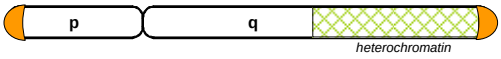


Y-STRs:
Markers, Mutations, and More

John M. Butler, Ph.D.
National Institute of Standards and Technology

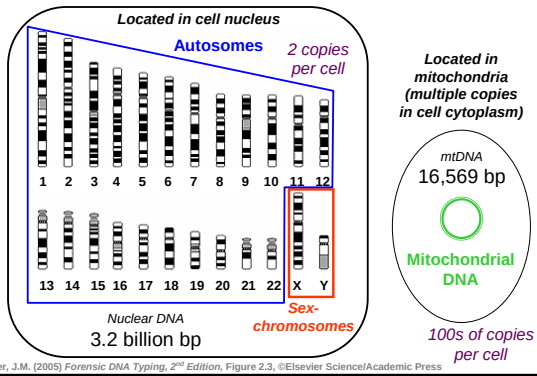
Presentation to the California Department of Justice DNA Laboratory
March 7, 2007

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
- Value of additional Y-STR loci (beyond the Yfiler 17)

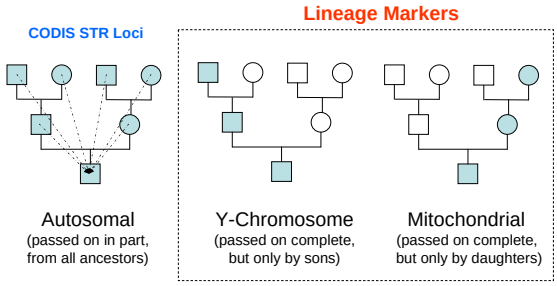
Human Genome

23 Pairs of Chromosomes + mtDNA



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 2.3, ©Elsevier Science/Academic Press

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

<u>Disadvantages</u>

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

Genetic Genealogy Companies

FamilyTreeDNA
http://www.familytreedna.com
http://www.dna-fingerprint.com

Sorenson Genomics
http://www.sorensongenomics.com

Relative Genetics / GeneTree
http://www.relativegenetics.com

oxford ancestors
http://www.oxfordancestors.com

DNA Heritage
http://www.dnaheritage.com

Ethnoancestry
http://www.ethnoancestry.com

GeoGene
http://www.geogene.com

The rapidly growing field of genetic genealogy is expanding the use of mtDNA and Y-STRs.

Perhaps the Real Reason Some Genetic Genealogy Is Performed…

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

171 labs

Questions #45a & #45b

- **Is your lab using or validating Y-STRs?**
  - **51 Yes** (30%)
    28 Yfiler, **15** PowerPlex Y, some both kits
  - **114** No
  - **6** no response

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

171 labs

Question #50

- Y-STR data can be entered in CODIS similar to entering the current STR loci in CODIS. **Do you think CODIS should include Y-STR loci in Popstats calculations?**
  - Yes – 116 (68%)
  - No – 18
  - No response – 37

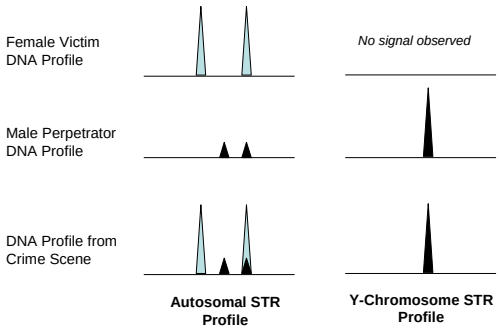
A *Law & Order* episode last week discussed the "CODIS Y-STR database" and its capabilities for familial searching…

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

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

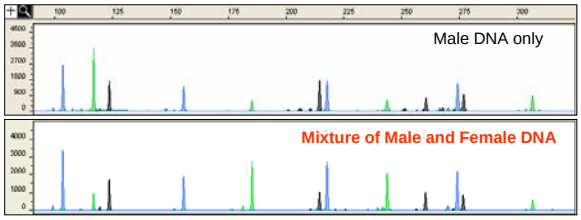


Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.2, ©Elsevier Science/Academic Press

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

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

A new section on the NIST STRBase website will be created on this topic soon

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
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

History of Y-STR Marker Discovery

"Extended Haplotype"

1992 - DYS19 (Roewer et al.)
1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)
1996 - DYS389I/II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)
1996 - DYF371, DYS425, DYS426 (Jobling et al.)
1997 - DYS288, DYS388 (Kayser et al.)
1998 - DYS385 a/b (Schneider et al.)

