




... working with industry to develop and apply technology, measurements and standards

Impact of Additional Y-STR Loci on Resolving Common Haplotypes and Closely Related Individuals

Peter M. Vallone
Amy E. Decker, Margaret C. Kline, John M. Butler


U.S. National Institute of Standards and Technology

September 30, 2006
V. International Forensic Y-User Workshop
Innsbruck, Austria




Presentation Outline

- Commercial Y-STR Kits and Beyond
- Common Y-STR Types
- Father-Son Samples
- Strategy for Selection of New Y-STR Loci



Going Beyond Commercial Y-STR Kits

- Most forensic DNA laboratories (certainly in the U.S.) will **only use commercially available kits** due to quality control issues
- Using these 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?
- Is it possible to regularly **resolve individuals from the same paternal lineage** (e.g., fathers and sons) if enough Y-STRs are examined?



Primary Commercial Y-STR Kits

Allele size range and locus dye colors

PowerPlex® Y
Released by Promega Corporation in Sept 2003

100 bp 200 bp 300 bp 400 bp

FL: DYS391, DYS389I, DYS439, DYS389II
 JOE: DYS438, DYS437, DYS19, DYS392
 TMR: DYS393, DYS390, DYS385a/b


3 dye colors
12-plex PCR

AmpFSTR® Yfiler™
Released by Applied Biosystems in Dec 2004

100 bp 200 bp 300 bp 400 bp


6-FAM: DYS456, DYS389I, DYS390, DYS389II
 VIC: DYS458, DYS19, DYS385a/b
 NED: DYS393, DYS391, DYS439, DYS635, DYS392
 PET: H4, DYS437, DYS438, DYS448

4 dye colors
17-plex PCR




Data Set Used to Examine Common Types

- Yfiler kit (**17 Y-STR loci**) run on all NIST male U.S. population samples
 - makes up ~20% of Applied Biosystems database
 - **submitted to the YHRD**
- Additional **20 Y-STR loci** run on full set of NIST population samples (*and several less polymorphic ones only on subset of samples*)
 - 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.



NIST Contribution to the YHRD

658 U.S. samples run with the Yfiler kit



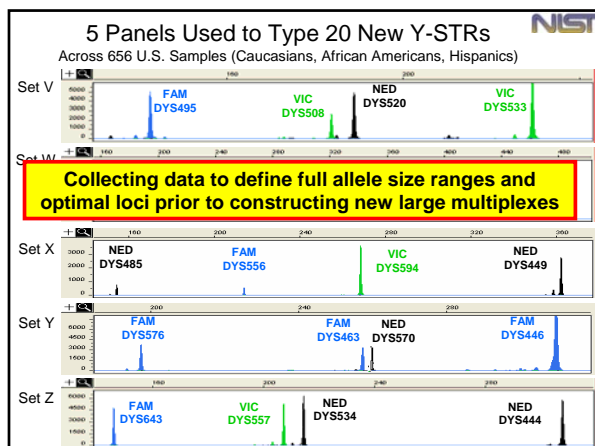
Address: <http://www.yhrd.org/index.html>

Search or list the YHRD contributors

Release 1.9:
41,965 Haplotypes in 357 populations

Contact: **John M. Butler, Amy E. Decker, Peter M. Vallone**
National Institute of Standards and Technology
Biotechnology Division
100 Bureau Dr., Mail Stop 8311
Gaithersburg MD 20899-8311
USA

Address: [redacted]
 EMail(s): jm.butler@nist.gov
 Phone: (301) 975-4049
 Fax: (301) 975-8505
 Haplotype count: 658
 Contrib. Populations: USA [African American], USA [European American], USA [Hispanic American]
 QC since: 20.12.2005
 Publication(s): Butler JM, Decker AE, Vallone PM, Kline MC (2005) Allele frequencies for 27 Y-STR Loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250-260.



Publication with Additional Y-STR Loci

Forensic Science International 156 (2006) 250-260

Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples

John M. Butler*, Amy E. Decker, Peter M. Vallone, Margaret C. Kline

Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8311, USA

Received 26 January 2005; received in revised form 22 February 2005; accepted 22 February 2005
Available online 30 March 2005

Full Haplotypes Available on <http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm>

# times haplotype observed	9
1	429
2	34
3	13
4	4
5	3
6	1
7	1
8	1
9	2
10	.
11	1
12	.
13	1
15	.
26	1

HD 0.996644
%DC 0.748476
HT 491

Total = 656 samples

With 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.

