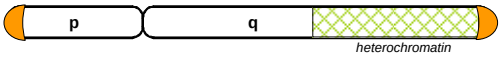


Y-STRs:
Why look at Y's?

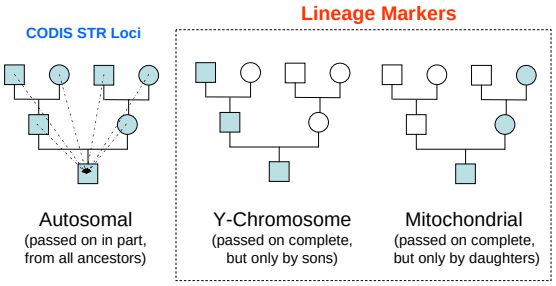
Amy E. Decker
National Institute of Standards and Technology

The First Annual Technological Advances in Human Identification
April 9, 2008

Presentation Outline

- Why Y is of interest in human identity testing
- Y-STR markers and kits available
- Different population databases and statistics for reporting matches
- Mutation rates, duplications, and deletions and their impact on interpretation

Different Inheritance Patterns



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Role of Y-STRs and mtDNA
Compared to Autosomal STRs

- **Autosomal STRs provide a higher power of discrimination and <u>are the preferred method</u> whenever possible**
- **Due to capabilities for male-specific amplification**, Y-chromosome STRs (**Y-STRs**) can be useful in extreme female-male mixtures (e.g., when differential extraction is not possible such as fingernail scrapings)
- **Due to high copy number**, mitochondrial DNA (**mtDNA**) may be the only source of surviving DNA in highly degraded specimens or low quantity samples such as hair shafts

A mtDNA result is better than no result at all…

Lineage Markers:
Y-STRs and mtDNA

<u>Advantages</u>

- Extend possible reference samples beyond a single generation (benefits missing persons cases and genetic genealogy)
- Family members have indistinguishable haplotypes unless mutations have occurred
- Potential to aid in familial searching to exclude partial matches from non-paternal relatives

<u>Disadvantages</u>

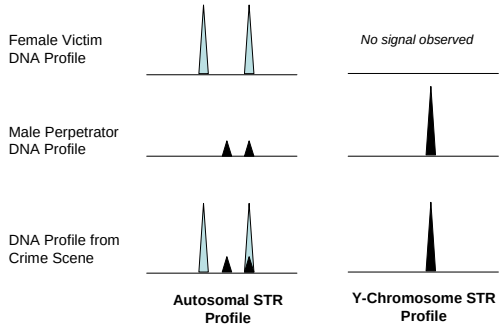
- Lower power of discrimination due to no genetic shuffling with recombination
- Family members have indistinguishable haplotypes unless mutations have occurred

Value of Y-Chromosome Markers

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

<u>Application</u>	<u>Advantage</u>
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

Y-STRs can permit simplification of male DNA identification in sexual assault cases

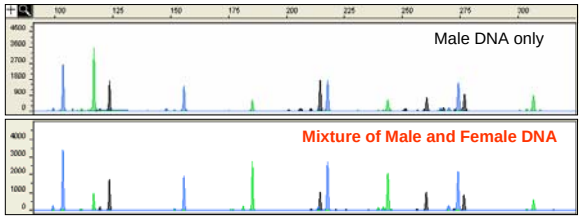


Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.2, ©Elsevier Science/Academic Press

Forensic Advantages of Y-STRs

- **Male-specific amplification** extends range of cases accessible to obtaining probative DNA results (e.g., fingernail scrapings, sexual assault without sperm)
- **Technical simplicity due to single allele profile**; can potentially recover results with lower levels of male perpetrator DNA because there is not a concern about heterozygote allele loss via stochastic PCR amplification; number of male contributors can be determined
- **Courts have already widely accepted STR typing**, instrumentation, and software for analysis (Y-STR markers just have different PCR primers)

Y-STRs Identify the Male Component even with Excess Female DNA



800X female DNA

Scenarios Where Y-STRs Can Aid Forensic Casework

- Sexual assaults by vasectomized or azoospermic males (no sperm left behind for differential extraction)
- Extending length of time after assault for recovery of perpetrator's DNA profile (greater than 48 hours)
- Fingernail scrapings from sexual assault victims
- Male-male mixtures
- Other bodily fluid mixtures (blood-blood, skin-saliva)
- Gang rape situation to include or exclude potential contributors
- Confirmation of amelogenin Y negative males

