Yfiler and Beyond: Mutation Rates and New Loci for Increased Haplotype Resolution

Amy E. Decker,
Peter M. Vallone, Angela D. Gorman, John M. Butler

National Institute of Standards and Technology

60th Annual American Academy of Forensic Sciences
Washington, D.C.
February 23, 2008

Outline

- Summary of Yfiler mutation study
- The value of additional Y-STR loci
- SRM 2395 update
- Allele nomenclature

Mutation Rate of Y-STRs

- Determining mutation rates is important for the interpretation of typing results in human identity and paternity testing and evolutionary studies
- The average mutation rate for Y-STRs has been reported to be ~0.2% per generation

-Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Kruger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M and Sajantila, A. (2000) *American Journal of Human Genetics*, 66, 1580-1588.
-Dupuy, B.M., Stenersen, M., Egeland, T. and Olaisen, B. (2004) *Human Mutation*, 23, 117-124.

- **Mutations impact paternity testing and missing persons investigations** but not forensic direct evidence-suspect matches…

Father/son sample pairs

- 399 father/son pairs (798 total samples)
  - U.S. Caucasians, African Americans, Hispanics and Asians

Samples were provided by DNA Diagnostics Center (Fairfield, OH) as buccal swabs from previously analyzed paternity tests with a male child and an alleged father that was not excluded as a biological parent. These samples were distributed anonymously to NIST for testing.

DNA Extraction → DNA Quantification → PCR Amplification Identifiler/Yfiler™ kits

DNA IQ™ | Alu qPCR assay | Samples run with Identifiler to confirm samples were male and to check for autosomal allele sharing

Yfiler Loci Mutation Rates Measured at NIST

- **389 father/son sample pairs**
  - 788 samples with full profiles
- **17 Y-STR loci** in the Yfiler kit
- **24 differences** between father and son
  - 13 mutations resulted in the gain of a repeat in the son
  - 11 resulted in a loss of a repeat
- All single step repeat mutations
  - except a two repeat loss at Y-GATA-H4
- **2 sample pairs were found to have two mutations**
  - African American pair: mutations at DYS458 and DYS635
  - Asian pair: mutations at DYS439 and Y-GATA-H4
- Also observed 4 duplications, 1 triplication, and 4 deletions that were seen in both father and son

Decker, A.E., Kline, M.C., Redman, J.W., Reid, T.M., Butler, J.M. (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *FSI Genetics (in press)*

Mutation summary for father:son pairs with Yfiler loci

Yfiler kit loci	Literature Summary			NIST Results			
Locus	Mutations	# Meioses	Mutation Rate	Mutations	# Meioses	Mutation Rate	TOTAL
DYS19	22	9241	0.238%	1	389	0.257%	0.239%
DYS389I	14	7445	0.188%	5	389	1.285%	0.243%
DYS389II	22	7432	0.296%	6	389	1.542%	0.358%
DYS390	21	8723	0.241%	1	389	0.257%	0.241%
DYS391	25	8672	0.288%	0	389	<0.003%	0.276%
DYS392	5	8636	0.058%	0	389	<0.003%	0.055%
DYS393	6	7425	0.081%	0	389	<0.003%	0.077%
DYS385a/b	30	13765	0.218%	0	389	<0.003%	0.212%
DYS438	2	4075	0.049%	0	389	<0.003%	0.045%
DYS439	22	4052	0.543%	5	389	1.285%	0.608%
DYS437	6	3971	0.151%	0	389	<0.003%	0.138%
DYS448	1	557	0.180%	0	389	<0.003%	0.106%
DYS456	4	557	0.718%	1	389	0.257%	0.529%
DYS458	6	557	1.077%	4	389	1.028%	1.057%
DYS635	6	1430	0.420%	3	389	0.771%	0.495%
GATA-H4	4	1593	0.251%	3	389	0.771%	0.353%

Orange = highest mutation rate observed Blue = lowest mutation rate observed

Literature References for Mutation Rate Summary

- B. Berger, A. Lindinger, H. Niederstatter, P. Grubwieser, and W. Parson, Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. Int J Legal Med 119 (2005) 241-246.
- J. J. Mulero, C. W. Chang, L. M. Calandro, R. L. Green, Y. Li, C. L. Johnson, and L. K. Hennessy, Development and Validation of the AmpFISTR Yfilertrade mark PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System. J Forensic Sci. 51 (2006) 64-75.
- M. L. Pontes, L. Caine, D. Abrantes, G. Lima, and M. F. Pinheiro, Allele frequencies and population data for 17 Y-STR loci (AmpFISTR Yfiler)in a Northern Portuguese population sample. Forensic Sci. Int. 170 (2007) 62-67.
- S. Turrina, R. Atzei, and L. D. De, Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. Int. J. Legal Med. 120 (2006) 56-59.
- L. C. Tsai, T. Y. Yuen, H. M. Hsieh, M. Lin, C. H. Tzeng, N. E. Huang, A. Linacre, and J. C. Lee, Haplotype frequencies of nine Y-chromosome STR loci in the Taiwanese Han population. Int. J. Legal Med. 116 (2002) 179-183.
- Y Chromosome Haplotype Reference Database (YHRD) as of 03/23/07 www.YHRD.org