"Minimal Haplotype"

1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)
2000 - DYS434, DYS435, DYS436, DYS437, DYS438, DYS439 (Ayub et al.)
2000 - G09411 (DYS462), G10123 (de Knijff unpublished)

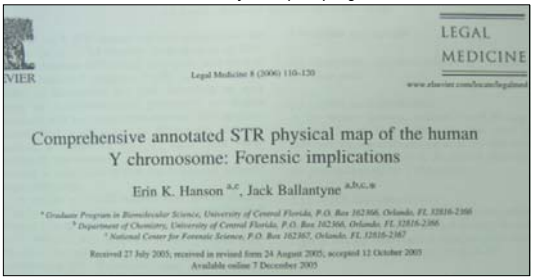
SWGDAM core

2001 - DYS441, DYS442 (Iida et al.)
2002 - DYS443, DYS444, DYS445 (Iida et al.); DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d (Redd et al.)
2002 – DYS468-DYS596 (129 new Y STRs; Manfred Kayser GDB entries)
2003 – DYS597-DYS645 (50 new Y STRs; Manfred Kayser GDB entries)
2004-2006 – DYS648-726 (GDB entries)

From J.M. Butler (2003) Recent developments in Y-STR and Y-SNP analysis. *Forensic Sci. Rev.* 15:91-111

Physical Map of the Human Y-Chromosome

Hanson, E.K. and Ballantyne, J. (2006) *Legal Med* 8: 110-120


LEGAL MEDICINE

Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications

Erin K. Hanson [a,c], Jack Ballantyne [a,b,c,*]

Describe the precise location of 417 Y-STRs

They note that not all will be useful due to low genetic variation or high X-chromosome homology

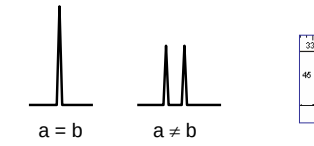
See also http://ncfs.ucf.edu/ystar/ystar.html

Y-STR Typing of Duplicated Regions

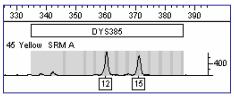
"multi-copy loci"

a repeat b repeat

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

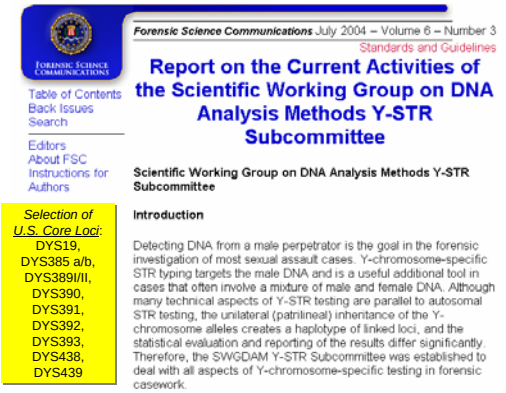


Y-PLEX™ 6 results



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


Forensic Science Communications July 2004 – Volume 6 – Number 3
Standards and Guidelines

Report on the Current Activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee

Table of Contents
Back Issues
Search
Editors
About FSC
Instructions for Authors

Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee

Selection of U.S. Core Loci: DYS19, DYS385 a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Introduction

Detecting DNA from a male perpetrator is the goal in the forensic investigation of most sexual assault cases. Y-chromosome-specific STR typing targets the male DNA and is a useful additional tool in cases that often involve a mixture of male and female DNA. Although many technical aspects of Y-STR testing are parallel to autosomal STR testing, the unilateral (patrilineal) inheritance of the Y-chromosome alleles creates a haplotype of linked loci, and the statistical evaluation and reporting of the results differ significantly. Therefore, the SWGDAM Y-STR Subcommittee was established to deal with all aspects of Y-chromosome-specific testing in forensic casework.