Confirmation of Amelogenin Negative Males

Normal AMEL X,Y male

AMEL X only

- **Often due to deletion of that entire region of the Y-chromosome rather than a primer binding site mutation**
- Most commonly seen in males of Indian subcontinent origin
- **Y-STRs help demonstrate that the AMEL X sample is really male**
- Chang *et al.* (2007) *Forensic Sci. Int.* 166: 115-120
  - 12/649 Malaysian males showed no AMEL Y
- Cadenas *et al.* (2007) *Forensic Sci. Int.* 166: 155-163
  - 5/77 Nepal males showed no AMEL Y

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

Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)
- **Paternal lineages possess the same Y-STR haplotype** (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another
- **Not as informative as autosomal STR results**
  - **More like addition (10 + 10 + 10 = 30) than multiplication (10 x 10 x 10 = 1,000)**

What has happened in the past few years…

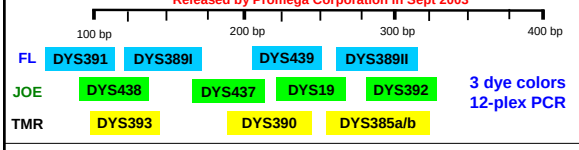
- "Full" Y-chromosome sequence became available in June 2003; over 350 Y-STR loci identified (only ~20 in 2000)
- **Selection of core Y-STR loci** (SWGDAM Jan 2003)
- **Commercial Y-STR kits released**
  - ~~Y-PLEX 6,5,12 (2001-03)~~, **PowerPlex Y** (9/03), **Yfiler** (12/04)
- Many population studies performed and databases generated with thousands of Y-STR haplotypes
  - U.S. consolidated Y-STR database (13,906 haplotypes with 11 loci) http://www.usystrdatabase.org/
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

Primary Commercial Y-STR Kits

Allele size range and locus dye colors

PowerPlex® Y

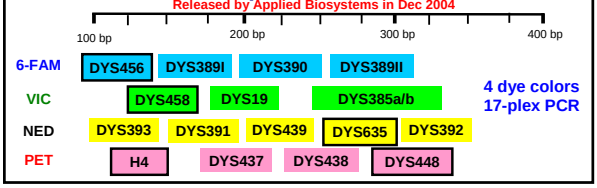
Released by Promega Corporation in Sept 2003



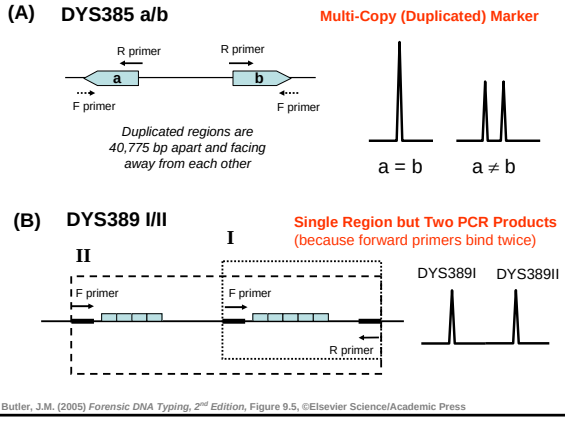
3 dye colors
12-plex PCR

AmpF/STR® Yfiler™

Released by Applied Biosystems in Dec 2004



4 dye colors
17-plex PCR

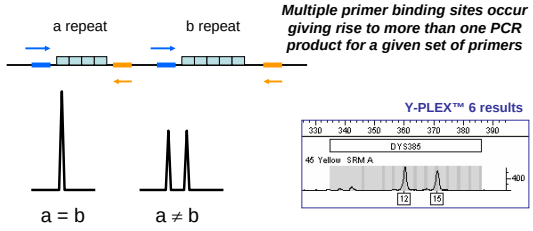


Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.5, ©Elsevier Science/Academic Press

Y-STR Typing of Duplicated Regions

"multi-copy loci"

Multiple primer binding sites occur giving rise to more than one PCR product for a given set of primers