# times haplotype observed	9	11
1	429	486
2	34	33
3	13	10
4	4	6
5	3	1
6	1	1
7	1	2
8	1	.
9	2	.
10	.	1
11	1	.
12	.	.
13	1	.
15	.	1
26	1	.

HD 0.996644 0.998529
%DC 0.748476 0.824695
HT 491 541

Total = 656 samples

With the 11 loci of the SWGDAM haplotype run on 656 samples, 15 samples had the most common type

# times haplotype observed	9	11	12
1	429	486	505
2	34	33	34
3	13	10	14
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	.	.
15	.	1	.
26	1	.	.

HD 0.996644 0.998529 0.999064
%DC 0.748476 0.824695 0.853659
HT 491 541 560

Total = 656 samples

With the 12 loci of the PowerPlex Y haplotype (PPY) run on 656 samples, 12 samples had the most common type

505 of the 656 had a unique haplotype with the PPY loci, 34 sample haplotypes were observed twice in the sample set, 13 sample haplotypes were observed three times, etc.

# times haplotype observed	9	11	12	17
1	429	486	505	626
2	34	33	34	12
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	.	.	.
15	.	1	.	.
26	1	.	.	.

HD 0.996644 0.998529 0.999064 0.999916
%DC 0.748476 0.824695 0.853659 0.97561
HT 491 541 560 640

Total = 656 samples

With the 17 loci in Yfiler across the 656 samples, there are 626 unique haplotypes, 12 haplotypes that were observed twice and 2 haplotypes that were observed three times

# times haplotype observed	9	11	12	17	ALL 37
1	429	486	505	626	652
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				
15		1			
26	1				

When all 37 loci (Yfiler + 20 new loci) are run on 656 samples, only two haplotypes are observed twice

These two sets of three unseparated Yfiler types will be examined next

	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 samples

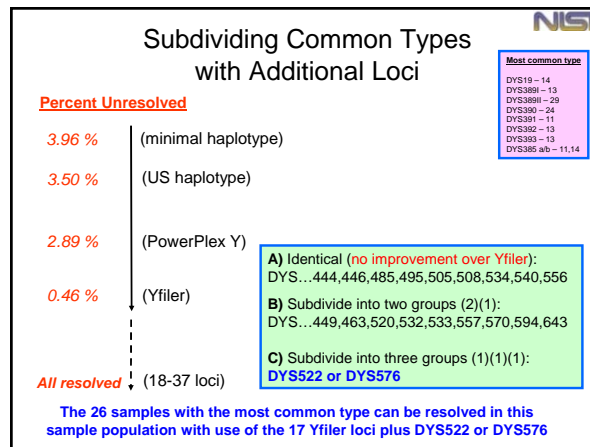
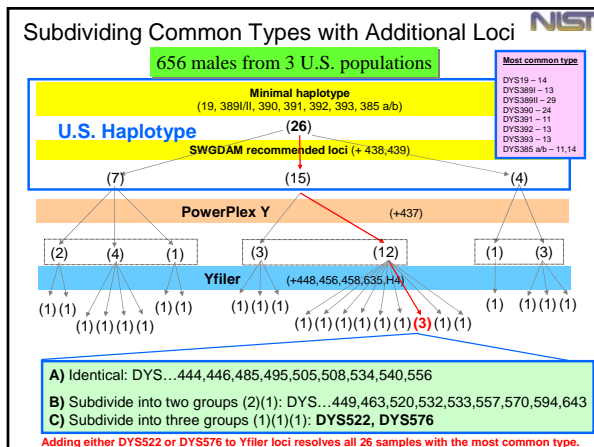
Subdividing Unresolved Yfiler Haplotypes (1)

Most Common Type

Sample Info	DYS 19	DYS 385a/b	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	H4
MT97185	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12
ZT79333	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12
TT51702	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12

Locus	MT97185	ZT79333	TT51702	Locus	MT97185	ZT79333	TT51702
DYS444	12	12	12	DYS532	14	14	13
DYS446	13	13	13	DYS533	13	12	13
DYS449	30	30	31	DYS534	15	15	15
DYS463	24	24	23	DYS540	12	12	12
DYS485	15	15	15	DYS556	11	11	11
DYS495	16	16	16	DYS557	15	17	17
DYS505	12	12	12	DYS570	16	17	17
DYS508	11	11	11	DYS576	17	20	18
DYS520	21	22	21	DYS594	9	10	10
DYS522	10	12	11	DYS643	10	11	10