Core Y-STR Characteristics

11 PCR products 9 primer sets

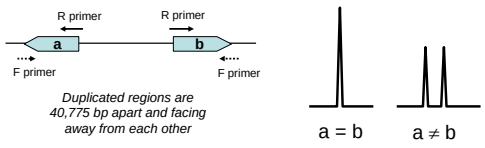
STR Marker	Position (Mb)	Repeat Motif	Allele Range	Mutation Rate
DYS393	3.17	AGAT	8-17	0.05%
DYS19	10.12	TAGA	10-19	0.20%
DYS391	12.54	TCTA	6-14	0.40%
DYS439	12.95	AGAT	8-15	0.38%
DYS389 I/II	13.05	[TCTG] [TCTA]	9-17 / 24-34	0.20%, 0.31%
DYS438	13.38	TTTTC	6-14	0.09%
DYS390	15.71	[TCTA] [TCTG]	17-28	0.32%
DYS385 a/b	19.19, 19.23	GAAA	7-28	0.23%
DYS392	20.97	TAT	6-20	0.05%

Positions in megabases (Mb) along the Y-chromosome were determined with NCBI build 35 (May 2004) using BLAT. Allele ranges represent the full range of alleles reported in the literature. Mutation rates summarized from YHRD (http://www.yhrd.org; accessed 6 Apr 2005).

Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci.*, 51(2): 253-265

(A) DYS385 a/b

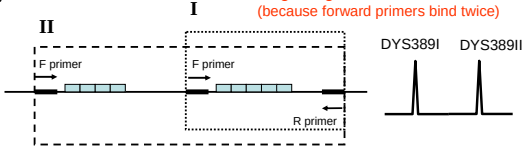
Multi-Copy (Duplicated) Marker



Duplicated regions are 40,775 bp apart and facing away from each other

(B) DYS389 I/II

Single Region but Two PCR Products
(because forward primers bind twice)



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

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

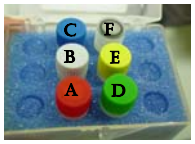
AmpFℓSTR® Yfiler™

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

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)**
- **22 Y STR markers sequenced**
- **9 additional Y STR markers typed**
- **42 Y SNPs typed** with Marligen kit

Certified for all loci in commercial Y-STR kits:

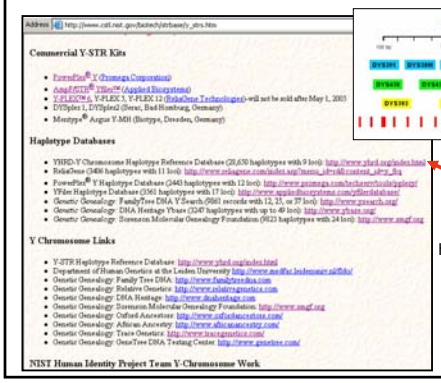
Y-PLEX 6
Y-PLEX 5
Y-PLEX 12
PowerPlex Y

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

Y-filer - adds DYS635 (C4); now sequenced

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

Y-Chromosome Information Resources on the NIST STRBase Website



Largest Y-STR Database
http://www.yhrd.org
41,965 haplotypes (9 loci)
14,835 haplotypes (11 loci)

Y-Chromosome Haplotype Reference Database (YHRD)

yhrd str database

http://www.yhrd.org

(357 populations)

As of 8/1/06: **41,965 haplotypes**

14,835 haplotypes
with all US required loci
(98 populations)

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/

PowerPlex Y

1311 Caucasians
325 Asians
894 Hispanics
1108 African Americans
366 Native Americans

4,004 total
(as of March 2005)

