Poster #32 at 14th International Symposium on Human Identification,
Phoenix, AZ, Sept 30-Oct 2, 2003

NIST SRM 2395 and Other Y Chromosome Work

John M. Butler, Richard Schoske*, Peter M. Vallone, Jannette W. Redman, and Margaret C. Kline

Biotechnology Division, National Institute of Standards and Technology, 100 Bureau Drive MS 8311, Gaithersburg, MD 20899-8311
***Current Address: Armed Forces Institute of Pathology, 6825 16th Street, NW Building #54 Rm 3079, Washington, DC 20306**

Email: john.butler@nist.gov
Phone: 301-975-4049

At the U.S. National Institute of Standards and Technology (NIST), we have developed and recently released a Human Y-Chromosome DNA Profiling Standard Reference Material, SRM 2395, to enable calibration of Y-STR results across laboratories worldwide. SRM 2395 includes 5 male DNA samples selected to exhibit a diverse set of alleles across 31 commonly used Y chromosome short tandem repeat (STR) and 42 single nucleotide polymorphism (SNP) markers. A female DNA sample is also included to serve as a negative control for male-specific DNA tests. In addition to the typing results from all commercially available Y-STR kits, the five male samples in SRM 2395 have been sequenced at 22 Y-STR loci to confirm allele calls.

Our group has also been actively involved in developing novel Y-STR multiplexes [1,2] and evaluating the diversity of numerous Y-STRs in U.S. populations. Most recently, we have conducted a study of 22 common Y-STRs in 647 males from U.S. Caucasian, African American, and Hispanic populations [3]. In addition, 50 Y-SNPs have been examined in 229 U.S. Caucasians and African Americans [4]. We continue to post helpful developments in Y chromosome work on the STRBase web site: http://www.cstl.nist.gov/biotech/strbase. Y-STR fact sheets that describe primer sequences, allele sizes and sequences, and references to population data are available for all commonly used Y-STR loci [5].

References

[1] Schoske, R., *et al.* (2003) *Anal. Bioanal. Chem*. 375: 333-343

[2] Butler, J.M., *et al.* (2002) *Forensic Sci. Int.* 129: 10-24

[3] Schoske, R., *et al.* (2003) *Forensic Sci. Int., in press.*

[4] Vallone, P.M. and Butler, J.M., *submitted.*

[5] Butler, J.M. (2003) *Forensic Sci. Rev.* 15:91-111

Y-STR Information and Data

Marker Name	Allele Range	Repeat Motif	GenBank Accession	Reference Allele	Present in SRM 2395	Rank in NIST Population Study
DYS464 a/b/c/d	9-20	CCTT	X17354	13	*	1
DYS385 a/b	7-28	GAAA	AC022486 (r&c) #	11	*	2
YCAII a/b	11-25	CA	AC015978	23	*	3
DYS458	13-20	GAAA	AC010902	16	*	4
DYS390	17-28	(TCTA) (TCTG)	AC011289	24	*	5
DYS447	22-29	TAA**W**A compound	AC005820	23	*	6
DYS389 II	II: 24-34	(TCTG) (TCTA)	AC004617 (r&c)	29	*	7
DYS448	20-26	AGAGAT	AC025227	22	*	8
DYS456	13-18	AGAT	AC010106	15	*	9
DYS438	6-14	TTTTC	AC002531	10	*	10
DYS19	10-19	TAGA	AC017019 (r&c)	15	*	11
DYS439	9-14	AGAT	AC002992	13	*	12
DYS437	13-17	TCTA	AC002992	16	*	13
Y-GATA-H4	8-13 (25-30)	TAGA	AC011751 (r&c)	12	*	14
DYS392	6-17	TAT	AC011745 (r&c)	13	*	15
DYS460 (A7.1)	7-12	ATAG	AC009235 (r&c)	10	*	16
DYS389 I	I: 9-17	(TCTG) (TCTA)	AC004617 (r&c)	12	*	17
DYS391	6-14	TCTA	AC011302	11	*	18
DYS426	10-12	GTT	AC007034	12	*	19
DYS450	8-11	TTTTA	AC051663	9	*	20
DYS393	9-17	AGAT	AC006152	12	*	21
DYS388	10-18	ATT	AC004810	12	*	22
DYS425	10-14	TGT	AC095380	10		*not examined*
DYS434	9-12	TAAT (CTAT)	AC002992	10		*not examined*
DYS435	9-13	TGGA	AC002992	9	*	*not examined*
DYS436	9-15	GTT	AC005820	12	*	*not examined*
DYS441	12-18	CCTT	AC004474	14		*not examined*
DYS442	10-14	TATC	AC004810	12		*not examined*
DYS443	12-17	TTCC	AC007274	13		*not examined*
DYS444	11-15	TAGA	AC007043	14		*not examined*
DYS445	10-13	TTTA	AC009233	12		*not examined*
DYS446	10-18	TCTCT	AC006152	14		*not examined*
DYS449	26-36	TTTC	AC051663	29		*not examined*
DYS452	27-33	**Y**ATAC compound	AC010137	31		*not examined*
DYS453	9-13	AAAT	AC006157	11		*not examined*
DYS454	10-12	AAAT	AC025731	11		*not examined*
DYS455	8-12	AAAT	AC012068	11		*not examined*
DYS459 a/b	7-10	TAAA	AC010682	9		*not examined*
DYS461 (A7.2)	8-14	(TAGA) CAGA	AC009235 (r&c)	12	*	*not examined*
DYS462	8-14	TATG	AC007244	11	*	*not examined*
DYS463	18-27	AA**R**GG compound	AC007275	24		*not examined*
Y-GATA-C4	20-25	T**S**TA compound	G42673	21		*not examined*
Y-GATA-A10	13-18	TAGA	AC011751	13		*not examined*

