

STRBase: 10 Years and Beyond

New Internet Resources for the Human Identity Testing Community

John M. Butler

Biochemical Science Division, National Institute of Standards and Technology, 100 Bureau Drive MS 8311, Gaithersburg, MD 20899-8311

General Information

Recent Updates Purpose of STRBase

NIJ-Funded Projects Training Materials Links to other web sites Glossary of commonly used terms

Forensic STR Information

Multiplex STR kits

Tri-Allelic Patterns

Variant Allele Reports

Published PCR primers

Amelogenin information

Forensic SNP information

Mitochondrial DNA (mtDNA)

Sex-typing markers

Cat STR Information

Dog STR information

Lab Resources and Tools

STR Allele Sequencing

Chromosomal Locations

Interlaboratory Studies

Validation information

Planned Future Additions

Training Materials

Y-chromosome STRs

Loci

STRs101: Brief Introduction to STRs

Sequence Information (annotated)

Mutation Rates for Common Loci

Publications and Presentations from NIST Project Team

Core Loci: FBI CODIS Core STR Loci and European Core

STR Fact Sheets (observed alleles and PCR product sizes)

miniSTRs (short amplicons) - including MiniFiler

STR Reference List - now >2900 references

Addresses for scientists working with STRs

Data from NIST U.S. Population Samples

New STR Markers under Development at NIST

NIST Mixture 2005 Interlab Study MIX05 Data

Mixture Interpretation – with worked examples

Technology for resolving STR alleles

Known STR Flanking Region Variation Slides showing applications for STR typing

DNA Advisory Board Quality Assurance Standards

DNA Quantitation - including information on SRM 2372

Statistical Analysis Resources - with worked examples

Population data - downloadable OmniPop program

NIST-Developed Software including mixSTR and Multiplex_QA

NIST Standard Reference Material for PCR-Based Testing

Null Alleles - discordance observed between STR kits

Other DNA Marker Information and Non-Human Resources

For more than a decade, the U.S. National Institute of Standards and Technology (NIST) has maintained the <u>Short Tandem Repeat DNA Internet DataBase</u> (STRBase), which is located at **http://www.cstl.nist.gov/biotech/strbase**/. The purpose of STRBase has been and continues to be an attempt to bring together the abundant literature and information in the human identity testing field in a cohesive fashion to make current and future work easier. STRBase recently became NIST Standard Reference Database (SRD) 130. New resources that are regularly added to STRBase can now be found quickly at **http://www.cstl.nist.gov/biotech/strbase/updates.htm**. A page was recently created to track null alleles detected through DNA testing with different primer sets (see **http://www.cstl.nist.gov/biotech/strbase/NullAlleles.htm**). In addition, a new software program enabling concordance checking with multiple data sets can be downloaded from **http://www.cstl.nist.gov/biotech/strbase/Software.htm**. Hundreds of new pages of information have been added in the past few years and numerous PowerPoint slides, NIST publications and presentations, software programs, and other useful information is available for download.