Yfiler

1276 Caucasians
330 Asians
597 Hispanics
985 African Americans
106 Native Americans
105 Filipino
59 Sub-Saharan Africans
103 Vietnamese

3,561 total
(as of December 2004)

Statistics with Y-STR Haplotypes

Most labs will probably go with the **counting method** (number of times a haplotype is observed in a database) as is typically done with mtDNA results

Questions under review: Should a theta correction be employed?
If so, what is the appropriate formula to use involving theta?
Would such a hypothetical theta value be universal for all subpopulations?
Would it really change the overall reporting statistic significantly?

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)
  - 7 matches in 27,773
- YHRD (11 loci)
  - 0 matches in 6,281
- ReliaGene (11 loci)
  - 0 matches in 3,403
- PowerPlex Y (12 loci)
  - 0 matches in 4,004
- Yfiler (17 loci)
  - 0 matches in 3,561

Y-Chromosome Haplotype Reference Database
www.YHRD.org

7 matches in 27,773 individuals from 236 worldwide populations

Release "15" from 2004-12-17 16:11:24

Minimal Haplotype Result

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

Population	#	Metapopulation
Bogotá, Colombia [European]	1 / 147	Eurasian MP / European MP
Central Portugal	1 / 230	Eurasian MP / European MP
Cologne, Germany	1 / 135	Eurasian MP / European MP
Leipzig, Germany	1 / 661	Eurasian MP / European MP
Liguna, Italy	1 / 81	Eurasian MP / European MP
London, UK	1 / 285	Eurasian MP / European MP
Lyon, France	1 / 125	Eurasian MP / European MP

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:

$$p = X/N$$

7 matches in 27,773

p = 7/27,773 = 0.000252 = **0.025%**

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

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

$$0.000252 + 1.96\sqrt{\frac{(0.000252)(1-0.000252)}{27{,}773}}$$

= 0.000252 + 0.000187 = 0.000439
= **0.044% (~1 in 2270)**

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 4,004

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

$1-\alpha^{1/N} = 1-(0.05)^{[1/4,004]} = 0.000748$
= **0.075% (~1 in 1340)**

If using database of 2,443, then the best you can do is 1 in 816

National U.S. Y-STR Population Database

- Efforts underway at the **University of Central Florida** (with NIJ funding) to consolidate all U.S. data on Y-STR loci for population
- Data from ReliaGene, Promega, Applied Biosystems being gathered plus any forensic lab population sample data available

Current Y-STR Databases

AGENCY	# MARKERS	# SAMPLES
NCFS	76	1,396
University of AZ	38	2,518
AB	17	3,561
Promega	12	4,004
Reliagene	11	4,623
Proposed National Y-STR Database		16,102
Proposed National Y-STR Database with YHRD		29,187 (54,863 MHL)

NIJ

Slide from Jack Ballantyne, CODIS Conference (Oct 2006) presentation

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)

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 Identifiler** (QC for gender and potential sample switches)
- Typed with Yfiler (17 Y-STRs) – **examined mutations**

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	**3.3% with 12 Y-STRs**
…			
40	0.89390382	**0.10609618**	

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

Separating Brothers with 47 Y-STRs

- Two suspected brothers (ZT79338 and ZT79339) are part of our ~660 U.S. sample dataset at NIST.
- Thus far, we have evaluated 47 Y-STR allele calls on these samples.
- **A mutation at DYS391 separates these individuals** (one contains allele 11 and the other allele 10).
- These samples share autosomal STR alleles and contain identical mtDNA sequences.

