Poster B39 at American Academy of Forensic Sciences,
Seattle, WA, February 23, 2006

Examination of Additional Y-STR Loci for Increased Resolution of Common Haplotypes

Email: amy.decker@nist.gov
Phone: 301-975-5205

Amy E. Decker, Peter M. Vallone, Margaret C. Kline, Janette W. Redman, and John M. Butler
National Institute of Standards and Technology, Biochemical Sciences Division, Gaithersburg, MD 20899-8311

Copy of poster available:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

The need for additional Y-STR loci has become increasingly important as the potential forensic uses of Y chromosome testing are revealed. One major challenge in forensic investigations has been the recovery of genetic information from perpetrators in sexual assault cases involving low amounts of male DNA mixed with high levels of female DNA. While several commercial Y-STR kits have been developed to focus on this issue, additional loci could assist in increasing the power of discrimination between closely related male lineages. At the same time, new Y-STR loci may also help to separate related males, such as fathers and sons.

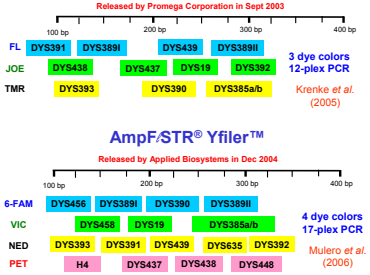
We have examined 27 Y-STR loci across approximately 660 U.S. Caucasian, African American, and Hispanic samples (Butler et al. 2006). An approach for selecting and evaluating these loci is described including their ability to resolve samples with a common type beyond commercial Y-STR kits. In addition, full concordance has been observed where the Y-STR loci overlap between in-house multiplex assays (Schoske et al. 2004) and commercial Y-STR kits.

Why use Y-STRs?

Y-chromosome STRs are attractive to the forensics community due to their male-specific amplification. This quality is useful in sexual assault situations where the female victim's DNA is in far excess to the male perpetrator's DNA. Y-STRs are also beneficial in difficult cases with no sperm evidence or only fingernail scrapings from the victim. Other applications for Y-STRs include paternity testing, tracing paternal lineages to aid in missing persons investigations, historical studies and to help link families through genetic genealogy.

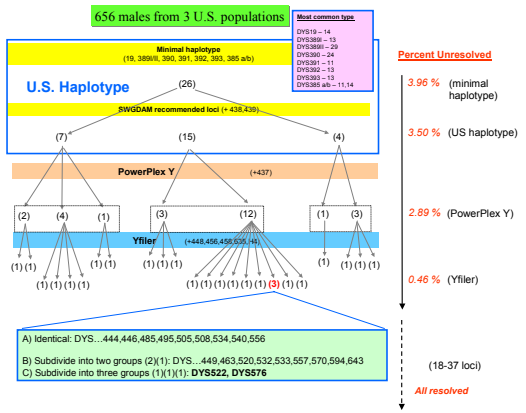
Y-STR Testing Kits

Two commercial kits are available in the U.S. for amplifying Y-STR loci. Powerplex Y amplifies 12 Y-STR loci in 3 dye colors whereas Yfiler amplifies 17 Y-STR loci in 4 dye colors. The schematic layouts display the locus size ranges of the PCR products.



Subdividing Common Types with Additional Loci

The 26 samples with the most common type can be resolved in this sample population with use of the 17 Yfiler loci plus DYS522 or DYS576.



26 of 656 population samples (3.96%) were found to have the most common type with the minimal haplotype loci. Adding the 2 SWGDAM loci breaks these samples into 3 groups. PowerPlex Y further resolves the samples into 7 groups, whereas Yfiler loci will resolve all but 3 samples with the most common type. Using the loci in group "A" will not improve resolution, however with any group "B" locus, one of the 3 samples is resolved. Adding either DYS522 or DYS576 to Yfiler loci resolves all 26 samples with the most common type.

