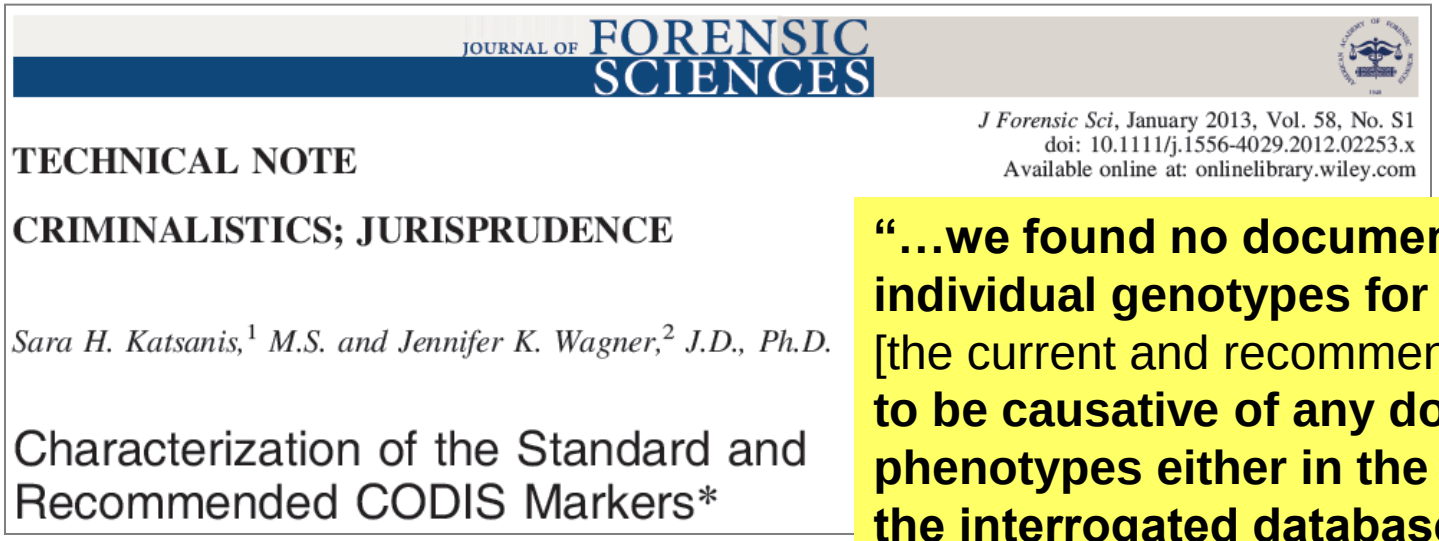
8. Butler, J.M., & Hill, C.R. (2013) *Topics on Forensic DNA Analysis: Current Practices and Emerging Technologies* (CRC Press). Chapter 9. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis *(in press).*

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

NIST Applied Genetics

Forensic STR loci are not linked to disease…

Katsanis, S.H., & Wagner, J.K. (2013) Characterization of the standard and recommended CODIS markers. *Journal of Forensic Sciences, 58(S1),* S169-S172.



JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2013, Vol. 58, No. S1
doi: 10.1111/j.1556-4029.2012.02253.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

CRIMINALISTICS; JURISPRUDENCE

Sara H. Katsanis,[1] M.S. and Jennifer K. Wagner,[2] J.D., Ph.D.

Characterization of the Standard and Recommended CODIS Markers*

"…we found no documentation of individual genotypes for the 24 STRs [the current and recommended CODIS loci] **to be causative of any documented phenotypes either in the literature or in the interrogated databases."**

See also on http://www.swgdam.org/

Open SWGDAM Letter Regarding the Claims Raised in State v. Abernathy that the CODIS Core Loci are Associated with Medical Conditions/Disease States

"The utility of the CODIS profile … is limited to identification purposes at this time."

"…we can affirm that individual genotypes are not at present revealing information beyond identification."

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

Mixture Section of STRBase

- **Training workshop slides** (thousands of slides of training materials available from 7 workshops)
- **SWGDAM Mixture Committee resource page** (contains worked mixture examples by Bruce Heidebrecht, Maryland State Police DNA Technical Leader)
- **Links to mixture interpretation software** (currently 12 links)
- **Literature references** (currently 144 articles)

Literature listing by topic for 144 articles

Topic category	# References
Mixture Principles & Recommendations	13
Setting Thresholds	11
Stutter Products & Peak Height Ratios	19
Stochastic Effects & Allele Dropout	18
Estimating the Number of Contributors	15
Mixture Ratios	9
Statistical Approaches	23
Low Template DNA Mixtures	8
Separating Cells to Avoid Mixtures	3
Software (plus 12 websites)	7
Probabilistic Genotyping Approach	11
General Information on Mixtures	7

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

NIST Applied Genetics

April 12 NIST DNA Mixture Webinar

DNA Analyst Training on Mixture Interpretation

Purpose:

The National Institute of Standards and Technology (NIST) will host a free one-day workshop on interpreting forensic DNA mixtures in casework. This workshop will be webcast live to maximize participation by forensic DNA analysts.

Agenda:

The workshop and webcast is expected to start at 8:30 AM ET and end at 5:00 PM ET on April 12, 2013. The agenda will be finalized closer to the workshop date.

Abstract:

DNA mixtures can be difficult to interpret and represent one of the biggest challenges faced by forensic laboratories today. With the success of DNA in the criminal justice system giving rise to expanded case acceptance policies, many DNA analysts are coping with complex mixtures arising from three or more contributors and/or low-level DNA

http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm

NIST Applied Genetics

Acknowledgments

Community feedback and contributors to the NIST STRBase website

Forensic DNA Team


John Butler


Mike Coble


Becky Hill


Margaret Kline


Dave Duewer

DNA Biometrics Team


Pete Vallone


Erica Butts


Kevin Kiesler

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

Funding from the **FBI Laboratory** through NIST Information Access Division

NIJ National Institute of Justice

Contact info:

john.butler@nist.gov

301-975-4049