Y-STR Mutation Rates for the 17 Yfiler Loci

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

Mutations Seen in 100 African American Father-Son Pairs

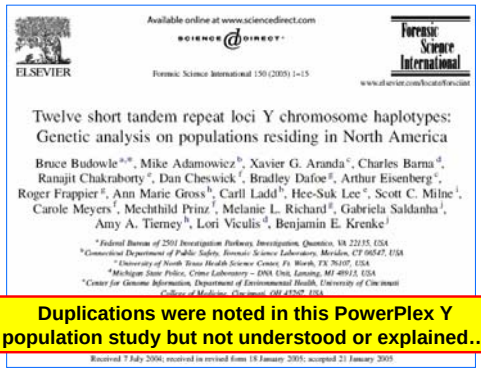
Ethnicity	Sample	locus	Allele (father)	Allele (child)	Comments
African American	65B	Y GATA H4	11	9	loss of 2 repeats
African American	46B	DYS389I and DYS389II	14,30	13,29	loss of 1 repeat
African American	58B	DYS389I and DYS389II	14,32	15,33	gain of 1 repeat
African American	18B	DYS390	24	23	loss of 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

Locus Duplication and Deletion

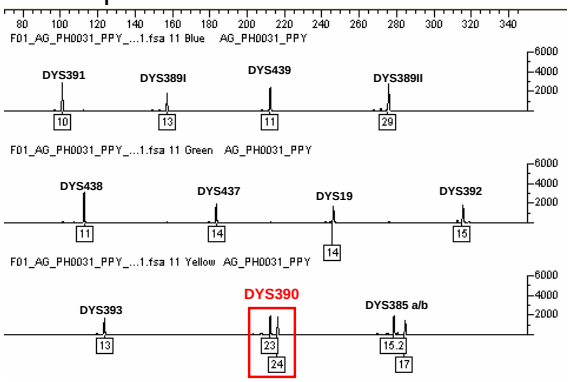
Events that impact Y-STR interpretation

PowerPlex Y Population Study



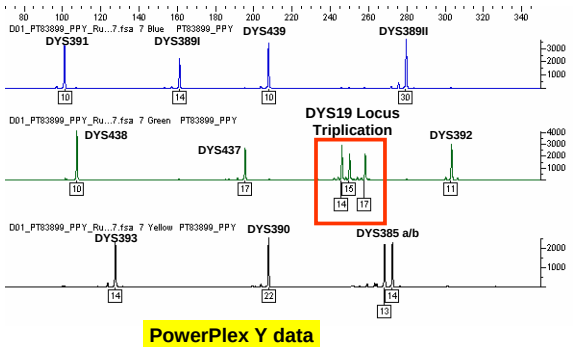
Duplications were noted in this PowerPlex Y population study but not understood or explained…

Duplication Observed at DYS390

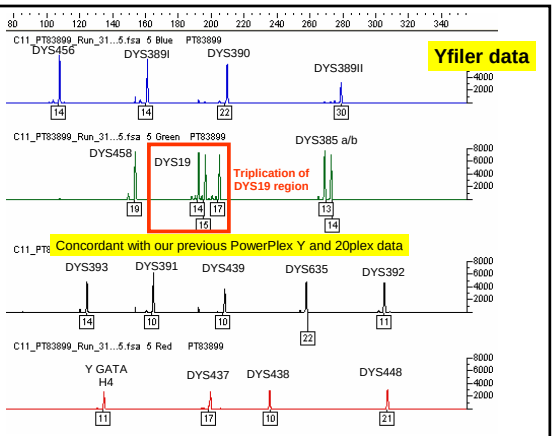


PowerPlex Y results on sample from Ann Marie Gross (MN BCA)

