Y-Chromosome and Mitochondrial DNA Analysis

The Human Y-Chromosome: Populations, Mutations, and Statistics

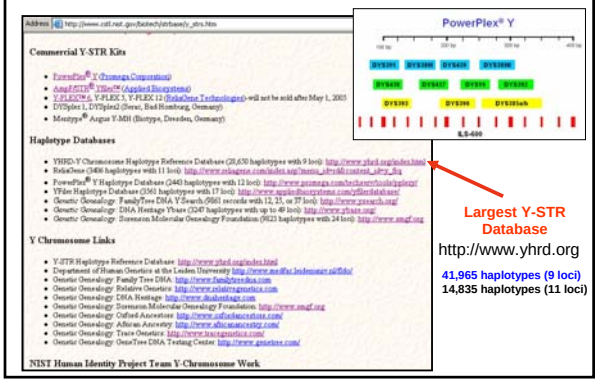
NEAFS 2006 Workshop
Rye Brook, NY
November 1, 2006

Northeastern Association of Forensic Scientists

Dr. John M. Butler
Dr. Michael D. Coble

john.butler@nist.gov
Michael.Coble@afip.osd.mil

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)
As of 12/17/04: 28,650 haplotypes
6,281 haplotypes
with all US required loci

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)

PowerPlex Y Haplotype Database

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

595 Caucasians
284 Asians
630 Hispanics
577 African Americans
357 Native Americans

2,443 total

March 2005

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

4,004 total

Yfiler Haplotype Database

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

Population	# Haplotypes
African American	985
Asian	330
Caucasian	1276
Filipino	105
Hispanic	597
Native American	106
Sub-saharan African	59
Vietnamese	103
All	**3561**

#Samples Contributed by NIST

259 African Americans
3 Asians
262 Caucasians

139 Hispanics

Data provided by NIST

663/3561 = **18.6%**

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

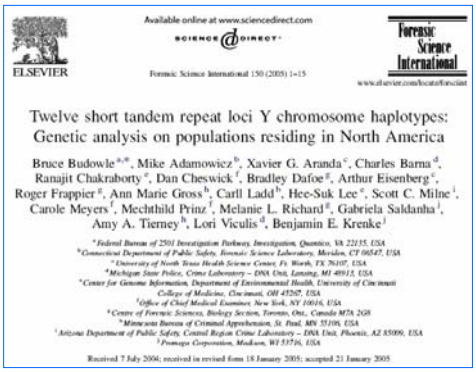
Slide from Jack Ballantyne, CODIS Conference (Oct 2006) presentation

A few recent Y-STR population studies

Population	# Samples	# Loci	Reference
5 North American groups	2,443	12	Budowle et al. (2005) FSI 150:1-15
U.S. Caucasians, African Americans, Hispanics	647	22 (27)	Schoske et al. (2004) FSI 139:107-121
Austrian	135	17	Berger et al. (2005) IJLM, in press (Yfiler)
91 European groups	12,700	7	Roewer et al. (2005) Hum Genet 116:279-291

More than 200 Y-STR population studies have been published (most of this data is deposited in the YHRD – Y Chromosome Haplotype Reference Database)

PowerPlex Y Population Study



Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Forensic Science International 150 (2005) 1–15
Forensic Science International
www.elsevier.com/locate/forsciint

Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America

Bruce Budowle [a,*], Mike Adamowicz [b], Xavier G. Aranda [c], Charles Barna [d], Ranajit Chakraborty [e], Dan Cheswick [f], Bradley Dafoe [g], Arthur Eisenberg [c], Roger Frappier [g], Ann Marie Gross [h], Carll Ladd [h], Hee-Suk Lee [e], Scott C. Milne [i], Carole Meyers [f], Mechthild Prinz [f], Melanie L. Richard [g], Gabriela Saldanha [j], Amy A. Tierney [h], Lori Viculis [d], Benjamin E. Krenke [j]

[a] Federal Bureau of 2501 Investigation Parkway, Investigation, Quantico, VA 22135, USA
[b] Connecticut Department of Public Safety, Forensic Science Laboratory, Meriden, CT 06547, USA
[c] University of North Texas Health Science Center, Ft. Worth, TX 76107, USA
[d] Michigan State Police, Crime Laboratory – DNA Unit, Lansing, MI 48913, USA
[e] Center for Genome Information, Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA
[f] Office of Chief Medical Examiner, New York, NY 10016, USA
[g] Centre of Forensic Sciences, Biology Section, Toronto, Ont., Canada M7A 2G8
[h] Minnesota Bureau of Criminal Apprehension, St. Paul, MN 55106, USA
[i] Arizona Department of Public Safety, Central Region Crime Laboratory – DNA Unit, Phoenix, AZ 85006, USA
[j] Promega Corporation, Madison, WI 53716, USA

