Y-STRs: Investigations, Mutations and Standardization

Amy Decker
National Institute of Standards and Technology

Mid-Atlantic Association of Forensic Scientists
Annual Meeting
May 7, 2009

Presentation Outline

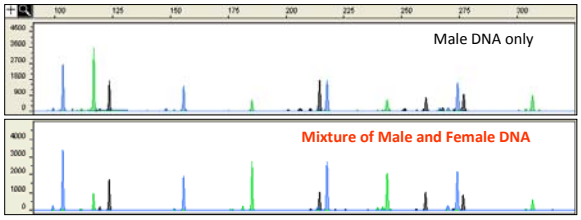
- Advantages of Y-STRs in human identity testing
- Y-STR markers and kits available
- Different population databases and statistics for reporting matches
- Mutation rates, duplications, and deletions and their impact on interpretation
- SWGDAM Guidelines on Y-STR Interpretation

Value of Y-STRs to Forensic Casework

Y-STRs can extend range of potential solvable forensic cases

- **Enabling detection of male DNA when mixed with excess female DNA**
  - Sexual assaults by vasectomized or azoospermic males (no sperm left behind to enable differential extraction)
  - Fingernail scrapings from sexual assault victims
  - Other bodily fluid mixtures (blood-blood, skin-saliva)
  - Extending length of time after assault for recovery of perpetrator's DNA profile (greater than 48 hours)
- Dealing with multiple male contributors
  - Gang rape situation to include or exclude potential contributors
- Gender clarification (with amelogenin Y null alleles)
- Extension of power of discrimination (with partial profiles)

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



800X female DNA

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)
- **Acceptance of statistical reports using the counting method** due to previous experience with mtDNA

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)**

Primary Commercial Y-STR Kits

Allele size range and locus dye colors

PowerPlex® Y

Released by Promega Corporation in Sept 2003

	100bp	200bp	300bp	400bp
FL	DYS391	DYS389I	DYS439	DYS389II
JOE	DYS438	DYS437	DYS19	DYS392
TMR	DYS393	DYS390	DYS385a/b	

3 dye colors
12-plex PCR

AmpFℓSTR® Yfiler™

Released by Applied Biosystems in Dec 2004

	100bp	200bp	300bp	400bp
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

Y-STR Databases and Statistics

Y-Chromosome Information Resources on the NIST STRBase Website

Largest Y-STR Database
http://www.yhrd.org

70,259 haplotypes (9 loci)
43,557 haplotypes (11 loci)

Y-Chromosome Haplotype Reference Database (YHRD)

YHRD.ORG3.0

http://www.yhrd.org

As of 4/30/09 (release 28): (538 populations)

72,055 total haplotypes in database

43,557 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

PowerP

1311
325
894
1108
366 Native Americans

4,004 total
(as of March 2005)

ucasians
panics
ican Americans
104 Native Americans

3,406 total

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

3,561 total
(as of December 2004)

U.S. Y-STR Database

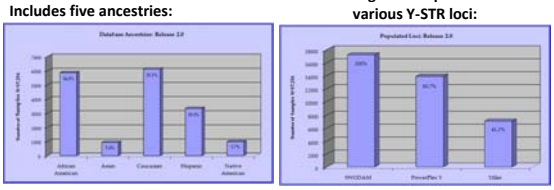
US Y-STR

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

As of 4/30/09 (Release 2.0): **17,216 haplotypes with 11 loci**

Includes five ancestries:

Percentage of samples run with various Y-STR loci:



17,216 samples with 11-locus SWGDAM haplotype
13,892 samples with 12-locus Powerplex Y haplotype
7,075 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)
  - 19 matches in 70,259
- YHRD (11 loci)
  - 3 matches in 43,557
- US Y-STR (11 loci)
  - 0 matches in 17,216

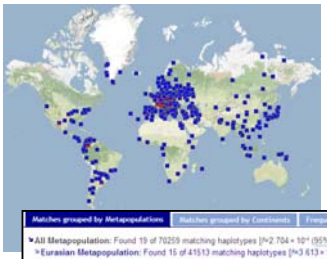
Searches performed 4/30/09

Y-Chromosome Haplotype Reference Database

www.YHRD.org

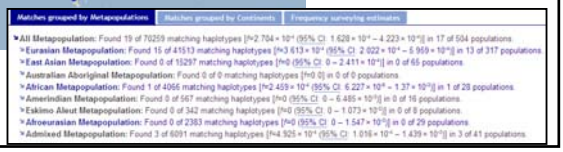
19 matches in 70,259 individuals from 504 worldwide populations

Release "28" from 2009-04-30 12:51



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 17,216

Ancestry	# of Haplotypes	Number of Haplotypes (with Selected Alleles)	Frequency	Frequency Upper Bound (95%)
African American	5858	0	0.000000	0.000511
Asian	935	0	0.000000	0.003198
Caucasian	6109	0	0.000000	0.000490
Hispanic	3331	0	0.000000	0.000898
Native American	983	0	0.000000	0.003042
Total	17216	0	0	0.000173

Overall Database Summary:

The selected haplotype is found in 0 of 17216 total individuals within the database with a frequency of 0. Applying the 95% upper confidence interval results in a frequency of 0.000173, which is equivalent to approximately 1 in every 5780 individuals.

The selected haplotype is found in 0 of 5858 **African American** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.000511, which is equivalent to approximately 1 in every 1957 individuals.

The selected haplotype is found in 0 of 935 **Asian** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.003198, which is equivalent to approximately 1 in every 313 individuals.

