Evaluation of Additional Y-STR Loci to Resolve Common Haplotypes

Poster #2 at Promega Meeting, Nashville, TN, October 10-12, 2006

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

Amy E. Decker, Margaret C. Kline, and John M. Butler

National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, MD 20899-8311

Copy of poster available:
http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2006_Decker.pdf

We are investigating the advantages of additional Y-STR loci beyond those currently available in commercial kits such as Promega's Powerplex Y [1] and Applied Biosystem's Yfiler [2]. The approach for selecting and evaluating these loci is detailed as well as their ability to resolve samples with common types that could not be resolved with the commercial Y-STR kits. A total of 82 Y-STRs have been characterized with a subset of U.S population samples [3,4]. From these studies, the Y-STR loci that appear to be the most forensically useful include DYS449, DYS481, DYS505, DYS522, DYS527 (a duplicated locus), DYS532, DYS534, DYS570, DYS576, DYS607, DYS650, DYS652, DYS709, DYS710, DYS712, DYS715, and DYS724 (a duplicated locus).

Another added benefit of these additional Y-STRs is their ability to increase the power of discrimination between closely related male individuals, such as fathers and sons. We have examined 350 father:son sample pairs from Caucasian, African American, Hispanic and Asian populations using the 17 Y-STR loci in the Yfiler kit. A total of 19 mutations were observed as well as several duplications and deletions [5]. In addition, we are updating SRM 2395 Human Y-chromosome DNA profiling standard certificate with additional Y-STR loci in order to support the genetic genealogy community which is continually adding new Y-STRs.

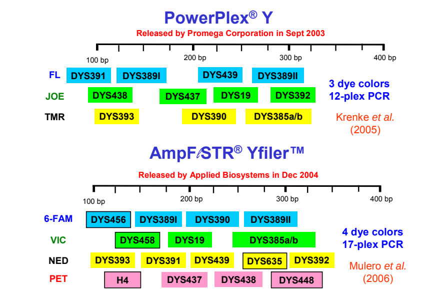
Why use Y-STRs?

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.

Most forensic DNA laboratories will only use commercially available kits due to quality control issues

Y-STR Testing Kits

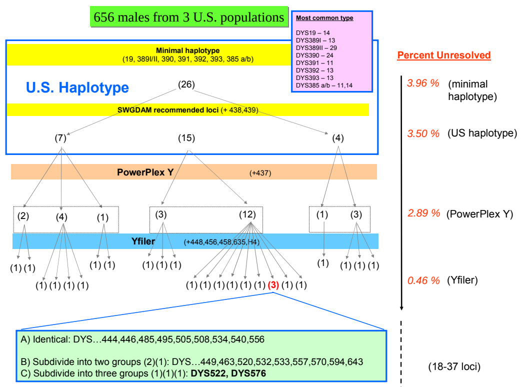
Two commercial kits are available in the U.S. for amplifying Y-STR loci. Powerplex Y amplifies 12 Y-STR loci in 3 dye colors whereas Yfiler amplifies 17 Y-STR loci in 4 dye colors. The schematic layouts display the locus size ranges of the PCR products.



Using commercial Y-STR kits as a starting point, are there additional loci that would be beneficial in separating samples with common types, which could be advocated to companies for possible future adoption in Y-STR kits?

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.