Received 7 July 2004; received in revised form 18 January 2005; accepted 21 January 2005

Initial 2,443 Samples in PowerPlex Y Haplotype Database

Table 1
Sample populations and number of individuals (or haplotypes) per sample

159 Canadian samples

Region	Location	African American	Caucasian	Hispanic	Asian	Asian Indian	Native American (Apache/Navajo)
Canada	Ontario	37	57		28	37	
Northeast US	Connecticut	182	164	160			
	New York	80	83	80	45		
Midwest US	Michigan	86	97	97			
	Minnesota			101	101		
South US	Texas	192	194	192	73		
Southwest US	Arizona						138/219
Total (N = 2443)		577	595	630	247	37	357

PowerPlex® Y Haplotype Database
http://www.promega.com/techserv/tools/pplexy/

Compilation of Y STR Population Data

Appendices in Rich Schoske's Ph.D. dissertation; available on STRBase

Locus	Population Ethnicity	Sample Size	Allele Range (Allele Frequencies) 10	11	12	13	14	15	16	17	18	19	Reference
DYS19	EUROPE												
	Innsbruck (Austria)	100				0.14	0.56	0.23	0.07	0.01			Kayser et al. 1997
	Germany	86		0.07			0.512	0.279	0.081	0.046	0.012		Carracedo et al. 2001
	Bavaria (Germany)	151				NL	NL	NL	NL	NL			Anslinger et al. 2000
	Cologne (Germany)	163				0.041	0.5982	0.268	0.0515	0.0515			Hidding et al. 2000
	South Wurtemberg (Germany)	218				0.1009	0.5321	0.2431	0.0826	0.0413			Graw et al. 2000
	Munster (Germany)	272				0.04	0.57	0.23	0.12	0.04			Kayser et al. 1997
	Koln (Germany)	100				0.04	0.62	0.25	0.06	0.03			Kayser et al. 1997
	Jena (Germany)	143				0.06	0.55	0.18	0.16	0.06			Kayser et al. 1997
	Heidelberg (Germany)	113				0.07	0.5	0.29	0.09	0.04			Kayser et al. 1997
	Hahnover (Germany)	53				0.07	0.49	0.23	0.13	0.07			Kayser et al. 1997
	Magdeberg (Germany)	210				0.05	0.47	0.25	0.17	0.06			Kayser et al. 1997
	Bradenburg (Germany)	233				0.03	0.45	0.27	0.15	0.1			Kayser et al. 1997
	Munich 1 (Germany)	125				0.06	0.46	0.26	0.18	0.03			Kayser et al. 1997
	Munich 2 (Germany)	269				0.16	0.45	0.17	0.19	0.03			Kayser et al. 1997
	Berlin 1 (Germany)	233				0.07	0.39	0.27	0.21	0.06			Kayser et al. 1997
	Bremen	49				0.1	0.59	0.16	0.14				Kayser et al. 1997
	Leiden	88				0.04	0.7	0.19	0.03	0.02			Kayser et al. 1997
	Leicester, pooled	339			0.01	0.04	0.46	0.26	0.16	0.07	0.01		Kayser et al. 1997
	British	41				0.05	0.8	0.1	0.02	0.02			Kayser et al. 1997
	Bratislava	57				0.07	0.19	0.25	0.31	0.21			Kayser et al. 1997
	Norway	300				0.027	0.527	0.313	0.1200	0.013			Kayser et al. 1997

- **Source: over 200 published population data papers**
- Helps define observed allele ranges, which aids in multiplex assay development (spacing between loci in the same dye color)
- Information is available to the community through the STRBase website – ***permits analysis of optimal markers for particular population***

Richard Schoske Dissertation

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Schoske2003dis.pdf

Rich Schoske
PhD student from American University
Funded by Air Force

- Worked at NIST from Nov 2000 to May 2003
- **270 page Ph.D. dissertation**
- Entitled "The design, optimization and testing of Y chromosome short tandem repeat megaplexes."
- Available for download on NIST STRBase website

Source of YCC Samples

Problems with male lineages in population databases (YCC 6/7, 12/13, 15/16, 29/30, 49/50, 8/37) - really need detailed pedigree information

- 11 !Kung
- 5 Pygmy
- 10 Bantu speakers
- 14 Europeans
- 5 Middle Easterns
- 3 Pakistanis
- 3 Chinese
- 3 Japanese
- 5 North Asians
- 1 Cambodian
- 2 Melanesians
- 12 Native Americans