Either DYS522 or DYS576 will fully resolve all three of these samples

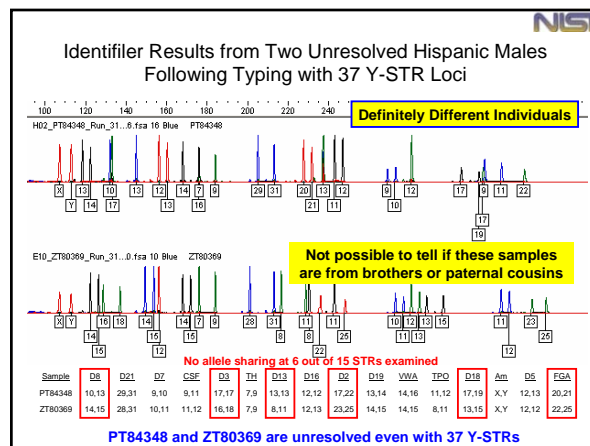


Subdividing Unresolved Yfiler Haplotypes(2)

Sample Info	DYS 19	DYS 385a/b	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	H4
PT83904	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12
PT84348	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12
ZT80369	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12

Locus	PT83904	PT84348	ZT80369	Locus	PT83904	PT84348	ZT80369
DYS444	12	12	12	DYS532	14	14	14
DYS446	12	12	12	DYS533	11	11	11
DYS449	31	31	31	DYS534	16	17	17
DYS463	16	16	16	DYS540	11	11	11
DYS485	15	15	15	DYS556	12	12	12
DYS495	12	12	12	DYS557	16	18	18
DYS505	11	11	11	DYS570	22	22	22
DYS508	11	11	11	DYS576	16	16	18
DYS520	19	19	19	DYS594	11	11	11
DYS522	12	12	12	DYS643	12	12	12

PT84348 and ZT80369 are unresolved even with 37 Y-STRs



NIST

Summary on Subdividing Common Types

- 640 haplotypes were observed in the 656 U.S. population samples with the Yfiler loci: 626 were unique, 2 were observed 3 times, and 12 haplotypes were observed twice.
- With the addition of 20 new Y-STR loci, all but two sample pairs are resolved.
- In this sample set, the 7 Y-STRs (**DYS532, DYS522, DYS576, DYS570, DYS505, DYS449, DYS534**) have the same ability to resolve the sample haplotypes as all 20 new loci.
- These 7 loci will be the focus of future studies and multiplex assays.

NIST

NIST Work with Father-Son Samples

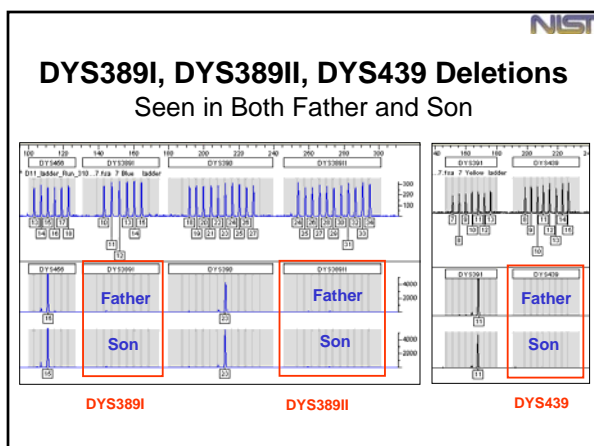
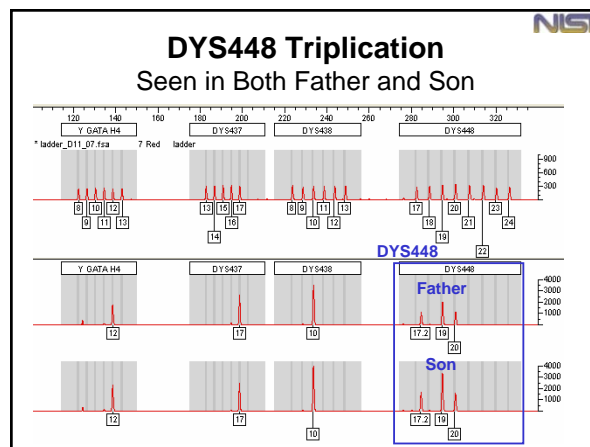
- Samples obtained from paternity testing laboratory as buccal swabs, extracted with DNA-IQ, quantified, diluted to 0.5 ng/uL
- To-date: **100 father-son pairs of U.S. Caucasian, African American, Hispanic, and Asian (800 samples)**
- Verified autosomal STR allele sharing with Identifier (QC for gender and potential sample switches)
- Typed with Yfiler (17 Y-STRs) – **examined mutations**