Resolving 656 U.S. Population Samples

From the 9 loci of the minimal haplotype (MHL) run on 656 samples, 26 samples had the most common type. 429 of the 656 had a unique haplotype with the MHL loci, 34 sample haplotypes were observed twice in the sample set, 13 sample haplotypes were observed three times, etc. As more markers are added to the MHL loci, the number of unique haplotypes increase, as does the haplotype diversity (HD) and the percent discrimination capacity (%DC). When all 37 loci (including MHL, SWGDAM, Powerplex Y, Yfiler, and 20 new Y-STRs) are run on 656 samples, only two haplotypes are observed twice (4 total samples) and therefore unresolved in the sample data set.

# times haplotype observed	MHL	SWGDAM	PPY	Yfiler	ALL 37	
1	429	486	505	626	652	Unique HT
2	34	33	34	12	2	
3	13	10	14	2	.	
4	4	6	3	.	.	
5	3	1	2	.	.	
6	1	1	.	.	.	
7	1	2	1	.	.	
8	1	.	.	.	.	
9	2	.	.	.	.	
10	.	1	.	.	.	
11	1	.	.	.	.	
12	.	.	1	.	.	
13	1	.	.	.	.	
14	.	.	.	.	.	
15	.	1	.	.	.	
16	.	.	.	.	.	
17	.	.	.	.	.	
18	.	.	.	.	.	
19	.	.	.	.	.	
20	.	.	.	.	.	
21	.	.	.	.	.	
22	.	.	.	.	.	
23	.	.	.	.	.	
24	.	.	.	.	.	
25	.	.	.	.	.	
26	1	.	.	.	.	
HD	0.995644	0.998529	0.999064	0.999916	0.999991	
%DC	0.748476	0.824695	0.853659	0.97561	0.996951	
# HT	491	541	560	640	654	total 656

Sample Info	DYS446	DYS486	DYS445	DYS448	DYS495	DYS505	DYS508	DYS522	DYS533	DYS534	DYS542	DYS556	DYS557	DYS570	DYS576	DYS594	DYS643
# of sample repetitions	f	3	2	3	2	0	1	3	6	7	4	3	2	3	5	3	2
MT07165	12	13	30	24	15	16	12	11	21	10	14	13	15	12	11	15	16	17	9	10
JT75333	12	13	30	24	15	16	12	11	22	12	14	12	15	12	11	17	17	20	10	11
T01702	12	13	31	29	15	16	12	11	21	11	13	13	15	12	11	17	16	18	10	10
PT85064	12	13	31	18	15	16	11	11	19	12	14	11	16	11	12	18	22	18	13	12
PT84048	12	12	31	18	15	12	11	11	19	12	14	11	17	11	12	18	22	18	11	12
JT85366	12	12	31	18	15	12	11	11	19	12	14	11	17	11	12	18	22	18	11	12
UT85837	12	13	30	24	15	16	11	11	21	11	13	12	16	12	11	15	17	19	10	10
JT51476	12	13	30	24	15	16	11	11	21	11	13	12	16	12	11	15	17	19	10	10
UC19777	12	13	29	24	15	16	12	11	21	11	14	13	14	12	11	16	18	18	10	10
DT07753	12	15	24	24	17	16	12	12	21	10	13	12	16	12	11	16	18	18	10	10
Y4	12	13	29	24	14	16	11	11	21	11	11	12.13	15.17	12	11	16	18	18	10	9
MT07715	13	13	26	24	15	16	11	11	21	11	13	12	14	12	10	15	17	18	10	10
ZT78994	12	14	31	24	15	16	11	11	21	10	15	12	16	12	11	17	16	20	10	10
ZT78995	12	14	31	24	15	16	11	11	21	10	15	12	17	12	11	17	16	20	10	10
WA29612	12	13	31	24	15	16	13	11	22	10	13	12	15	13	11	16	17	16	9	10
WT92462	12	13	31	24	15	16	13	11	22	10	13	12	15	13	11	16	17	16	9	10
WT92479	13	13	28	21	12	17	11	10	21	12	12	11	16	12	11	15	23	16	10	12
MT94827	13	14	28	21	12	17	11	10	20	10	13	11	16	12	12	16	22	17	10	12
TT91696	11	13	29	21	16	16	14	10	23	11	10	11	15	12	12	14	17	15	10	8
PT84213	11	13	30	22	16	16	11	10	22	11	11	11	15	12	12	14	17	16	10	8
Y27	12	13	29	24	16	16	12	11	21	10	12	12	14	12	11	15	17	19	10	10
Y28	12	13	29	24	16	16	12	11	21	10	12	12	14	12	11	15	17	19	10	10
MT94829	13	12	31	24	15	15	13	11	23	11	12	12	13	11	12	15	17	16	11	10
MT87158	13	12	31	24	15	15	13	11	23	12	12	12	13	11	12	15	17	16	11	10
PT84253	13	12	29	20	14	15	11	13	23	11	11	11	15	10	11	20	19	15	11	14
ZT78204	13	12	29	20	14	15	12	14	23	11	11	12	15	10	11	20	19	15	11	14
Y17	12	13	30	20	14	15	12	14	23	11	13	11	15	12	12	21	19	14	11	14
Y16	12	13	30	20	14	15	12	14	23	11	13	11	15	12	12	21	19	14	11	14
ZT78500	12	14	31	21	14	15	13	12	23	11	12	11	15	12	12	17	17	17	11	13
JT01484	12	14	30	21	14	15	13	12	23	11	12	11	15	12	12	19	16	16	11	13