- **74 male cell lines**
- 2 females (YCC1 and 54)
- established in 1991 by Mike Hammer and Nathan Ellis

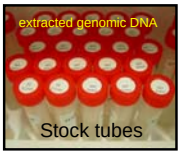
http://ycc.biosci.arizona.edu/

Standard U.S. Population Dataset

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

260 Caucasians, 260 African Americans, 140 Hispanics, 3 Asians = **663 males**

DNA extracted from whole blood (anonymous; self-identified ethnicities) received from Interstate Blood Bank (Memphis, TN) and Millennium Biotech Inc. (Ft. Lauderdale, FL)



To date: (>100,000 allele calls)
Identifiler (15 autosomal markers + Amelogenin) (10,608)
Roche Linear Arrays (HV1/HV2 10 regions) (6,630)
Y STRs 22 loci—27 amplicons (17,388)
Y STRs 27 new loci (14,535)
Yfiler kit 17 loci (11,237)
Y SNPs 50 markers on sub-set of samples (11,498)
Orchid 70 autosomal SNPs on sub-set (13,230)
miniSTR testing-new loci and CODIS concordance (9,228)
New miniSTR loci – for 26 loci, 17,238 genotypes
mtDNA full control region sequences by AFDIL

Genotypes with various human identity testing markers

U.S. Population Data on 22 Y-STRs

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Forensic Science International 139 (2004) 107–121
Forensic Science International
www.elsevier.com/locate/forsciint

High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays

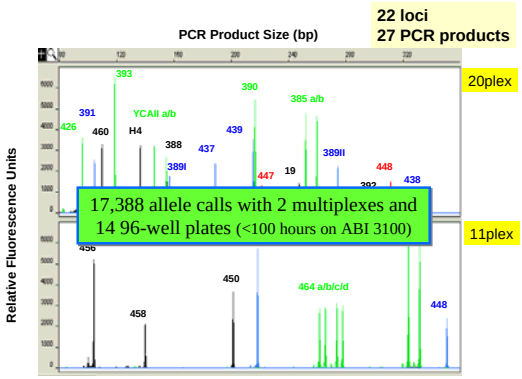
Richard Schoske[a,b], Peter M. Vallone[a], Margaret C. Kline[a], Janette W. Redman[a], John M. Butler[a,*]
[a]Biotechnology Division, National Institute of Technology, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899, USA
[b]Department of Chemistry, American University, Washington, DC 20016, USA

Received 29 April 2003; received in revised form 25 September 2003; accepted 1 October 2003

pdf file available at http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

NIST Multiplexes for High-Throughput Y STR Typing

22 loci
27 PCR products



Schoske *et al.* (2004) High-throughput Y-STR typing of U.S. populations…, *Forensic Sci. Int., 139:107-121*

US haplotype (Reliagene kits)

Yfiler (ABI)

PowerPlex Y (Promega)

+C4

Y-STR	Pooled Population STR diversity (N=647)	Rank	African American STR diversity (N=260)	Rank	Caucasian STR diversity (N=244)	Rank	Hispanic STR diversity (N=143)	Rank
DYS464 a/b/c/d	0.956	1	0.954	1	0.934	1	0.937	1
DYS385 a/b	0.912	2	0.942	2	0.838	2	0.901	2
YCAII a/b	0.790	3	0.797	3	0.701	5	0.772	4
DYS458	0.765	4	0.758	5	0.743	3	0.793	3
DYS390	0.764	5	0.664	10	0.701	5	0.665	13
DYS447	0.747	6	0.767	4	0.683	7	0.748	5
DYS389II	0.736	7	0.722	6	0.675	8	0.734	6
DYS448	0.721	8	0.722	6	0.595	11	0.704	8
DYS456	0.700	9	0.671	9	0.731	4	0.695	9
DYS438	0.691	10	0.560	15	0.594	12	0.690	10
DYS19	0.676	11	0.722	6	0.498	19	0.672	12
DYS439	0.656	12	0.636	11	0.639	9	0.717	7
DYS437	0.637	13	0.499	17	0.583	13	0.624	14
H4	0.611	14	0.612	12	0.562	14	0.609	15
DYS392	0.609	15	0.434	20	0.596	10	0.673	11
DYS460	0.570	16	0.568	14	0.555	15	0.556	18
DYS389I	0.549	17	0.531	16	0.538	17	0.596	16
DYS391	0.534	18	0.447	19	0.552	16	0.577	17
DYS426	0.519	19	0.375	21	0.482	20	0.522	19
DYS450	0.489	20	0.487	18	0.177	22	0.414	21
DYS393	0.485	21	0.586	13	0.363	21	0.448	20
DYS388	0.365	22	0.246	22	0.501	18	0.312	22