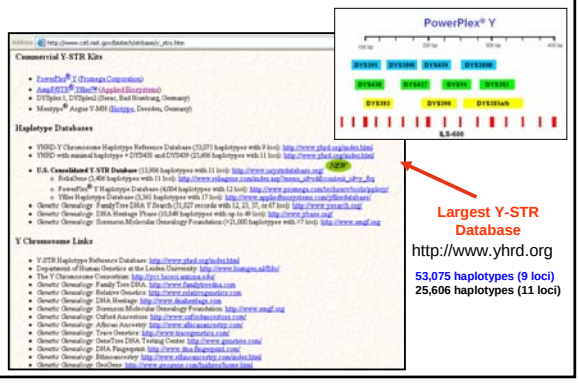


DYS385 a/b and YCAII a/b

Y-STR loci are often counted by the number of amplicons rather than the number of PCR primer pairs

Y-STR Databases and Statistics

Y-Chromosome Information Resources on the NIST STRBase Website



Largest Y-STR Database

http://www.yhrd.org

53,075 haplotypes (9 loci)
25,606 haplotypes (11 loci)

Y-Chromosome Haplotype Reference Database (YHRD)

http://www.yhrd.org

(477 populations)

As of 1/15/08: **53,075 haplotypes**

25,606 haplotypes
with all US required loci

Commercial Y-STR kits exist to amplify all of the core loci in a single reaction (plus a few additional markers)

Run only with minimal haplotype

DYS19
DYS389I/II
DYS390
DYS391
DYS392
DYS393
DYS385 a/b

US haplotype requires 2 additional loci:
DYS438
DYS439

Haplotype Databases for Y-STR Kits

http://www.promega.com/techserv/tools/pplexy/

http://www.appliedbiosystems.com/yfilerdatabase/

http://www.reliagene.com/index.asp?menu_id=rd&content_id=y_frq

This data has now been consolidated into the U.S. Y-STR Database

Power

1311
325
894
1108
366 Native Americans

4,004 total
(as of March 2005)

106 Native Americans
105 Filipino
59 Sub-Saharan Africans
103 Vietnamese

3,561 total
(as of December 2004)

3,406 total

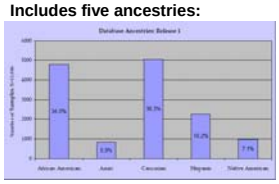
U.S. Y-STR Database

US Y-STR

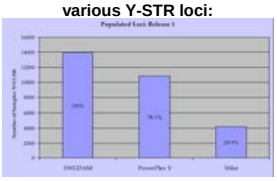
http://www.usystrdatabase.org/newdefault.aspx

As of 12/31/07: **13,906 haplotypes with 11 loci**

Includes five ancestries:



Percentage of samples run with various Y-STR loci:



13,906 samples with 11-locus SWGDAM haplotype
10,865 samples with 12-locus Powerplex Y haplotype
4,163 samples with 17-locus Yfiler haplotype

Example Y-STR Haplotype

Core US Haplotype

- DYS19 – 14
- DYS389I – 13
- DYS389II – 29
- DYS390 – 24
- DYS391 – 11
- DYS392 – 14
- DYS393 – 13
- DYS385 a/b – 11,15
- DYS438 – 12
- DYS439 – 13

Matches by Databases

- YHRD (9 loci)
  - 17 matches in 53,075
- YHRD (11 loci)
  - 2 matches in 26,586
- US Y-STR (11 loci)
  - 0 matches in 13,906

Searches performed 3/20/08

Y-Chromosome Haplotype Reference Database

www.YHRD.org

Release "23" from 2008-01-15 14:44:25

17 matches in 53,075 individuals from 442 worldwide populations

Minimal Haplotype Result

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 14
DYS393 – 13
DYS385 a/b – 11,15

US Y STR Database Search Results (with 11 loci)

0 matches in 13,906

Ancestry	# of Haplotypes	Number of Haplotypes (with Selected Alleles)	Frequency	Frequency Upper Bound (95%)
African American	4796	0	0.000000	0.000624
Asian	820	0	0.000000	0.003646
Caucasian	5047	0	0.000000	0.000593
Hispanic	2260	0	0.000000	0.001324
Native American	983	0	0.000000	0.003042
Total	13906	0	0	0.000215

Overall Database Summary:

The selected haplotype is found in 0 of 13906 total individuals within the database with a frequency of 0. Applying the 95% upper confidence interval results in a frequency of 0.000215, which is equivalent to approximately 1 in every 4651 individuals.

The selected haplotype is found in 0 of 4796 **African American** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.000624, which is equivalent to approximately 1 in every 1603 individuals.

The selected haplotype is found in 0 of 820 **Asian** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.003646, which is equivalent to approximately 1 in every 274 individuals.