640 haplotypes were observed in the 656 U.S. population samples with the Yfiler loci. 626 of these haplotypes were unique, 2 haplotypes were observed 3 times, and 12 haplotypes were observed 2 times (see chart above). With the addition of 20 new Y-STR loci, all but two sample pairs (highlighted in blue) are resolved.

The Y-STR loci in boxes demonstrated the best performance in separating the unresolved Yfiler haplotypes. In this population set, these 7 Y-STRs have the same ability to resolve the sample haplotypes as all 20 new loci. Therefore these loci will be the focus for future studies and for multiplexing.

Selection and Characterization of New Y-STR Loci

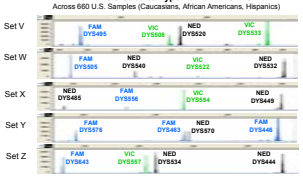
Potential new Y-STR loci were selected based on best candidates from previous studies (Redd et al. 2002 and Kayser et al. 2004) and mapped using primer sequences present in the Genome Database (http://www.gdb.org). The exact Y-chromosome locations were determined using hgBLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and the July 2003 human genome reference sequence. The loci were then divided into small multiplexes based on the range of alleles found in sample testing. Multiple alleles were sequenced for each Y-STR locus to aid in determining repeat number from the observed base pair size.

Example of PCR primer positions using DNAStar software

Y-Chromosome Mapping

Y-STR locus	Y-chromosome position (Mb)
DYS444	17.664
DYS446	3.175
DYS449	8.261
DYS463	7.687
DYS485	20.488
DYS490	3.487
DYS495	13.450
DYS504	2.947
DYS505	3.684
DYS508	16.232
DYS520	7.774
DYS522	7.459
DYS525	7.120
DYS532	8.423
DYS533	16.831
DYS534	16.831
DYS540	17.004
DYS556	20.940
DYS557	21.573
DYS570	6.905
DYS575	7.480
DYS576	7.097
DYS594	20.045
DYS632	14.347
DYS635	13.690
DYS641	14.573
DYS643	15.864

5 Panels Used to Type 20 New Y-STRs
Across 660 U.S. Samples (Caucasians, African Americans, Hispanics)



Y-STR Multiplex PCR Protocol

Reaction volume of 10 µL: 1 unit AmpliTaq Gold® DNA polymerase; 1X Gold Buffer (Applied Biosystems); 1.5 mM $MgCl_2$, 250 µM dNTPs, 0.16 µg/µL BSA, and 1 µM primers.