Locus Triplication at DYS19

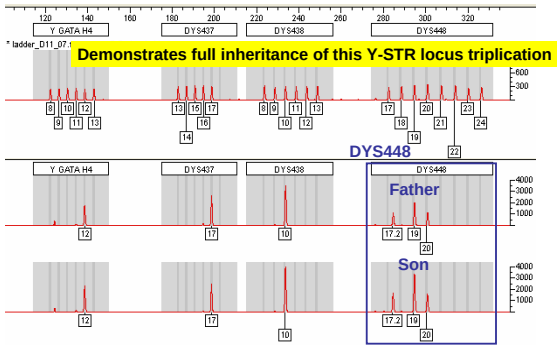


PowerPlex Y data



DYS448 Triplication

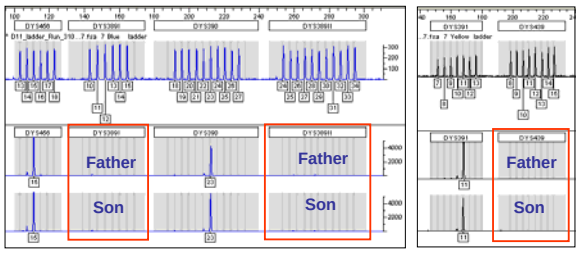
Seen in Both Father and Son



DYS389I, DYS389II, DYS439 Deletions

Seen in Both Father and Son

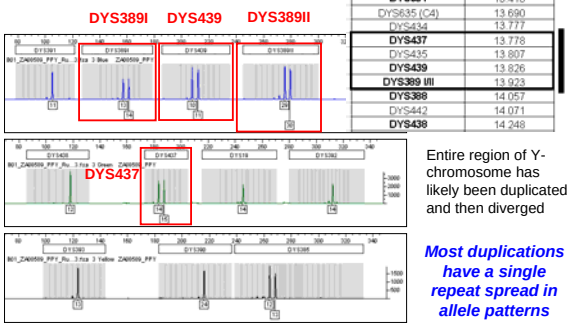
Yfiler data



Full inheritance of these Y-STR locus deletions

Duplication at Multiple Loci with Single-Source Sample

PowerPlex Y data



Y-chromosome mapping

q-arm

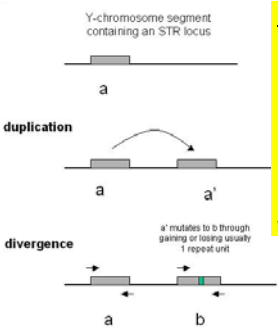
Y STR Marker	Position (Mb)
DYS391	13.413
DYS635 (C4)	13.690
DYS434	13.777
DYS437	13.778
DYS435	13.807
DYS439	13.826
DYS389 I/II	13.923
DYS388	14.057
DYS442	14.071
DYS438	14.248

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

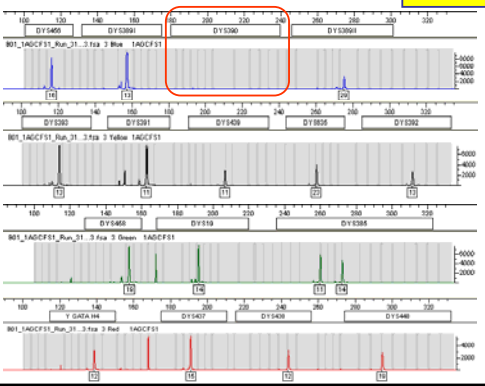
- ***Note the number of loci containing >1 allele (other than multi-copy DYS385)***
- ***Consider relative position on the Y-chromosome if multiple loci have two alleles***
- ***See if repeat spread is >1 repeat unit***
- ***Examine DYS385 for presence of >2 alleles***