The selected haplotype is found in 0 of 5047 **Caucasian** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.000593, which is equivalent to approximately 1 in every 1686 individuals.

The selected haplotype is found in 0 of 2260 **Hispanic** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.001324, which is equivalent to approximately 1 in every 755 individuals.

The selected haplotype is found in 0 of 983 **Native American** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.003042, which is equivalent to approximately 1 in every 329 individuals.

Frequency Estimate Calculations

In cases where a Y-STR profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

$p = X/N$

17 matches in 53,075

p = 17/53,075 = 0.00032 = **0.032%**

An upper bound confidence interval can be placed on the profile's frequency using:

$$p + 1.96\sqrt{\frac{(p)(1-p)}{N}}$$

$$0.00032 + 1.96\sqrt{\frac{(0.00032)(1-0.00032)}{53{,}075}}$$

= 0.00032 + 0.000152 = 0.000472

= **0.047% (1 in 2118)**

When there is no match…

In cases where the profile has not been observed in a database, the upper bound on the confidence interval is

$1-\alpha^{1/N}$

0 matches in 13,906

where α is the confidence coefficient (0.05 for a 95% confidence interval) and N is the number of individuals in the database.

$1-\alpha^{1/N} = 1-(0.05)^{[1/13{,}906]} = 0.000215$
= **0.022% (1 in 4651)**

A simplified calculation would be 3/N.

In this example: 3/13906 = 0.000216 = 0.022% (1 in 4630)

The Meaning of a Y-Chromosome Match

Conservative statement for a match report:

The Y-STR profile of the crime sample matches the Y-STR profile of the suspect (at *xxx* number of loci examined). Therefore, ***we cannot exclude the suspect*** as being the donor of the crime sample. In addition, we cannot exclude all patrilineal related male relatives and an unknown number of unrelated males as being the donor of the crime sample.

Y-STR Mutations

Mutations will impact kinship testing involving Y-STRs

(e.g., use of a paternal relative as a reference for a missing persons case)

Probability of Finding No Mutation or at Least One Mutation Between Two Y-STR Haplotypes in a Single Generation

Using average mutation rate of 0.28% (Kayser et al. AJHG 2000, 66:1580-1588)

# STRs	Prob. no mutation	Prob. at least one mutation
1	0.99720000	0.00280000
2	0.99440784	0.00559216
3	0.99162350	0.00837650
4	0.98884695	0.01115305
5	0.98607818	0.01392182
6	0.98331716	0.01668284
7	0.98056387	0.01943613
8	0.97781829	0.02218171
9	0.97508040	0.02491960
10	0.97235018	0.02764982
11	0.96962760	0.03037240
12	0.96691264	0.03308736
…		
40	0.89390382	**0.10609618**

3.3% with 12 Y-STRs

Gusmão, L., Butler, J.M., et al. (2006) *Forensic Sci. Int.* 157:187-197

NIST Work with Father-Son Samples

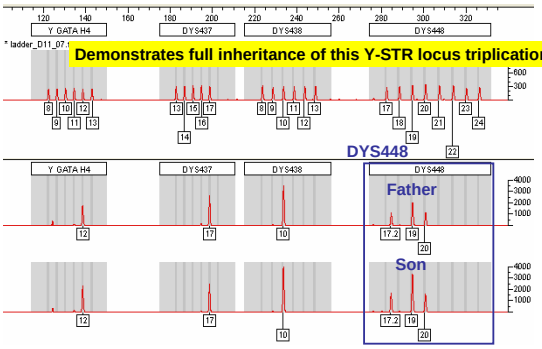
- Samples obtained from paternity testing laboratory as buccal swabs, extracted with DNA-IQ, quantified, diluted to 0.5 ng/uL
- 399 father/son pairs (798 total samples)
  - U.S. Caucasians, African Americans, Hispanics and Asians
- **Verified** autosomal STR allele sharing **with Identifiler** (QC for gender and potential sample switches)
- Typed with Yfiler (17 Y-STRs) – **examined mutations**

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

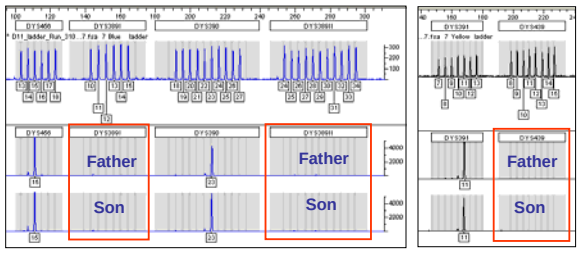
Decker, A.E., Kline, M.C., Redman, J.W., Reid, T.M., Butler, J.M. (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *FSI Genetics (in press)*