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

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. 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	0.996644	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	DYS444	DYS446	DYS449	DYS463	DYS485	DYS495	DYS505	DYS508	DYS520	DYS522	DYS532	DYS533	DYS534	DYS540	DYS556	DYS557	DYS570	DYS576	DYS594	DYS643
# of sample separations	1	3	5	3	2	0	5	1	3	6	7	4	4	2	2	3	5	5	0	2
MT97185	12	13	30	24	15	16	12	11	21	10	14	13	15	12	11	15	16	17	9	10
ZT79333	12	13	30	24	15	16	12	11	22	12	14	12	15	12	11	17	17	20	10	11
TT51702	12	13	31	23	15	16	12	11	21	11	13	13	15	12	11	17	16	18	10	10
PT83904	12	12	31	18	15	12	11	11	19	12	14	11	16	11	12	18	22	18	11	12
PT84348	12	12	31	18	15	12	11	11	19	12	14	11	17	11	12	18	22	18	11	12
ZT80369	12	12	31	18	15	12	11	11	19	12	14	11	17	11	12	18	22	18	11	12
UT58337	12	13	30	24	15	16	11	11	21	11	13	12	16	12	11	16	17	19	10	10
JT51476	12	13	30	24	15	16	12	11	21	11	13	13	16	12	11	16	17	17	10	10
UC10177	12	13	29	24	15	16	12	11	21	11	14	13	14	12	11	16	18	18	10	10
OT07753	12	15	24	24	17	16	12	12	21	10	13	12	15	12	11	16	17	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
MT97176	13	13	26	24	15	16	12	11	21	11	13	12	14	12	10	16	17	18	10	10
ZT79994	12	14	31	24	15	16	11	11	21	10	15	12	16	12	11	17	16	20	10	10
ZT79995	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
WT52482	12	13	31	24	15	16	13	11	22	10	13	12	15	13	11	16	17	16	9	10
WT52470	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
TT51696	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	10	15	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	16	17	19	10	10
Y28	12	13	29	24	16	16	13	11	21	14	12	12	14	12	11	16	17	19	10	10
MT94839	13	12	31	24	15	15	13	11	23	11	12	12	13	11	12	15	17	16	11	10
MT97158	13	12	31	24	15	15	13	11	23	12	12	12	13	12	12	15	17	16	11	10
PT84253	13	12	29	20	14	15	11	13	23	11	11	11	15	13	11	20	19	15	11	14
ZT79304	13	12	29	20	14	15	13	13	23	11	12	11	15	12	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	14	11	15	12	12	21	19	14	11	14
ZT79620	12	14	31	21	14	15	13	12	23	11	12	11	15	12	12	17	17	17	11	13
JT51484	12	15	32	20	14	15	13	12	23	11	12	11	15	12	12	19	19	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.

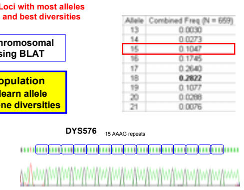
The Y-STR loci in boxes demonstrated the best performance in separating the unresolved Yfiler haplotypes. In this population set, these 7 Y-STRs have the same ability to resolve the sample haplotypes as all 20 new loci. Therefore these loci will be the focus for future studies and for multiplexing.

References

1) 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.
2) Mulero, J., *et al.* (2006) Development and validation of the AmpFℓSTR® Yfiler™ PCR Amplification Kit: a male specific, single amplification 17 Y-STR multiplex system. *J. Forensic Sci.* 51(1): 64-75.
3) 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.
4) 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.
5) 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.
6) Redd, A.J., *et al.* (2002) Forensic value of fourteen novel STRs on the human Y chromosome. *Forensic Sci. Int*. 130: 97-111
7) Kayser, M., *et al.* (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am. J. Hum. Genet.* 74: 1183-1197.
8) Nicklas, J.A. and Buel, E. (2003) Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. *J. Forensic Sci.* 48(5): 936-944.
9) Gusmão, L., Butler, J.M., et al. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci. Int.* 157:187-197.
10) Hanson, E.K. and Ballantyne, J. (2006) Comprehensive annotated STR physical map of the human Y chromosome: forensic implications. *Legal Med*. 8: 110-120.
11) Leat, N., et al. (2006) Properties of novel and widely studied Y-STR loci in three South African populations. *Forensic Sci. Int., in press*.

Selection and Characterization of New Y-STR Loci

Select best loci from previous published studies (often little population data available)
Loci with most alleles and best diversities
BLAST to obtain genome sequence using GDB primers
Determine chromosomal location using BLAT
Divide loci into small multiplexes based on size (GDB primers – no redesign)
Perform population testing to learn allele ranges and gene diversities
Sequence at least two alleles to define repeat nomenclature
Create bins and panels for GeneMapper sample analysis
Construct allelic ladders



We plan to develop large megaplex reactions with the best additional Y-STRs

Potential new Y-STR loci were selected based on best candidates from previous studies (Redd et al. 2002, Kayser et al. 2004, Leat et al. 2006). The selected loci were examined in a U.S. population screen consisting of 31 Caucasians, 32 African Americans, and 32 Hispanics. The number of alleles and calculated gene diversity values are shown below for 82 Y-STR loci. Those loci highlighted in green have been sequenced and are included in the SRM 2395 certificate. We are currently sequencing loci highlighted in yellow.