#r&c refers to reverse and complement sequence orientation in order to maintain common repeat structure

Comparison of different Y-STR marker combinations in U.S. populations

Y-STR Marker Combinations	260 African Americans		244 U.S. Caucasians		143 U.S. Hispanics	
	HD	**RMP**	**HD**	**RMP**	**HD**	**RMP**
Y-PLEX 6 kit (6 loci, 7 amplicons)	0.9974	0.0026	0.9914	0.0086	0.9934	0.0066
"minimal" haplotype (7 loci, 9 amplicons)	0.9982	0.0018	0.9946	0.0053	0.9957	0.0043
"extended" haplotype (8 loci, 11 amplicons)	0.9988	0.0012	0.9971	0.0029	0.9975	0.0025
"U.S. haplotype" (9 loci, 11 amplicons)	0.9993	0.0007	0.9974	0.0026	0.9986	0.0014
NIST Y-STR 11plex (7 loci, 11 amplicons)	0.9993	0.0007	0.9987	0.0013	0.9992	0.0008
NIST Y-STR 20plex (17 loci, 20 amplicons)	0.9998	0.0002	0.9998	0.0002	0.9998	0.0002
11plex + 20plex (22 loci, 27 amplicons)	**0.9999**	**0.0001**	**0.9999**	**0.0001**	**0.9999**	**0.0001**
Top 10 loci (w/o YCAII a/b)	0.9999	0.0001	0.9999	0.0001	0.9999	0.0001

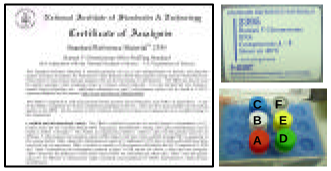
HD: haplotype diversity; **RMP**: random match probability

From Schoske et al. (2003) *Forensic Sci. Int., in press*

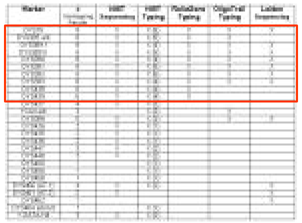
SRM 2395 Information and Data

- **5 male samples + 1 female sample (neg. control)**
- **100 ng of each component (50 µL at ~2 ng/µL)**
- **31 Y-STRs markers typed (22 are sequenced)**
- **42 Y-SNPs typed**

- Can be used to verify results with *any primer sets*
- Will help U.S. labs meet DAB/FBI Standard 9.5
- All commercial Y-STR and Y-SNP loci are present

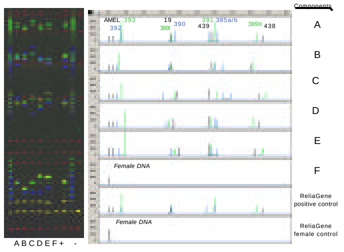


Now available from NIST Standard Reference Material office (http://www.nist.gov/srm)

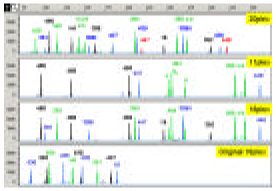


The 11 U.S. core Y-STR loci (boxed region) are well-characterized

FMBIO III and ABI 3100 Results with Y-PLEX™ 12



Results from SRM 2395 Component B with NIST Y-STR assays

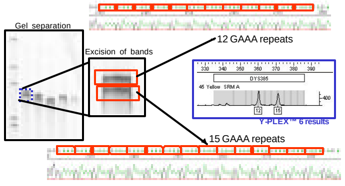


Our group at NIST has been active with developing new multiplex assays for typing Y-STR loci. Full details on the 20plex [2], 11plex [3], 18plex [3,5], and 10plex [1] have been published including primer sequences (see STRBase: http://www.cstl.nist.gov/biotech/strbase/y_strs.htm).