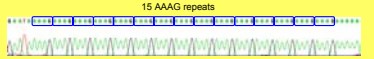
Thermal cycling: 95 ºC for 10 minutes; 30 cycles (94 ºC-1 min, 55 ºC-1 min, 72 ºC-1 min); 60 ºC-45 min, 25 ºC-hold

DNA template: 1-2 ng of genomic DNA

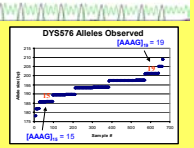
3100 conditions: POP-6 polymer, 36cm capillary array, GS-500 LIZ size standard, GS500 analysis parameters

At least two alleles were sequenced for each Y-STR locus to define the allele nomenclature

DYS576
15 AAAG repeats



Allele	Combined Freq (N = 659)
13	0.0030
14	0.0273
15	0.1047
16	0.1745
17	0.2640
18	0.2822
19	0.1077
20	0.0288
21	0.0076

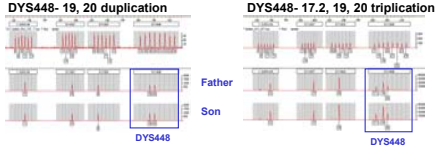


Examination of Father-Son Pairs

We have examined 300 father:son pairs (600 samples). Buccal swabs were extracted with DNA IQ, quantified with an Alu qPCR assay (Buel et al. 2003), and typed with Identifiler and Yfiler STR kits to obtain information on 15 autosomal STRs and 17 Y-STRs.

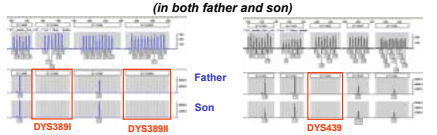
As noted previously (Butler et al. 2005), duplications and deletions can occur on the Y-chromosome, which may be seen in both father and son.

Duplication/ Triplication



Deletions

DYS389I, DYS389II, DYS439 were all deleted in these samples (in both father and son)



Y-STR Mutation Rates

In almost 300 father:son sample pairs, we observed 14 differences between father and son with the 17 Y-STR loci in the Yfiler kit. Five mutations resulted in the loss of a repeat in the son and 9 loci gained a repeat. All samples resulted in single repeat mutations except one sample which was a two repeat loss at Y-GATA-H4. Also, one sample pair was found to have two mutations (DYS635 and DYS458). Additional mutations in father and son pairs for the 17 Y-STR loci have been reported in the literature.

Yfiler kit loci	Literature Summary *			NIST Results			
Locus	Mutations	# Meioses	Mutation Rate	Mutations	# Meioses	Mutation Rate	TOTAL
DYS19	12	7272	0.165%	0	297	0.000%	**0.159%**
DYS389I	11	5476	0.201%	3	297	1.010%	**0.243%**
DYS389II	12	5463	0.220%	3	297	1.010%	**0.260%**
DYS390	16	6824	0.234%	1	293	0.341%	**0.239%**
DYS391	23	6702	0.343%	0	297	0.000%	**0.329%**
DYS392	4	6668	0.060%	0	297	0.000%	**0.057%**
DYS393	4	5456	0.073%	0	298	0.000%	**0.070%**
DYS385a/b	22	9980	0.220%	0	297	0.000%	**0.214%**
DYS438	1	2434	0.041%	0	297	0.000%	**0.037%**
DYS439	12	2409	0.498%	2	296	0.676%	**0.518%**
DYS437	5	2395	0.209%	0	296	0.000%	**0.186%**
DYS448	0	143	0.000%	0	294	0.000%	**<0.23%**
DYS456	1	143	0.699%	1	296	0.338%	**0.456%**
DYS458	3	143	2.098%	2	297	0.673%	**1.136%**
DYS635	3	1016	0.295%	3	298	1.007%	**0.457%**
GATA-H4	3	1179	0.254%	2	296	0.676%	**0.339%**