The selected haplotype is found in 0 of 6109 **Caucasian** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.00049, which is equivalent to approximately 1 in every 2041 individuals.

The selected haplotype is found in 0 of 3331 **Hispanic** individuals within the database, with a frequency of 0.000000. Applying the 95% upper confidence interval results in a frequency of 0.000898, which is equivalent to approximately 1 in every 1114 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:

$$p = X/N$$

19 matches in 70,259

p = 19/70,259 = 0.00027 = **0.027%**

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

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

$$0.00027 + 1.96\sqrt{\frac{(0.00027)(1-0.00027)}{70{,}259}}$$

= 0.00027 + 0.000121 = 0.000391

= **0.039% (1 in 2554)**

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 17,216

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

$1-\alpha^{1/N} = 1-(0.05)^{[1/17{,}216]} = 0.000174$

= **0.017% (1 in 5747)**

A simplified calculation would be 3/N.

In this example: 3/17216 = 0.000174 = 0.017% (1 in 5747)

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

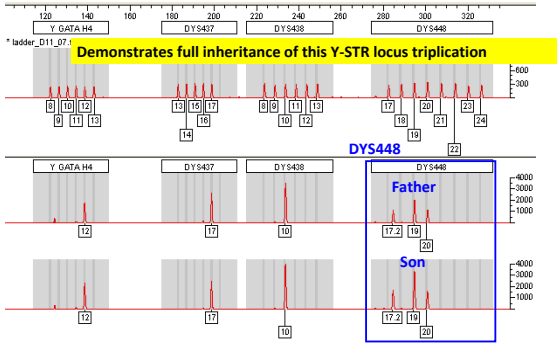
Gusmão, L., Butler, J.M., et al. (2006) *Forensic Sci. Int.* 157:187-197

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

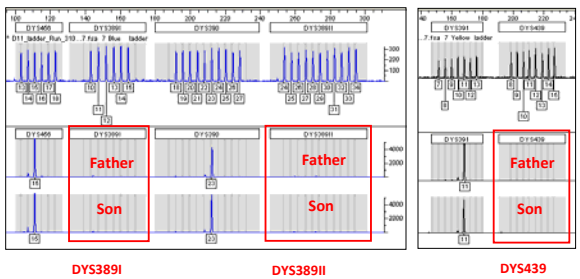
Decker, A.E., Kline, M.C., Redman, J.W., Reid, T.M., Butler, J.M. (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *FSI Genetics 2(3): e31-e35.*

DYS448 Triplication Seen in Both Father and Son



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

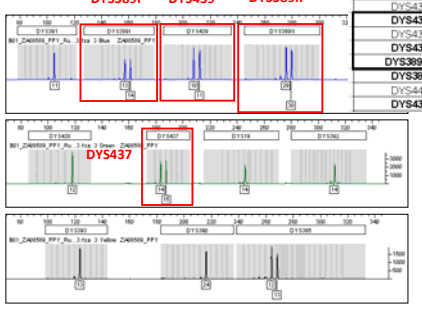
Yfiler data



Full inheritance of these Y-STR locus deletions

Duplication at Multiple Loci with Single-Source Sample

PowerPlex Y data



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

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

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

Recent News: Mutation Rates of Yfiler loci in the literature

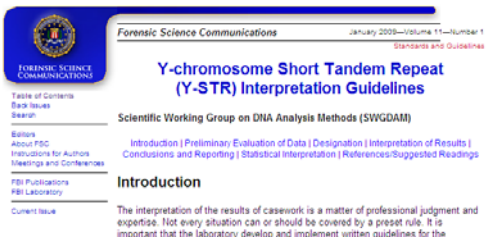
- Goedbloed et al 2009 observed a father:son pair with **THREE Y-STR mutations** in the 17 Yfiler loci
  - This is the **first report** of this number of mutations and **updates previous conclusions** on the threshold for the number of allelic differences to conclude an "exclusion".
- Ge et al 2009 observed a **three-step mutation** in a father:son pair with Yfiler kit at DYS456
- Amorim 2008: **Cautions against combining evidence** from autosomal loci with mito or Y-chromosome loci

SWGDAM Y-STR Interpretation Guidelines

http://www.fbi.gov/hq/lab/fsc/current/backissu.htm

Approved July 15, 2008 by SWGDAM

Published in *Forensic Sci. Comm.* Jan 2009, Volume 11, Number 1


Forensic Science Communications

Y-chromosome Short Tandem Repeat (Y-STR) Interpretation Guidelines

Scientific Working Group on DNA Analysis Methods (SWGDAM)

Introduction | Preliminary Evaluation of Data | Designation | Interpretation of Results | Conclusions and Reporting | Statistical Interpretation | References/Suggested Readings

Introduction

The interpretation of the results of casework is a matter of professional judgment and expertise. Not every situation can or should be covered by a preset rule. It is important that the laboratory develop and implement written guidelines for the

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.1) Population Databases

- Loci on NRY should be considered linked as a single locus
- **Source of population database should be documented**
- Relevant population(s) for which the frequency will be estimated should be identified
- **Consolidated US Y-STR database should be used for population frequency estimation**

http://www.usystrdatabase.org

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

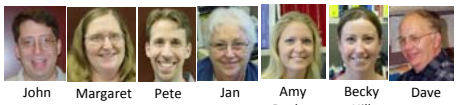
(5.5) Joint Match Probability

- **The product rule may be utilized** to combine the autosomal STR genotype match probability and Y-STR haplotype frequency information
- Citation to Walsh et al. (2008) Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Sci. Int.* 174: 234-238

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 Redma, 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