Value of Additional Loci

Value of Y-Chromosome Markers

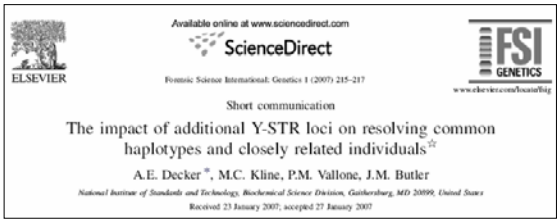
J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition; Table 9.1

Application	Advantage
Forensic casework on sexual assault evidence	**Male-specific amplification** (can avoid differential extraction to separate sperm and epithelial cells)
Paternity testing	Male children can be tied to fathers in motherless paternity cases
Missing persons investigations	**Patrilineal male relatives may be used for reference samples**
Human migration and evolutionary studies	Lack of recombination enables comparison of male individuals separated by large periods of time
Historical and genealogical research	Surnames usually retained by males; can make links where paper trail is limited

Why go beyond Yfiler?

- **Yfiler** (17 loci)
  DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 and Y-GATA-H4
- **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

Are there advantages to typing additional loci beyond the PowerPlex Y 12 or the Yfiler 17 Y-STRs?



Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 215–217
ELSEVIER
FSI GENETICS
www.elsevier.com/locate/fsig

Short communication

The impact of additional Y-STR loci on resolving common haplotypes and closely related individuals☆

A.E. Decker*, M.C. Kline, P.M. Vallone, J.M. Butler
National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, MD 20899, United States

Received 23 January 2007; accepted 27 January 2007

Full 37 locus haplotypes available on STRBase:
http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/HispanicsHaplotype37.pdf
http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/CaucasiansHaplotype37.pdf
http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/AfricanAmericansHaplotype37.pdf

# times haplotype observed	12 loci PPY	17 loci Yfiler	37 loci ALL 37	NIST U.S. Pop (C, AA, H) Total # samples: 656
1	505	626	652	unique
2	34	12	2	unresolved haplotypes
3	14	2	.	
4	3	.	.	
5	2	.		
6	.	.		
7	1	.		
8	.	.		
9	.	.		
10	.	.		
11	.	.		
12	1	.	.	
HD	0.99906	0.99992	0.99999	
DC	0.85366	0.97561	0.99695	
# HT	560	640	654	

One set of three unseparated Yfiler types will be examined next

Most common type is observed 12 times

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

95% (626/656) Yfiler haplotypes were unique

Subdividing Unresolved Yfiler Haplotypes

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

Lessons Learned from NIST Data Set

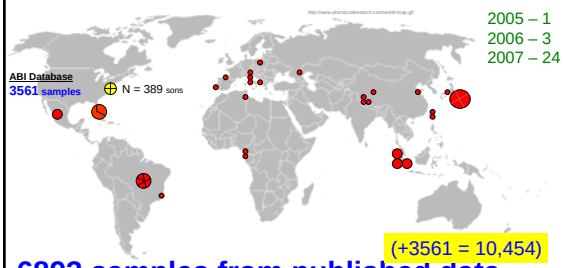
- Some Y-STRs are more useful than others in sub-dividing common haplotypes (e.g., DYS576)
- You don't gain much by typing additional Y-STRs (most unresolved types only occur twice)
- 95% of 17 locus Yfiler haplotypes are unique

What additional population data exists with Yfiler?

And how does it compare to our NIST data?