NIJ-Funded Projects at NIST

And the second se	anner oth Varian 118 Typing Iversit 12 Baiw Juleger C. Om Relati bitels and Provid Tyles
THE R P. LEWIS CO., LANSING MICH.	Terris E. Marine Garden R. Marine W. Anny E. Danier, Proc. 14, Springer V. Kine, and Select M. Barley
	Annothesis furthered the singlet in Earth Stage 1984 (see 1987). The same
	Approximate STM Access of Control Access of Cont
tint Contraction	Annual US Lot Seyed to COM Makers
and a second sec	Performant Control. Today MI, Maland S. Collector or 47521; Pere U. Value, Magner I. Kim and Annu K. Barlo
B All prime Property	Ensame by Sarage Decks 273, Institute States
county ways of 2019 w of atomation Name 71	francisco Inne Angel Participantes (San Marian Charles and Angel San Angel S
Name provide at the set	and if it is not been a second to been a second to the of "language with intervention process (21) are as and the second second and the language of the second sec
Annalyzed an added on	interages Programs The derivative replace plants have two program in 2014 and 1014 of plants 2014 as consentants of 1 spain. If a plant set of the spain of 1 spain, if a plant set of the spain of 1 spain. The spain of 1 spain of 1 spain of 1 spain of 1 spain. The spain of 1 spain. The spain of 1 spain of 1 spain of 1 spain of 1 spain. The spain of 1 spain of 1 spain of 1 spain. The spain of 1 spain of 1 spain of 1 spain of 1 spain. The spain of 1 spain of
122014	a plane or bing higher is astrone improvement of our best weight many forwards (2017) and (2018) and interactions of the conserve of 1.1. Proof better as Dates provides republicly Monarias Properties Astronetics in the Band Hand hang higher is approximately a static sectoring increasing control on the static sectoring and out of the state of the states of the plane of a static sectoring increasing control on the static sectoring and out of the states of the states of the plane of the plane of the static sectoring in the states of the states of the states of the states of the states of the plane of the plane of the states of the states of the states of the states of the states of the states of the states of the states of the states of the states of the states of the states
and set of the set of	3.9 A second second regards a constrained results of the second secon
	en constante de la constante de La constante de la constante de
	Read in 107 Name and Party in 177 and

Projects Performed by NIST Human Identity Project Team ABI 3100 performance with various STR typing systems ABI 31300 upgrade evaluation AutoDimer: software to enable rapid multiplex PCR design Autosamal STR loci: bayond the CODIS markers Biomatrica dry storage device DNA stability studies Cat STR assay, development of the "MeoryNec" DNA development of the "MeoryNec" DNA stability rom blocktains on various filter papers DNA stability studies

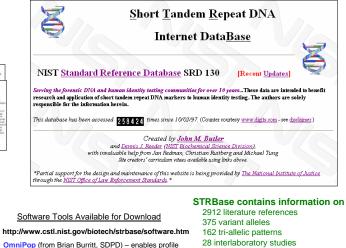
Dog STR assay development Expert Systems: FSS-i3 software evaluation Genetic variation in U.S. populations Interlaboratory studies: NIST Mixed Stain Study #1 & #2 Interlaboratory studies: NIST Mixed Stain Study #1 Interlaboratory studies: NIST Mixed Stain Study #3 Interlaboratory studies: DNA quantitation (QS04) Interlaboratory studies: STR mixture interpretation (MIX05) Literature collection on topics related to forensic DNA typing Locus-specific brackets (LSBs) for Y-STR typing Luminex system for Y-SNP typing Mass spectrometry SNP typing MiniFiler concordance study miniSTR assays for recovery of information from degraded DNA miniSTR: new loci and assays/support to other labs Mitochondrial DNA: LINEAR ARRAY evaluation Mitochondrial DNA: coding region SNP assay developmen Mitochondrial DNA: Affymetrix MitoChip evaluation mixSTR: a tool to aid mixture DNA cases Multiplex PCR assay development (2000-present) Multiplex QA: development of data quality assessment software Nuclear DNA from hair shafts Quality control methods for PCR primers SNP typing: ancestry informative marker assays SNP typing: identification assays SNP website for possible forensic SNP loci SRM 2391, 2391a, 2391b): PCR-based DNA Profiling Standard SRM 2395: Human Y-Chromosome DNA Profiling Standard SRM 2372: Human DNA Quantitation Standard STRBase: a human identity testing community resource STR kit primer sequence analysis Supplying samples to collaborators and colleagues Training materials: STR typing using capillary electrophoresis Training materials: mixture interpretation and LCN DNA testing Validation resources, standardization, and training Variant allele sequencing to determine basis for allele dropout Variant antice sequencing to detail the basis to antice dropout Y-chromosome: Duplication characterization Y-chromosome: Fxamination of the DVS464 multi-copy marker Y-chromosome: Y-SNP variation in U.S. populations Y-chromosome: Y-SNP kit evaluations Y-chromosome: Y-STR multiplex assay developmen Y-chromosome: Y-STR multiplex assay developmen Y-chromosome: Y-STR value of additional loci





John Margaret Jan Pete Becky Amy Dave Butler Kline Redman Vallone Hill Decker Duewer





match probability calculations across 202 populations

mixSTR - compares reference profiles to mixtures for

310/3100/3130 electropherogram data with quality metrics

STR MatchSamples - compares STR types from multiple

Downloadable PowerPoint files (>800 slides) and

theory and troubleshooting, low-copy number DNA

testing, and Y-chromosome and mtDNA analysis.