SRM components are all distinguishable from one another with these variable Y-SNPs

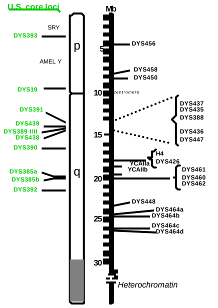
SRM 2395	AMEL	P25	M45	M89	DYS391	M2	M170	M172	M201	YCC
		(C/A)	(G/A)	(C/T)	(C/G)	(A/G)	(A/C)	(T/G)	(G/T)	Haplogroup
Component A	XY	A	A	T	C	A	A	T	G	**R1b**
Component B	XY	C	G	T	C	A	A	G	G	**J2**
Component C	XY	C	G	C	G	G	A	T	G	**E3a**
Component D	XY	C	G	T	C	A	A	T	T	**G**
Component E	XY	C	G	T	C	A	C	T	G	**I**
Component F	XX									

Results obtained with Signet Y-SNP Identification System Kit (Marligen Biosciences, Ijamsville, MD)



Alleles from multi-copy Y-STRs were first separated by gel electrophoresis and then sequenced individually. However, we were unsuccessful at sequencing the dinucleotide YCAII alleles using this approach.

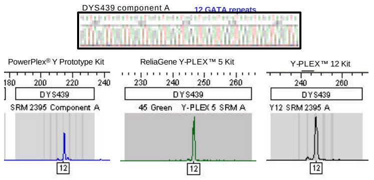
Y-STR Marker Positions for Loci Typed in SRM 2395

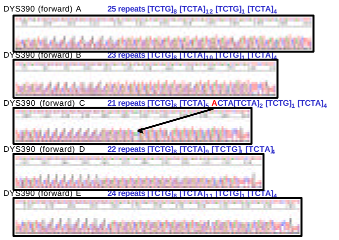


Reference materials help establish universal allele nomenclatures

Y-STR allele nomenclatures can differ in the literature making analysis of new loci confusing. For example, three different methods for designating DYS439 alleles have been published. SRM 2395 components can be used to help establish a uniform nomenclature in laboratories worldwide, which will greatly aid with development of databases for new Y-STR markers. We will continue to add information from new loci as they become accepted by the Y-chromosome typing community.

SRM 2395 Results





All five male components have different alleles for the compound tetranucleotide Y-STR locus DYS390. We discovered a previously unreported variant allele that occurs in the center of the repeat region for component C where a "T" becomes an "A" (see arrow).

Materials and Methods

NIST Y-STR Assays: Usually 20uL PCR volume (also 10 or 25uL); 2 U TaqGold; 1X Gold Buffer; 1.75mM $MgCl_2$; 300 uM dNTPs; 0.2-1.8 uM primers (see Refs. [1,2,3]); 5% (v/v) glycerol; 0.16ug/uL BSA; 0.05-5 ng DNA template; PCR as in Ref. [1,2,3] using GeneAmp 9700 with either 28 or 32 cycles; ABI 310 or ABI 3100 with appropriate dye matrices; Data analysis with GeneScan 3.7 and Genotyper 3.7
Commerical Kits: Followed manufacturer protocols; sometimes with 1/2X or 1/4X reduced volume PCR
Sequencing: Followed manufacturer protocols for BigDye v.3 kit after generating PCR products using unlabeled primers that were designed to anneal outside the assay primer positions
Y-SNP Typing (allele-specific hybridization): Followed manufacturer protocols for Marligen Signet Y-SNP kit on Luminex 100 platform
Y-SNP Typing (primer extension): Primers and PCR conditions in Ref. [4]

For more information on Y STRs, see STRBase:
http://www.cstl.nist.gov/biotech/strbase

Interlaboratory Studies Conducted with SRM 2395 Components

- **ReliaGene Technologies Inc.** (New Orleans, LA)
  - Y-PLEX 6: DYS19, DYS385 a/b, DYS389 II, DYS390, DYS391, DYS393
  - Y-PLEX 5: DYS389 I/II, DYS392, DYS438, DYS439
- **OligoTrail LLC** (Evanston, IL)
  - Locus-specific brackets: DYS19, DYS385 a/b, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS388, YCAII a/b
- **Peter de Knijff's Lab** (Leiden University, The Netherlands)
  - Sequence analysis: DYS19, DYS388, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS460, DYS461, DYS462