Sources of Yfiler Worldwide Population Data

28 published population studies with Yfiler data



(+3561 = 10,454)

6893 samples from published data

6514 haplotypes (discrimination capacity 94.5%)

6257 unique haplotypes **(96.0% unique)**

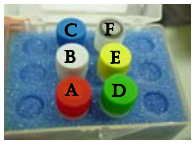
Summaries of Recent Worldwide Yfiler Data

- **10,454 Yfiler profiles** now available
  - 3561 current Yfiler database + 6893 published data
- ~95% of the time a complete 17 locus Yfiler profile will be unique
- However, just like mtDNA, common types do exist so many of the remaining Yfiler haplotypes are shared (present in multiple individuals)

For full list of sources: Butler, J.M., Hill, C.R., Decker, A.E., Kline, M.C., Reid, T.M., Vallone, P.M. (2007) New autosomal and Y-chromosome STR loci: characterization and potential uses. Proceedings of the Eighteenth International Symposium on Human Identification. See http://www.promega.com/geneticidproc/

NIST SRM 2395 update

Y-Chromosome Standard NIST SRM 2395



Human Y-Chromosome DNA Profiling Standard

- **5 male samples** + 1 female sample (neg. control)
- **100 ng of each (50 µL at ~2 ng/µL)**
- **42 Y SNPs typed** with Marligen kit

$316

- **41 Y STR markers now typed and sequenced**

Certified for all loci in commercial Y-STR kits:

PowerPlex Y
Yfiler

SWGDAM recommended loci: DYS19, DYS385 a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Helps meet FBI Standard 9.5 (and ISO 17025)…traceability to a national standard

Y-STR Loci Being Added to NIST SRM 2395 Certificate

- DYS635 (to complete Yfiler loci)
- **DYS449**
- **DYS481**
- **DYS570**
- **DYS576**

These 19 Y-STR loci have been sequenced across all 5 male components in SRM 2395

- DYS492, DYS522, DYS527, DYS532, DYS534, DYS572, DYS607, DYS650, DYS652, DYS709, DYS710, DYS712, DYS715, DYS717

- **Also examining nomenclature on additional Y-STR loci used by genetic genealogy companies**

http://www.cstl.nist.gov/biotech/strbase/srm2395.htm

DYS635 Information for SRM 2395 Update

Component	Type	DNA Sequence of STR Repeat Region
A	**23**	$(TCTA)_4$ $(TGTA)_2$ $(TCTA)_2$ $(TGTA)_2$ $(TCTA)_2$ $(TGTA)_2$ $(TCTA)_9$
B	**21**	$(TCTA)_4$ $(TGTA)_2$ $(TCTA)_2$ $(TGTA)_2$ $(TCTA)_{11}$
C	**23**	$(TCTA)_4$ $(TGTA)_2$ $(TCTA)_2$ $(TGTA)_2$ $(TCTA)_{13}$
D	**21**	$(TCTA)_4$ $(TGTA)_2$ $(TCTA)_2$ $(TGTA)_2$ $(TCTA)_{11}$
E	**21**	$(TCTA)_4$ $(TGTA)_2$ $(TCTA)_2$ $(TGTA)_2$ $(TCTA)_{11}$

- We have added information to SRM 2395 certificate for DYS635 as it is present in the Yfiler kit (released Dec 2004)

SRM 2395 Male Component Typing Results Certified Values

SRM 2395	DYS 19	DYS 388	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385a/b	DYS 438	DYS 439
A	14	12	13	29	25	11	13	13	12-15	12	12
B	14	15	13	28	23	11	11	12	14-17	9	12
C	16	12	14	32	21	12	11	13	17-20	11	11
D	15	12	12	28	22	10	11	14	14-15	11	11
E	17	13	14	31	24	10	12	14	13-15	10	11

SRM 2395	DYS 426	DYS 435	DYS 436	DYS 437	DYS 447	DYS 448	DYS 456	DYS 458	DYS 460	DYS 461	DYS 635	H4
A	12	12	12	15	24	19	15	16	11	12	23	12
B	11	11	12	14	25	21	15	15	10	13	21	12
C	11	11	12	14	25	21	15	17	9	13	23	12
D	11	11	12	16	23	21	15	16	11	11	21	12
E	11	11	12	14	26	20	15	16	11	12	21	11

Yfiler loci in blue

Results with New Y-STR Loci

SRM 2395	DYS 449	DYS 481	DYS 522	DYS 532	DYS 534	DYS 570	DYS 576	DYS 607	DYS 715	DYS 717
A	28	22	10	15	15	17	18	15	14	16
B	32	23	11	11	15	18	16	15	11	16
C	30	28	10	12	15	18	17	15	12	16
D	28	23	12	15	14	17	18	13	13	19
E	27	28	12	9	14	18	17	14	12	13

SRM 2395	DYS 492	DYS 527	DYS 572	DYS 650	DYS 652	DYS 709	DYS 710	DYS 712
A	12	21,23	11	18	24	13	36	23
B	12	22	11	18	25	16	34.2	22.3
C	11	17,20	9	16	25	15	35.2	21
D	11	15,21	9	18	23	17	33.2	26
E	12	22,23	11	24	26	16	31	19

Nomenclature Issues for Y STR Alleles

Why do we need SRM 2395?