NIST Publications Available for Download

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

pdf presentations covering topics such as CE

STR ConvertFormats - transforms data from columns

Training Materials Available for Download

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

AutoDimer - aids multiplex assay design

Multiplex QA – enables comparison of ABI

data sets to aid concordance checking

rapid allele inclusion/exclusion

into rows

- 94 NIST publications (since 2000)
- 257 NIST presentations (since 2000)
- 342 weblinks
- 45 organizations 41 journals
- 18 academic and forensic institutes
- 117 commercial sites
- 21 genetic genealogy labs
- 57 paternity testing labs 9 legal sites dealing with forensic DNA
- 35 useful information sites

Variant Allele and Tri-Allelic Patterns

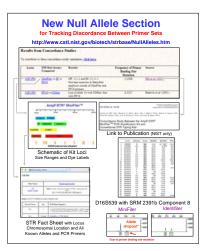
STR Locus	# Variant Alleles Reported	# Tri-Allelic Patterns Reported		
CSF1PO	17	5		
FGA	93	22		
TH01	12	1		
TPOX	15	15		
VWA	8	18		
D3S1358	26	6		
D5S818	7	4		
D7S820	24	7		
D8S1179	10	11		
D13S317	15	8		
D16S539	12	8		
D18S51	33	17		
D21S11	25	19		
D2S1338	7	2		
D19S433	18	2		
Penta D	28	5		
Penta E	20	11		
F13A01	1	0		
FES/FPS	1	1		
SE33	1	0		
D1S1677	1	0		
D14S1434	1	0		

Poster #6 at 18th International Symposium on Human Identification, Hollywood, CA, October 2-4, 2007

> Acknowledgments: Jan Redman for entry of variant alleles, references, and address information. Funding support from the National Institute of Justice through interagency agreement 2003-IJ-R-029 to the NIST Office of Law Enforcement Standards.

References

- Ruitberg, C.M., Reeder, D.J., Butler, J.M. (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res.* 29: 320-322.
- [2] Butler, J.M. and Reeder, D.J. (1997) STRBase: a short tandem repeat DNA database for the DNA typing community. Profiles in DNA, Promega Corporation, Volume 1, No. 2, p. 10. http://www.promega.com/profiles/102/Profiles/iDNA/T02_10.pdf
- [3] Butler, J.M., Ruitberg, C.M., Reeder, D.J. (1998) STRBase: a short tandem repeat DNA internetaccessible database. Proceedings of the Eighth International Symposium on Human Identification 1997, Promega Corporation, pp. 38-44. http://www.promega.com/geneticidipro/cussymp8proc/13.html
- [4] Butler, J.M. (2007) New resources for the forensic genetics community available on the NIST STRBase website. Progress in Forensic Genetics 12 (in press).