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

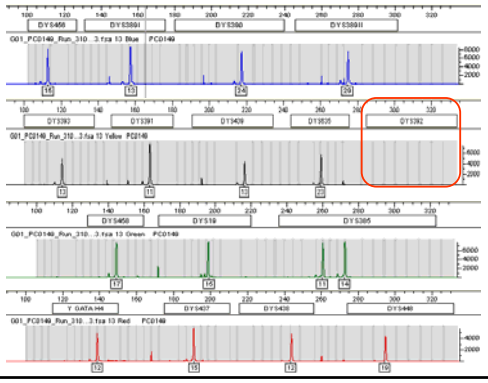
CFS-1 with Y filer

DYS390 is deleted



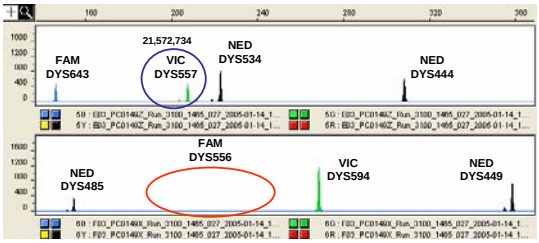
Sample PC0149 with Yfiler

DYS392 is deleted



PC0149 with Additional Y-STRs

One of the closest available loci fails



DYS556 ~32,000 bp away from DYS392 is **missing**

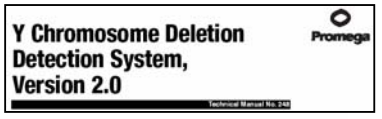
DYS557 ~600,000 bp away from DYS392 is present

Deletions of some Y-STRs can be an inadvertent diagnosis of male infertility

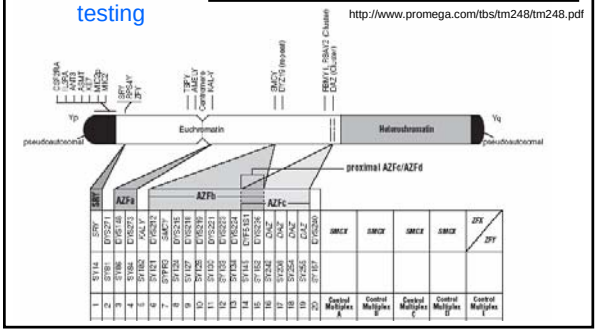
King *et al.* (2005) Inadvertent diagnosis of male infertility through genealogical DNA testing. *J. Med. Genet*. 42:366-368

- ***AZFa* deletion** (<1 in 100,000 men): expected to lack **DYS389I/II, DYS437, DYS438, DYS439**
- ***AZFb* deletion** (very rare): expected to lack **DYS385 and DYS392**
- ***AZFc* deletion** (1 in 4,000 men): expected to lack DYS464
- Possible that "incomplete" haplotypes are not being submitted to the Y-STR haplotype databases
- **Thus, Y-STRs are not neutral with respect to fertility information**

Promega sells a Y-deletion test for infertility testing



http://www.promega.com/tbs/tm248/tm248.pdf



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!

Value of Additional Loci

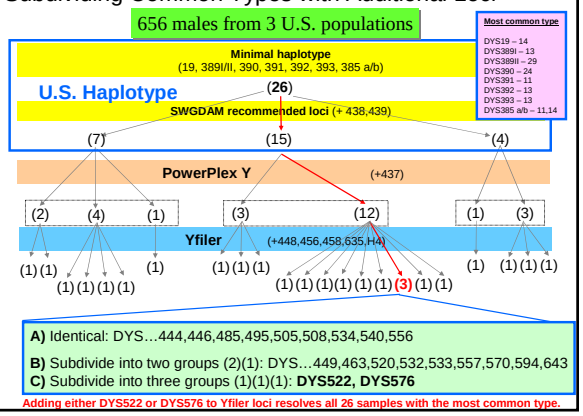
Going Beyond Commercial Y-STR Kits

- 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?

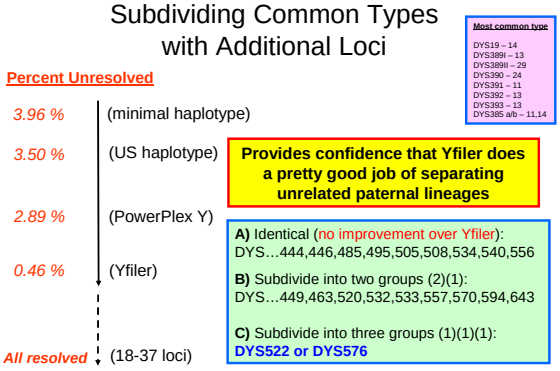
Data Set Used to Examine Common Types

- 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.

Subdividing Common Types with Additional Loci



Subdividing Common Types with Additional Loci



# times haplotype observed	9 MHL
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

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.