* Literature summary from www.YHRD.org and papers in press

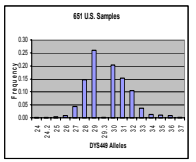
Examining Additional Loci

27 new Y-STR loci have been characterized recently including DYS635 found in the ABI Yfiler kit. 21 of these loci were typed with over 650 U.S. Caucasians, African Americans and Hispanics. Six loci were examined with only a subset of 94 samples due to low variation, poor performance or X-chromosome homology.

All haplotype data available at
http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm

The Allele Frequency Distribution Observed with U.S. Samples for DYS449

Locus	Allele	Combined Freq (N = 651)	Cau freq (N = 261)	Afr Am freq (N = 253)	Hisp freq (N = 137)
DYS449	24	0.0015	0.0038	0.0000	0.0000
	24.2	0.0015	0.0038	0.0000	0.0000
	25	0.0046	0.0000	0.0000	0.0219
	26	0.0077	0.0115	0.0040	0.0073
	27	0.0445	0.0421	0.0553	0.0292
	28	0.1459	0.1724	0.1186	0.1460
	29	**0.2611**	**0.2969**	0.1818	0.3358
	29.3	0.0015	0.0038	0.0000	0.0000
	30	0.2028	0.2490	0.1779	0.1606
	31	0.1536	0.1149	**0.2016**	**0.1387**
	32	0.1045	0.0651	0.1542	0.0876
	33	0.0369	0.0230	0.0435	0.0511
	34	0.0123	0.0038	0.0198	0.0146
	35	0.0108	0.0038	0.0237	0.0000
	36	0.0077	0.0038	0.0119	0.0073
	37	0.0031	0.0000	0.0079	0.0000



Remaining allele frequency data available: Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR Loci with U.S. Caucasian, African American, and Hispanic samples. Forensic Sci. Int. 156:250-260.

Conclusions

1. An approach for examining new Y-STR loci has been established.
2. Studies with father and son sample pairs will aid in estimating relative mutation rates of various Y-chromosome STR loci and will assist in understanding which and how many Y-STRs may be optimal for differentiating between fathers and sons.
3. The addition of more Y-STR loci beyond the minimal haplotype has increased the ability to resolve samples from one another in a sample population.

Acknowledgments and Disclaimer

Special thanks to David Duewer for his assistance with data analysis and to Tom Reid at DNA Diagnostics Center for providing the father-son samples.

These projects were funded by the National Institute of Justice through interagency agreement 2003-IJ-R-029 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 recommendation or endorsement 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 purpose.

References

1) Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR Loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250-260.

2) Butler, J.M., Decker, A.E., Kline, M.C., Vallone, P.M. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation. *J. Forensic Sci.* 50(4): 853-859.

3) Schoske, R., Vallone, P.M., Kline, M.C., Redman, J.W., Butler, J.M. (2004) High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Sci. Int.* 139: 107-121.

4) Redd, A.J., *et al.* (2002) Forensic value of fourteen novel STRs on the human Y chromosome. *Forensic Sci. Int.* 130: 97-111

5) Kayser, M., *et al.* (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am. J. Hum. Genet.* 74: 1183-1197.

6) Krenke, B., *et al.* (2005) Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci. Int.* 148(1):1-14.

7) Mulero, J., *et al.* (2006) Development and validation of the AmpFtSTR® Yfiler™ PCR Amplification Kit: a male specific, single amplification 17 Y-STR multiplex system. *J. Forensic Sci.* 51(1): 64-75.

For more information, see STRBase:
http://www.cstl.nist.gov/biotech/strbase/y_strs.htm