D16S539		http	://www.	cstl.nist.	.gov/biotech	strbase/str	_D16S539.htm	CSF1PO (US core) F13A1 F13B
Other Names Chromosomal Location					cation	GenBa	ink Accession	FES/FPS
D16 16q24.1				6-24.1	G07925; has 11 repeats			FGA (FIBRA) (US core) HPRTB
UniSTS: 45590 Chr 16; 84.94			16; 84.944	144 Mb (May 2004, NCBI AC024591.3 build 35)			.3; has 11 repeats	LPL Penta D
Repeat (GATA)	= GenBa	nik top str	and					Penta E
Reported Primers	Ref.	PCR	Primer Sequ	ences		SE33 (German core)		
Set 1	Promeg	5.00	GOTUTA	GAGCTEG	werPlex 16 (JOE la ITAAAAAG-3' TCTGTAAGCATG	TH01 (US core) TPOX (US core) VWA (US core)		
Set 2	CHLC	5-GA	TCCCAAGE	Terreer	CTT-F			D1S1656 (European recomme
	Web Sit		GTTTGTGT					D2S441 (European recommend
Set 3 Set 4	ABI				Plos (5-FAM label nobility modifier)	02)		D2S1338 (European locus)
	Jun				acces, aberna)			D3S1358 (US core) D5S818 (US core)
CR Product Si	zes of Ob	served A	lleles					D6S1043
Allele (Repeat #)	Set 1	Set 2	Set 3	Set 4*	Repeat S	tructure	Ref.	D7S820 (US core) D8S1179 (US core) D10S1248 (European recomme
4	260 bp	129 bp	229 bp	248 bp			variant allele	D12S391 (European recommer
5	264 bp	133 bp	233 bp	252 bp	[GATA],		721	D13S317 (US core) D14S1434
6	268 bp	137 bp	237 bp	256 bp			variant allele	D1431434 D16S539 (US core)
7	272 bp	141 bp	241 bp	260 bp			variant allele	D18S51 (US core)
8	276 bp	145 bp	245 bp	264 bp	[GATA] _x		721	D19S433 (European locus)
8.3	279 bp	148 bp	248 bp	267 bp			variant allele	D21S11 (US core) D22S1045 (European recomme
9	280 bp	149 bp	249 bp	268 bp	[GATA] ₉		721	DYS19 (European core)
9.3	283 bp	152 bp	252 bp	271 bp			variant allele	DYS385 a/b (European core) DYS388
10	284 bp	153 bp	253 bp	272 bp	[GATA]20		721	DYS389 I/II (European core)
11	288 bp	157 bp	257 bp	276 bp	[GATA]11		721	DYS390 (European core)
11.3	291 bp	160 bp	260 bp	279 bp			variant allele	DYS391 (European core) DYS392 (European core)
12	292 bp	161 bp	261 bp	280 bp	[GATA]12		721	DYS393 (European core)
12.1	293 bp	162 bp	262 bp	281 bp			variant allele	DYS434
12.2	294 bp	163 bp	263 bp	282 bp			variant allele	DYS437 DYS438 (SWGDAM)
13	296 bp	165 bp	265 bp	284 bp	[GATA] ₁₃		721	DYS439 (SWGDAM)
13.1	297 bp	166 bp	266 bp	285 bp			variant allele	DYS447
13.3	299 bp	168 bp	268 bp	287 bp			variant allele	DYS448 DYS456
14	300 bp	169 bp	269 bp	288 bp	[GATA]14		721	DYS458
14.3	303 bp	172 bp	272 bp	291 bp			variant allele 721	DYS460
15	304 bp 308 bp	173 bp 177 bp	273 bp 277 bp	292 bp 296 bp	[GATA] ₁₃		721 variant allele	DYS464 DYS635
Allelic Ladde	rs: Comm tiplexes: a: 0.11% STR Fact HOMEPA	ercially a PowerPle Sheet Pl	vailable fro ix 1.1, Pov	m Prome;	ga and Applied , COffler, SGM F	Biosystems has, Identifiler	- an out the second	Y-GATA-A4 Y-GATA-A7.1 Y-GATA-A7.2 Y-GATA-A10 Y-GATA-H4

Disclaimer

This project was funded by the National Institute of Justice through interagency agreement 2003-JH-R-20 to the NIST Office of Law Enforcement Standards. Points of view are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. 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 recommediation or endocresement 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 purces.

NIST publications and presentations: http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm NIST-developed software: http://www.cstl.nist.gov/biotech/strbase/software.htm NIST-collected population data: http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm

NIST-collected population data: http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm Poster available for download from STRBase: http://www.cstl.nist.gov/biotech/strbase/pub_pres/Butler_Promega2007poster.pdf