Schoske *et al.* (2004) High-throughput Y-STR typing of U.S. populations…, *Forensic Sci. Int., 139:107-121*

Statistical Calculations on Y-STR Data

- **Locus (gene) Diversity** = **$(n/n-1)(1 - \Sigma p_i^2)$** where n is the number of samples in the dataset and p_i is the frequency of the i^{th} allele
- **Haplotype Diversity** **(HD) = $(n/n-1)(1 - \Sigma p_i^2)$** where n is the number of samples in the dataset and p_i is the frequency of the i^{th} haplotype
- **Random Match Probability** (RMP) = 1 - HD
- **Discrimination Capacity** (DC) – total number of observed haplotypes divided by the total number of individuals in the dataset
- **Unique Haplotypes** (UH) – number of haplotypes that occur only once in the dataset

Calculating Gene (STR) Diversity

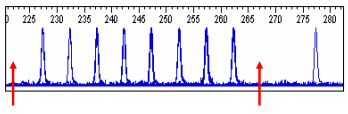
Locus	Allele	Size Range (bp)	Count	Combined Freq (N = 661)
DYS463	17	222.45	1	0.0015
	18	227.34-227.44	27	0.0408
	19	232.30-232.39	7	0.0106
	20	237.24-237.44	151	0.2284
	21	242.21-242.41	67	0.1014
	22	247.12-247.40	74	0.1120
	23	252.13-252.33	35	0.0530
	24	257.05-257.49	256	**0.3873**
	25	262.01-262.26	37	0.0560
	26	267.05-267.21	5	0.0076
	28	277.22	1	0.0015
		failure	2	
				STR diversity
		TOTAL	**661**	**0.7684**

D = $(n/n-1)(1 - \sum x^2)$

Comparison of Different Datasets to Evaluate Variability in Locus

DYS463-FAM

9 alleles



9 Alleles Observed

Combo 1 plate (95 samples)
31 Caucasians
32 African Americans
32 Hispanics

11 Alleles Observed

7 plates (663 samples)
260 Caucasians
260 African Americans
140 Hispanics
3 Asians

Our allele ranges are well defined after a 95 sample screen

Extended Haplotype vs. US Haplotype

Extended : 19, 389I/II, 390, 391, 392, 393, 385 a/b + **YCAII a/b**
US haplotype: 19, 389I/II, 390, 391, 392, 393, 385 a/b + **438, 439**

Y-STR Marker Combinations	260 African Americans		244 Caucasians		143 Hispanics	
	HD	RMP	HD	RMP	HD	RMP
"minimal" haplotype	0.9982	0.0018	0.9946	0.0053	0.9957	0.0043
"extended" haplotype	0.9988	0.0012	0.9971	0.0029	0.9975	0.0025
"U.S. haplotype"	0.9993	0.0007	0.9974	0.0026	0.9986	0.0014

HD = haplotype diversity
RMP = random match probability (1-HD)

$HD = (n/n-1)(1 - \Sigma p_i^2)$

U.S. haplotype is as good as extended haplotype in all major U.S. populations

Y-STR Marker Combinations	260 African Americans		244 Caucasians		143 Hispanics	
	DC	UH	DC	UH	DC	UH
"minimal" haplotype	88.5%	213	75.8%	161	81.1%	100
"extended" haplotype	91.9%	227	83.6%	184	89.5%	120
"U.S. haplotype"	91.9%	222	82.3%	176	93.3%	121

DC = discrimination capacity (number of haplotypes/number of samples)
UH = unique haplotype (occurs only once in a given population)

Schoske *et al.* (2004) High-throughput Y-STR typing of U.S. populations…, *Forensic Sci. Int., 139:107-121*

Y-STR Marker Combinations	**260 African Americans**		**244 Caucasians**		**143 Hispanics**	
	HD	RMP	HD	RMP	HD	RMP
Y-PLEX 6 kit	0.9974	0.0026	0.9914	0.0086	0.9934	0.0066
"minimal" haplotype	0.9982	0.0018	0.9946	0.0053	0.9957	0.0043
"extended" haplotype	0.9988	0.0012	0.9971	0.0029	0.9975	0.0025
"U.S. haplotype"	0.9993	0.0007	0.9974	0.0026	0.9986	0.0014
Y-STR 11plex	0.9993	0.0007	0.9987	0.0013	0.9992	0.0008
Y-STR 20plex	0.9998	0.0002	0.9998	0.0002	0.9998	0.0002
22 Y-STRs	0.9999	0.0001	0.9999	0.0001	0.9999	0.0001
Top 10 (w/o YCAII a/b)	0.9999	0.0001	0.9999	0.0001	0.9999	0.0001