Locus (# samples)	# Alleles	Diversity
DYS724 a/b (CD Y) (93)	36	0.9691
DYS464 a/b/c/d (91)	42	0.9646
DYS527 a/b (93)	32	0.9388
DYS710 (93)	17	0.9236
DYS385 a/b (94)	29	0.9179
DYS481 (93)	11	0.8359
DYS449 (90)	12	0.8345
DYS712 (95)	12	0.8340
DYS490 (92)	18	0.8201
DYS504 (94)	9	0.8101
DYS576 (93)	8	0.8046
DYS570 (94)	10	0.8042
YCAII a/b (91)	13	0.7993
DYS557 (93)	7	0.7887
DYS534 (93)	9	0.7882
DYS643 (92)	7	0.7862
DYS458 (94)	8	0.7808
DYS635 (94)	8	0.7779
DYS652 (95)	10	0.7742
DYS650 (95)	10	0.7740
DYS459 a/b (95)	6	0.7680
DYS463 (95)	9	0.7680
DYS447 (91)	9	0.7636
DYS390 (94)	6	0.7632
DYS715 (94)	7	0.7628
DYS532 (94)	7	0.7541
DYS389II (94)	5	0.7447
DYS709 (95)	8	0.7402
DYS456 (94)	5	0.7355
DYS607 (95)	7	0.7355

Locus	# Alleles	Diversity
DYS438 (94)	5	0.7211
DYS19 (94)	5	0.7113
DYS508 (93)	7	0.7106
DYS446 (94)	7	0.7014
DYS448 (94)	6	0.6937
DYS723 (94)	4	0.6891
DYS485 (93)	8	0.6821
DYS522 (94)	4	0.6792
DYS495 (94)	5	0.6747
DYS716 (95)	4	0.6524
DYS452 (93)	7	0.6487
Y-GATA-H4 (94)	5	0.6461
DYS505 (93)	5	0.6454
DYF406S1 (DYS555)	5	0.6421
DYS437 (94)	5	0.6417
DYS439 (94)	4	0.6388
DYS520(94)	6	0.6381
Y-GATA-A10 (95)	4	0.6336
DYS492 (93)	5	0.6335
DYS444 (88)	6	0.6264
DYS533 (94)	6	0.6264
DYS460 (91)	4	0.5973
DYS392 (94)	7	0.5962
DYS389I (94)	3	0.5692
DYS572 (93)	4	0.5676
DYS462 (95)	6	0.5669
DYS537 (95)	3	0.5648
DYS594 (93)	5	0.5617
DYS391 (94)	4	0.5502
DYS531 (95)	6	0.5357

Locus	# Alleles	Diversity
DYS556 (93)	4	0.5346
DYS721 (95)	4	0.5234
DYS426 (91)	3	0.5221
DYS565 (95)	3	0.5165
DYS578 (95)	3	0.5165
DYS525 (93)	7	0.5157
DYS450 (91)	3	0.5070
DYS632 (94)	2	0.5017
DYS726 (94)	4	0.4907
DYS540 (94)	4	0.4871
DYS393 (94)	4	0.4770
DYS717 (95)	7	0.4531
DYS388 (91)	8	0.4498
DYS719 (94)	6	0.3606
DYS425 (95)	3	0.2278
DYS454 (95)	5	0.1957
DYS645 (95)	3	0.1820
DYS455 (95)	5	0.1781
DYS641 (94)	3	0.1219
DYS434 (95)	3	0.0824
DYS575 (94)	2	0.0213
DYS472 (95)	1	0.0000

Gene Diversity (D) = $(n/n-1)(1-\sum p_i^2)$, where n is the sample size and p_i is the allelic frequency

Is it possible to regularly resolve individuals from the same paternal lineage (e.g., fathers and sons) if enough Y-STRs are examined?

Examination of Father-Son Pairs

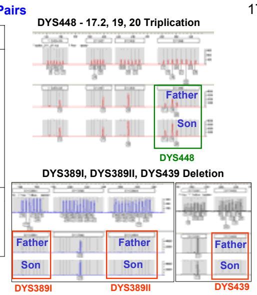
To-date, we have examined 350 father:son pairs (700 samples). Buccal swabs were extracted with DNA IQ, quantified with an Alu qPCR assay (Nicklas and Buel 2003), and typed with Identifiler and Yfiler STR kits to obtain information on 15 autosomal STRs and 17 Y-STRs. Autosomal allele sharing confirmed paternity.

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.

Mutations Seen in 100 African American Father-Son Pairs

Sample	Y-STR Locus	Allele (father)	Allele (child)	Comments
65B	Y GATA H4	11	9	loss of 2 repeats
46B	DYS389I and DYS389II	14,30	13,29	loss of 1 repeat
58B	DYS389I and DYS389II	14,32	15,33	gain of 1 repeat
18B	DYS390	24	23	loss of 1 repeat
90B	DYS456	15	16	gain of 1 repeat
16B	**DYS458**	18	19	gain of 1 repeat
39B	DYS458	18	19	gain of 1 repeat
16B	**DYS635**	23	22	loss of 1 repeat
47B	DYS635	22	23	gain of 1 repeat
72B	DYS635	22	23	gain of 1 repeat
22B	DYS448	19,20	19,20	Duplication
72B	DYS448	19,20	19,20	Duplication
97B	DYS448	17.2,19,20	17.2,19,20	Triplication
33B	DYS389I and DYS389II			Deletion
33B	DYS439			Deletion