DYS448 Triplication
Seen in Both Father and Son



DYS389I, DYS389II, DYS439 Deletions
Seen in Both Father and Son

Yfiler data



DYS389I DYS389II DYS439

Full inheritance of these Y-STR locus deletions

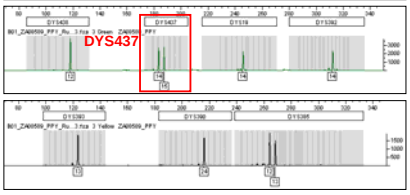
Duplication at Multiple Loci with Single-Source Sample

PowerPlex Y data

DYS389I DYS439 DYS389II

Y-chromosome mapping

Y STR Marker	Position (Mb)
DYS391	13.413
DYS635 (C4)	13.690
DYS434	13.777
DYS437	13.778
DYS435	13.807
DYS439	13.826
DYS389 I/II	13.923
DYS388	14.057
DYS442	14.071
DYS438	14.248

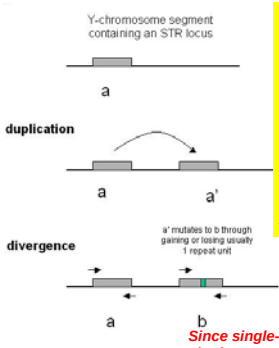


Entire region of Y-chromosome has likely been duplicated and then diverged

Most duplications have a single repeat spread in allele patterns

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Duplication and Divergence Model



Locus	# dup*	>1 repeat
DYS19	23	2
DYS389I	5	0
DYS389II	9	2
DYS390	1	0
DYS391	3	1
DYS392	0	0
DYS393	3	0
DYS385a/b	17	0

*from www.yhrd.org, literature, and our work

92% have single repeat difference

Since single-step mutations are most common, then single repeat spacing in duplicated alleles is expected

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Deciphering between a Mixture of Multiple Males and Locus Duplication

- ***Note the number of loci containing >1 allele (other than multi-copy DYS385)***
- ***Consider relative position on the Y-chromosome if multiple loci have two alleles***
- ***See if repeat spread is >1 repeat unit***
- ***Examine DYS385 for presence of >2 alleles***

Locus duplication along the Y-chromosome is in many ways analogous to heteroplasmy in mitochondrial DNA, which depending on the circumstances can provide greater strength to a match between two DNA samples.

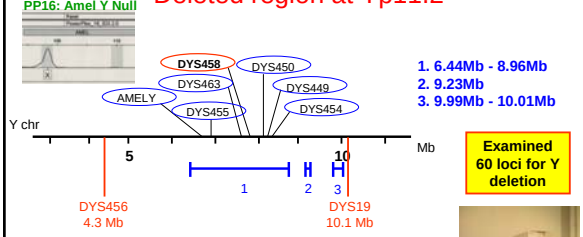
Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Practical Information on Y Deletions

- If DYS458 is deleted in Yfiler, then your sample is likely to lack an Amelogenin Y amplicon as DYS458 and AMEL Y are 1.13 Mb apart on the short arm of the human Y-chromosome
  - Chang *et al.* (2007) *Forensic Sci. Int.* 166: 115-120
- Many Y-chromosomes are more complicated than originally thought!

Y Chromosome Deletions

Deleted region at Yp11.2

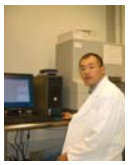


1. 6.44Mb - 8.96Mb
2. 9.23Mb
3. 9.99Mb - 10.01Mb

Examined 60 loci for Y deletion

- Identified 3 deleted regions in an Amel Y negative male from Gifu, Japan
- This pattern has not been reported before

Tomohiro Takayama Ph.D. - Research Specialist from Criminal Investigation Laboratory, Gifu Pref. Police H.Q., Japan
Worked at NIST from Sept-Nov. 2007