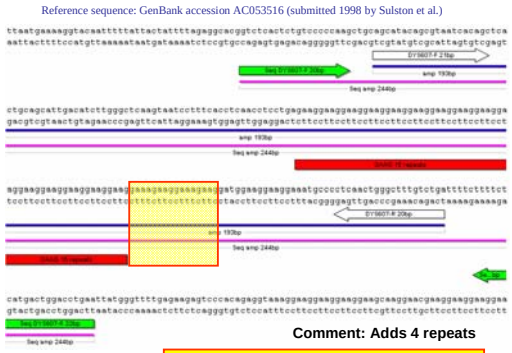
- Differences in nomenclature can lead to confusion or problems with database matches… **NIST SRMs are viewed positively for aiding nomenclature disputes**
- **Standardizing the nomenclature** and encouraging (or requiring) all testing laboratories to calibrate their results to this standard will aid in nomenclature issues

Categories for Y-STR markers included in SRM 2395 update

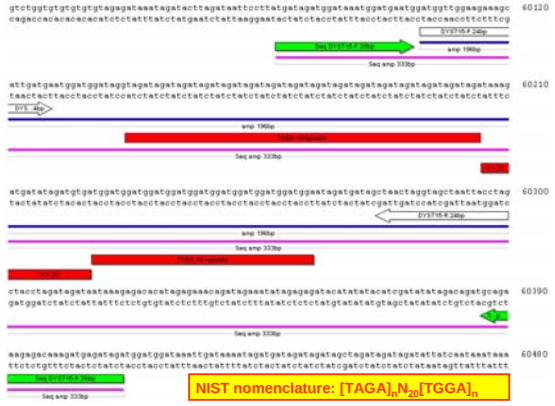
Category	Example Repeat Structure	21 Y-STR loci
simple repeats	(GATA)(GATA)(GATA)	DYS456, DYS458, DYS481, DYS492, DYS522, DYS532, DYS534, DYS570, DYS572, DYS576
simple repeats with non-consensus alleles	(GATA)(GAT-)(GATA)	DYS712 Component B
compound repeats	(GATA)(GACA)(GATA)	DYS527, DYS607, DYS635, DYS650, DYS652, DYS712, DYS717
complex repeats	(GATA)(GACA)(CA)(CATA)	DYS710
repeats containing non-variable non-repetitive region	(GATA)N_n(GATA)	DYS449, DYS715, Y-GATA-H4

Categories based on paper by **Urquhart *et al.* (1994) *Int. J. Legal Med.* 107:13-20**

DYS607



DYS715



ISFG Guidelines for Y-STR Allele Nomenclature

Gill et al. (2001) *Forensic Sci. Int*. 124: 5-10

- Number of complete repeats
- A partial repeat (variant allele) is designated by number of complete repeats separated by a dot followed by the number of bases in the incomplete repeat (e.g., 17.3)
- Some locus nomenclatures take into account the total number of repetitive units (non-variant plus variant) while others have taken into account only the variable repetitive stretches
  - "If a nomenclature is already in use, it is recommended that it should be continued. However, to encourage consistency for newly reported STRs, it is recommended that alleles should be named according to the total number of repeat units of the DNA that comprises both variant and non-variant repeats"
- Duplicated systems such as DYS385 have to be treated as genotypes and alleles should be separated by a hyphen (e.g., "11-14")

Attempts to Resolve Nomenclature Issues

- Mulero, J.J., Budowle, B., **Butler, J.M.**, Gusmão, L. (2006) Letter to the Editor--Nomenclature and allele repeat structure update for the Y-STR locus GATA H4. *J. Forensic Sci.* 51(3): 694.
- 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.

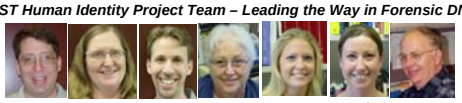
Summary

- We have examined mutation rates in 389 father/son pairs with Yfiler
- Yfiler does a good job at resolving unrelated male haplotypes (95% unique for NIST sample set which is comparable to published data). Simply adding more loci does not guarantee more unique haplotypes (certain loci will be more useful than others).
- We have characterized new Y-STR loci and reviewed nomenclature to update SRM 2395 which will assist genetic genealogy companies looking beyond commercially available loci

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 – Leading the Way in Forensic DNA…



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

Tom Reid (DNA Diagnostics Center) – supplying the father-son samples for mutation rate analysis