Mutations in both DYS458 and DYS635 were seen in father and son 16B



Y-STR Mutation Rates

In almost 350 father:son sample pairs, we observed 19 differences between father and son with the 17 Y-STR loci in the Yfiler kit. Eight mutations resulted in the loss of a repeat in the son and 11 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	346	0.000%	0.158%
DYS389I ≠	11	5476	0.201%	5	346	1.445%	0.275%
DYS389II ≠	12	5463	0.220%	6	346	1.734%	0.310%
DYS390	16	6824	0.234%	1	342	0.292%	0.237%
DYS391	23	6702	0.343%	0	346	0.000%	0.326%
DYS392	4	6668	0.060%	0	346	0.000%	0.057%
DYS393	4	5456	0.073%	0	347	0.000%	0.069%
DYS385a/b	22	9980	0.220%	0	346	0.000%	0.213%
DYS438	1	2434	0.041%	0	346	0.000%	0.036%
DYS439	12	2409	0.498%	3	345	0.870%	0.545%
DYS437	5	2395	0.209%	0	345	0.000%	0.182%
DYS448	0	143	0.000%	0	343	0.000%	0.000%
DYS456	1	143	0.699%	1	345	0.290%	0.410%
DYS458	3	143	2.098%	3	346	0.867%	1.227%
DYS635	3	1016	0.295%	3	347	0.865%	0.440%
GATA-H4	3	1179	0.254%	2	345	0.580%	0.328%

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

≠ DYS389I is a subsection of DYS389II

Y-STR Loci in Use for Genetic Genealogy

The rapidly growing field of genetic genealogy is expanding the number of Y-STR loci in use.

http://www.isogg.org
http://www.familytreedna.com
http://www.dna-fingerprint.com
http://www.sorensongenomics.com
http://www.relativegenetics.com
http://www.ethnoancestry.com
http://www.oxfordancestors.com
http://www.dnaheritage.com
http://www.geogene.com

Family Tree DNA (12, 37, or 67 loci) – DYS19, DYS385 a/b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS439, DYS437, DYS447, DYS448, DYS449, DYS454, DYS455, DYS458, DYS459 a/b, DYS464 a/b/c/d, DYS438, DYS442, DYS460, GATA-H4, YCA II a/b, DYS456, DYS570, DYS576, DYS607, DYS724 a/b (CDY a/b), DYF395S1a/b, DYF406S1, DYS413 a/b, DYS425, DYS436, DYS444, DYS446, DYS450, DYS472, DYS481, DYS487, DYS490, DYS492, DYS511, DYS520, DYS531, DYS534, DYS537, DYS557, DYS565, DYS568, DYS572, DYS578, DYS590, DYS594, DYS617, DYS640, DYS641

Relative Genetics (18, 26, or 43 loci) – DYS19, DYS385 a/b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS426, DYS438, DYS439, DYS447, DYS448, DYS460, YCAII a/b, GATA-H4, DYS393, DYS437, DYS454, DYS455, DYS461, DYS462, GGAAT1B07, GATA-A10, DYS635, DYS441, DYS442, DYS444, DYS445, DYS446, DYS449, DYS452, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d

Ethnoancestry (18 loci) –DYS481, DYS487, DYS490, DYS494, DYS505, DYS522, DYS531, DYS533, DYS549, DYS556, DYS575, DYS578, DYS589, DYS594, DYS636, DYS638, DYS641, DYF406S1 + *Y-SNPs*

Oxford Ancestors (10 loci) – DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS425, DYS426

GeoGene (6 loci) – DYS19, DYS388, DYS390, DYS391, DYS392, DYS393 + *Y-SNPs*

9/26/2006: Family Tree DNA announced that their DNA databases contain **111,139** records: **3547** SURNAME PROJECTS, **55,610** unique surnames, **77,543** Y-DNA records in the database, **20,531** distinct 12-marker haplotypes, **25,921** distinct 25-marker haplotypes, **19,840** distinct 37-marker haplotypes, **33,596** mtDNA records in the database

Summary

- We are **examining new Y-STR loci** beyond those available in commercial kits.
- As expected, **more Y-STR loci increase the ability to resolve samples from one another** particularly those with a most common type.
- **Studies with father and son sample pairs are on-going** to measure mutation rates and to assist understanding which and how many Y-STRs may be optimal for differentiating between closely related individuals.

Acknowledgments and Disclaimer

Special thanks to Peter Vallone and David Duewer for their 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.

For more information, see STRBase:

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