Literature summary of AMELY null allele

Country	Population	No. of null/no. individuals	Frequency (%)	Reference
Sri Lanka	Sri Lankan	2/24	8.3	Santos et al (1998), Jobling et al. (2007)
India	Indian	5/270	1.9	Thangaraj et al. (2002)
India (whole)	Indian (caste and tribes)	10/4,257	0.23	Kashyap et al (2006)
Nepal	Nepalese	5/77	6.5	Cadenas et al (2006)
	Nepalese	9/769	1.2	Parkin et al (2007), Jobling et al (2007)
Austria	Austrian	5/28,182	0.018	Steinlechner et al (2002)
Italy	Italian	1/13,000	0.008	Lattanzi et al. (2005), Jobling et al. (2007)
Spain	Spanish	1/768	0.13	Bosch et al. (2002), Jobling et al. (2007)
Israel		1/96	1.0	Michael and Brauner (2004)
Australia	Mixed	22/109,000	0.02	Mitchell et al. (2006), Jobling et al. (2007)
Malaysia	Indian	10/315	3.2	Chang et al. (2007),
	Malay	2/334	0.6	Jobling et al. (2007), Yong, Gan, Chang et al. (2007)
	Chinese	0/331	0	
Singapore	Indian	3/175	1.76	Yong, Gan, Coble et al. (2007),
	Malay	1/182	0.6	Yong, Gan, Chang et al. (2007)
	Chinese	0/210	0	

Yong RY, Gan LS, Chang YM, Yap EP. Molecular characterization of a polymorphic 3-Mb deletion at chromosome Yp11.2 containing the AMELY locus in Singapore and Malaysia populations.Hum Genet. 2007 Nov;122(3-4):237-49.

Key References for Y-STR Information

- Balding, DJ. (2005) *Weight-of-Evidence for Forensic DNA Profiles.* John Wiley & Sons, Hoboken, NJ, pp. 99-101.
- Buckleton J, Walsh S, Harbison S. (2005) Chapter 9, "Nonautosomal Forensic Markers" In *Forensic DNA Evidence Interpretation*. (eds., Buckleton J, Triggs CM, Walsh SJ) CRC Press, Boca Raton, FL, pp.299-331.
- Budowle B, Jianye G, Chakraborty R. (2007) Basic principles for estimating the rarity of Y-STR haplotypes derived from forensic evidence. Proceedings of the Eighteenth International Symposium on Human Identification. http://www.promega.com/geneticidproc/ussymp18proc/oralpresentations/Budowle.pdf
- Butler JM. (2003) Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *Forensic Sci. Rev*. 15:91-111.
- Butler, J.M. and Schoske, R. (2004) Duplication of DYS19 flanking regions in other parts of the Y chromosome. *Int. J. Legal Med.*, 118: 178-183.
- Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C. DNA Commission of the International Society of Forensic Genetics: Recommendations on forensic analysis using Y-chromosome STRs. *Int. J. Legal Med*. 114, 305-9 (2001) and *Forensic Sci. Int.,* 124, 5-10 (2001).

Key References for Y-STR Information

- Gusmão, L., Butler, J.M., Carracedo, A., Gill, P., Kayser, M., Mayr, W.R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C., Schneider, P.M. (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. *Int. J. Legal Med.* 120: 191-200.
- Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, Walsh B, Redd AJ. (2006) Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci. Int.* 164(1):45-55.
- Krenke BE, Viculis L, Richard ML, Prinz M, Milne SC, Ladd C, Gross AM, Gornall T, Frappier JRH, Eisenburg AJ, Barna C, Aranda XG, Adamowicz M, Budowle B (2005) Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci. Int.* 148:1-14.
- Mulero JJ, Chang CC, Calandro LM, Green RL, Li Y, Johnson CC, et al. (2006) Development and validation of the AmpFlSTRs YfilerTM PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J. Forensic Sci*. 51(1):64–75.
- Prinz M, Ishii A, Coleman A, Baum HJ, Shaler RC (2001) Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Sci. Int*. 120:177-188.
- Walsh B, Redd AJ, Hammer MF. (2008) Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Sci. Int*. 174(2-3): 234-238.

Summary

- Y-STRs can aid in forensic casework and can be another useful tool in the courtroom (along with autosomal STRs)
- A Y-STR database consists of haplotype frequencies rather than allele frequencies because Y-STR loci are located on the non-recombining part of the Y-chromosome and are therefore considered linked.
- Haplotype frequency is based on the counting method. Applying a confidence interval corrects for database size and sampling variation.
- Mutation rates for Y-STRs are similar to autosomals (~0.2%). Regions of the Y-chromosome can be duplicated or deleted causing Y-STRs to be duplicated or deleted.

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

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