With the 9 loci of the minimal haplotype (MHL) run on 656 samples, 26 samples had the most common type

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM
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

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

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY
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

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

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY	17 Yfiler
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

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

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY	17 Yfiler	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**	.	.	.	.
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

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

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
~~DYS444~~	~~12~~	~~12~~	~~12~~
~~DYS446~~	~~13~~	~~13~~	~~13~~
DYS449	30	30	31
DYS463	24	24	23
~~DYS485~~	~~15~~	~~15~~	~~15~~
~~DYS495~~	~~16~~	~~16~~	~~16~~
~~DYS505~~	~~12~~	~~12~~	~~12~~
~~DYS508~~	~~11~~	~~11~~	~~11~~
DYS520	21	22	21
DYS522	10	12	11

Locus	MT97185	ZT79333	TT51702
DYS532	14	14	13
DYS533	13	12	13
~~DYS534~~	~~15~~	~~15~~	~~15~~
~~DYS540~~	~~12~~	~~12~~	~~12~~
~~DYS556~~	~~11~~	~~11~~	~~11~~
DYS557	15	17	17
DYS570	16	17	17
DYS576	17	20	18
DYS594	9	10	10
DYS643	10	11	10

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

Subdividing Unresolved Yfiler Haplotypes(2)

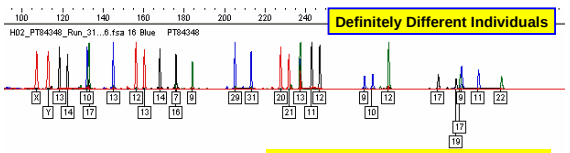
Sample Info	DYS 19	DYS 385a/b	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	H4
PT83904	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12
PT84348	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12
ZT80369	13	13,14	15	31	24	9	11	13	10	10	14	20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

Locus	PT83904	PT84348	ZT80369
~~DYS444~~	~~12~~	~~12~~	~~12~~
~~DYS446~~	~~12~~	~~12~~	~~12~~
~~DYS449~~	~~31~~	~~31~~	~~31~~
~~DYS463~~	~~18~~	~~18~~	~~18~~
~~DYS485~~	~~15~~	~~15~~	~~15~~
~~DYS495~~	~~12~~	~~12~~	~~12~~
~~DYS505~~	~~11~~	~~11~~	~~11~~
~~DYS508~~	~~11~~	~~11~~	~~11~~
~~DYS520~~	~~19~~	~~19~~	~~19~~
~~DYS522~~	~~12~~	~~12~~	~~12~~

Locus	PT83904	PT84348	ZT80369
~~DYS532~~	~~14~~	~~14~~	~~14~~
~~DYS533~~	~~11~~	~~11~~	~~11~~
DYS534	16	17	17
~~DYS540~~	~~11~~	~~11~~	~~11~~
~~DYS556~~	~~12~~	~~12~~	~~12~~
~~DYS557~~	~~18~~	~~18~~	~~18~~
~~DYS570~~	~~22~~	~~22~~	~~22~~
~~DYS576~~	~~18~~	~~18~~	~~18~~
~~DYS594~~	~~11~~	~~11~~	~~11~~
~~DYS643~~	~~12~~	~~12~~	~~12~~

PT84348 and ZT80369 are unresolved even with 37 Y-STRs

Identifiler Results from Two Unresolved Hispanic Males Following Typing with 37 Y-STR Loci



No allele sharing at 6 out of 15 STRs examined

Sample	D8	D21	D7	CSF	D3	TH	D13	D16	D2	D19	VWA	TPO	D18	Am	D5	FGA
PT84348	10,13	29,31	9,10	9,11	17,17	7,9	13,13	12,12	17,22	13,14	14,16	11,12	17,19	X,Y	12,13	20,21
ZT80369	14,15	28,31	10,11	11,12	16,18	7,9	8,11	12,13	23,25	14,15	14,15	8,11	13,15	X,Y	12,12	22,25

PT84348 and ZT80369 are unresolved even with 37 Y-STRs

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**.

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