NIST

Mutations Seen in 100 African American Father-Son Pairs

Ethnicity	Sample	locus	Allele (father)	Allele (child)	Comments
African American	65B	Y GATA H4	11	9	loss of 2 repeats
African American	46B	DYS389I and DYS389II	14,30	13,29	loss of 1 repeat
African American	58B	DYS389I and DYS389II	14,32	15,33	gain of 1 repeat
African American	18B	DYS390	24	23	loss of 1 repeat
African American	90B	DYS456	15	16	gain of 1 repeat
African American	16B	DYS458	18	19	gain of 1 repeat
African American	39B	DYS458	18	19	gain of 1 repeat
African American	16B	DYS635	23	22	loss of 1 repeat
African American	47B	DYS635	22	23	gain of 1 repeat
African American	72B	DYS635	22	23	gain of 1 repeat
African American	22B	DYS448	19,20	19,20	Duplication
African American	72B	DYS448	19,20	19,20	Duplication
African American	97B	DYS448	17,2,19,20	17,2,19,20	Triplication *
African American	33B	DYS389I and DYS389II			Deletion *
African American	33B	DYS439			Deletion *

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

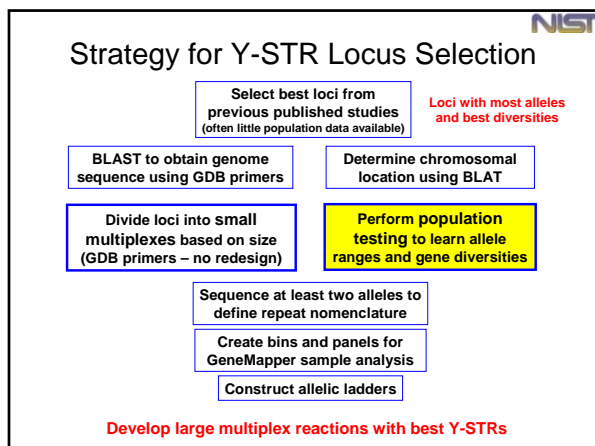


NIST

Y-STR Mutation Rates for Yfiler 17 Loci

Yfiler kit loci	Literature Summary *			NIST Results			TOTAL
	Mutations	# Meioses	Mutation Rate	Mutations	# Meioses	Mutation Rate	
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



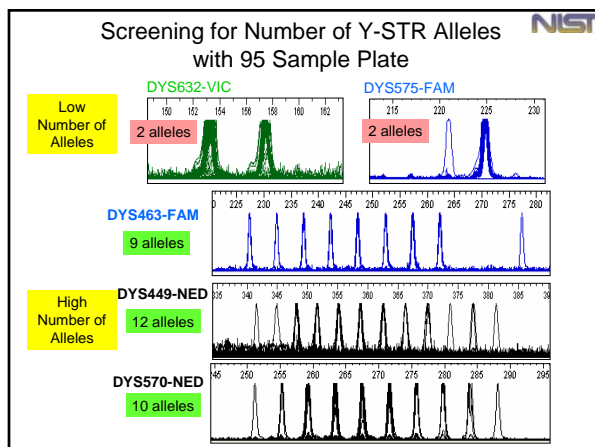
ARTICLE IN PRESS

Table 2
Gene diversity values for 19 loci selected from Y-chromosome sequence data

Locus	Number of alleles	Gene diversity
DYS710 ^a	15	0.90
DYS711 ^a	12	0.89
DYS626	10	0.85
DYS712 ^a	11	0.85
DYS713 ^a	10	0.80
DYS481	7	0.80
DYS818	7	0.79
DYS870	8	0.78
DYS714 ^a	8	0.77
DYS857	7	0.75
DYS864	8	0.74
DYS812	6	0.74
DYS607	7	0.70
DYS644	9	0.70
DYS715 ^a	5	0.64
DYS485	6	0.56
DYS716 ^a	3	0.55
DYS521	5	0.28
DYS717 ^a	3	0.24

Loci selected from:
Redd et al. 2002
Kayser et al. 2004
Leat et al. 2006

Samples from 46 Caucasians were analyzed using uniplex PCR followed by PAGE and silver staining. Gene diversity was calculated as $1 - \sum P_i^2$, where P_i is the allele frequency. The 14 loci in bold were selected for further analysis using dye-labeled primers and multiplex PCR.



- ### Conclusions
- 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

Funding from interagency agreement 2003-IJ-R-029 between the National Institute of Justice and the NIST Office of Law Enforcement Standards

NIST Human Identity Project Team

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

Tom Reid (DNA Diagnostics Center) – supplying the father-son samples