Y-SNP Information and Data

See **YCC 2002 Tree**; Y Chromosome Consortium (2002) *Genome Res.* 12:339-348

Y position (Mb)	SNP Name	YCC Hg Defined	ultiplex	Polymorphism
2,562,931	SRY+465	O2b	4	C->T
2,564,927	SRY10831a,b	B-R,R1a	5	A->G, G->A
2,566,620	SRY9138	K 1	4	C->T
2,642,605	M130(RPS4Y)	C	3	C->T
13,407,330	M2	E3a	2	A->G
13,413,670	DYS391	E 3	2	C->G
14,124,138	M168	CR	1	C->T
14,157,939	M170	I	3, A	A->C
14,179,223	M182	B 2	2	C->T
14,232,730	Tat	N 3	4	T>C
14,264,427	M174	D	3, A	T>C
14,279,781	M172	J2	3, A	T>G
14,337,676	M201	G	3	G->T
14,340,899	M198	R1a1	C	C->T
14,818,875	M175	O	1	-5bp (+/-)
14,892,152	M207	R	1	A->G
18,343,057	M3	Q 3	4, B	C->T
20,806,634	M5	M	4	C->T
20,807,275	P 3	A 2	2	G->A
20,903,048	M153	R1b3d	5	T>A
20,926,945	M9	K-R	B	C->G
20,927,335	M11	L	3, B	A->G
20,929,850	M18	R1b1	5	+2bp (-/+)
20,936,442	M31	A 1	2	G->C
20,937,124	M32	A 3	2	T>C
20,937,138	M33	E 1	2, B	A->C
20,938,391	M35	E3b	2, B	G->C
20,948,239	M37	R1b3a	5	C->T
20,949,887	M52	H	3	A->C
20,959,373	M119	O 1	4, A	A->C
20,961,146	M137	J2c	C	T>C
20,961,189	M124	R 2	4, C	C->T
20,961,274	M123	E3b3	C	G->A
20,961,362	M122	O 3	C	T>C
20,961,382	M166	J2f2	C	G->A
20,964,647	M112	B2b	A	G->A
20,975,939	M146	B 1	2	A->C
21,063,528	M42	B-R	1	A->T
21,064,475	M45	P-R	1	G->A
21,064,542	M157	R1a1b	5	A->C
21,066,207	M150	B2a	2	C->T
21,074,761	M60	B	1	+1bp (-/+)
21,086,865	M75	E 2	2, A	G->A
21,090,746	M69	H	B	T>C
21,102,797	M87	R1a1c	5	T>C
21,114,001	M89	F-R	1	C->T
21,134,846	M94	B-R	1	C->A
21,135,132	M95	O2a	4	C->T
22,653,504	P 4	A 2	2	C->T
24,142,855	P25a	R1b	5	C->A
25,868,376	P25b			
26,600,000	P25c			

P25 is a multicopy marker that occurs 3 times on the Y-chromosome

50 Y-SNP Markers Examined

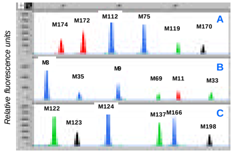
Samples Studied
- 114 U.S. Caucasians
- 115 African Americans

Two Typing Methods
- **Allele-specific Hybridization**
  42 Y SNPs + amelogenin using Marligen Signet Y SNP Identification kit
- **Allele-specific Primer Extension**
  18 Y SNPs with SNaPshot assays developed at NIST

10 Overlapping loci between methods (boxed)
>2,000 allele calls compared between two methods
- **Complete Concordance Seen!**

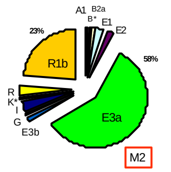
18 Y SNPs with SNaPshot assays developed at NIST

Primer extension length (nucleotides) relative to GS120 LIZ sizestandard

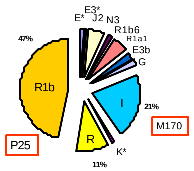


As of Fall 2003, over 250 Y-SNP markers have been characterized on the Y-chromosome. We have typed 50 of these Y-SNPs to enable coverage of all major haplogroups. A total of 18 different haplogroups (from a possible 45) were observed in 229 U.S. males from African American and Caucasian ancestry. Many of the haplogroups are present in both population groups. However, note that the M2 marker is only observed in the African American population. On the other hand, the primary Caucasian marker, P25, is seen in 23% of the African Americans.

Y SNP haplogroups for 115 African Americans



Y SNP haplogroups for 114 Caucasians



18 different haplogroups observed in 229 males
(6 haplotypes in both populations and 6 unique for each)

Acknowledgments
- Funding from the National Institute of Justice through the NIST Office of Law Enforcement Standards
- Funding from the NIST Office of Standard Reference Materials Program
- Much of this work was conducted by Rich Schoske while he was working at NIST as an American University graduate student under the advisement of Dr. Jim Girard with funding by the United States Air Force, through the Air Force Institute of Technology
- Mecki Prinz and Howard Baum for initial discussions on requirements for Y chromosome standards
- Alan Redd and Mike Hammer for valuable collaborations on Y-chromosome projects and helpful discussions
- Dave Carlson for advice with his Marligen Y-SNP assay on the Luminex 100 platform