Y-STR Marker Combinations	**260 African Americans**		**244 Caucasians**		**143 Hispanics**	
	DC	UH	DC	UH	DC	UH
Y-PLEX 6 kit	82.3%	188	68.9%	136	78.3%	97
"minimal" haplotype	88.5%	213	75.8%	161	81.1%	100
"extended" haplotype	91.9%	227	83.6%	184	89.5%	120
"U.S. haplotype"	91.9%	222	82.3%	176	93.3%	121
Y-STR 11plex	93.1%	227	88.5%	198	94.4%	127
Y-STR 20plex	98.5%	252	97.2%	230	98.6%	139
22 Y-STR Markers	98.9%	254	99.6%	242	99.3%	141
Top 10 (w/o YCAII a/b)	96.9%	244	97.5%	232	99.3%	141

Schoske *et al.* (2004) High-throughput Y-STR typing of U.S. populations…, *Forensic Sci. Int., 139:107-121*

# times haplotype observed	MHL	SWGDAM	PPY	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	.	.	.	.
14	.	.	.	.	.
15	.	1	.	.	.
16	.	.	.	.	.
17	.	.	.	.	.
18	.	.	.	.	.
19	.	.	.	.	.
20	.	.	.	.	.
21	.	.	.	.	.
22	.	.	.	.	.
23	.	.	.	.	.
24	.	.	.	.	.
25	.	.	.	.	.
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

N = 656

Haplotype Diversity (HD) vs. Discrimination Capacity (DC)

$HD = (N/N-1)(1 - \sum x^2)$

x = frequency of each haplotype

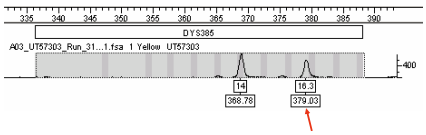
DC = (#HT)/N

Y-STR Variants and Mutations

DYS385 Flanking Region Deletion

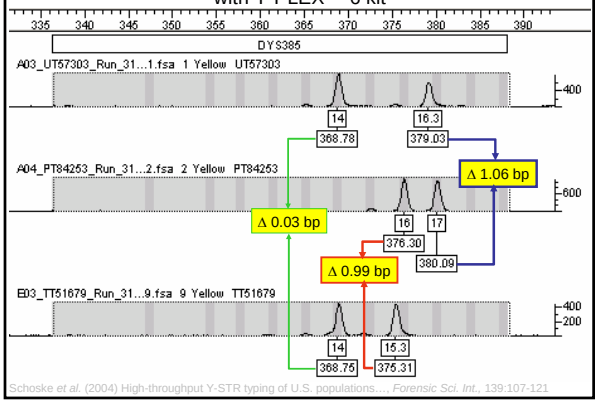
Schoske *et al.* (2004) High-throughput Y-STR typing of U.S. populations…, *Forensic Sci. Int.*, 139:107-121

- In a population of 244 Caucasians, one 15.3 and six 16.3 alleles were observed (~3%)
  - Alleles 15.3 and 16.3 are caused by a deletion in the flanking region that can be *outside* of some primer sets
  - Thus two different primer sets can produce seemingly discordant results (one primer set will report an "on-ladder" allele while another will result in an "off-ladder" allele)
- Could present challenges to DNA databases…

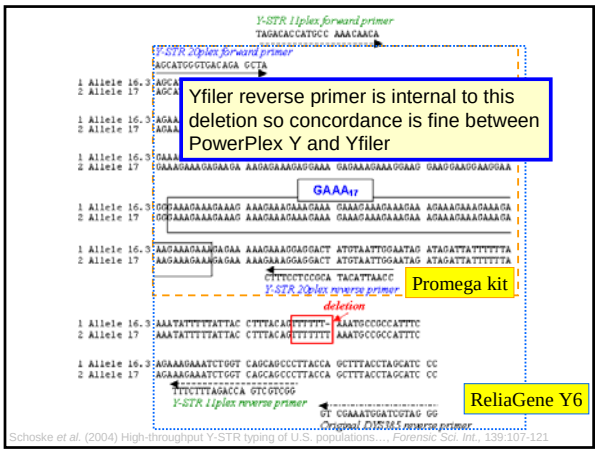


Really 17 repeats minus 1 bp deletion

Verification of DYS385 15.3 and 16.3 Alleles with Y-PLEX™ 6 kit

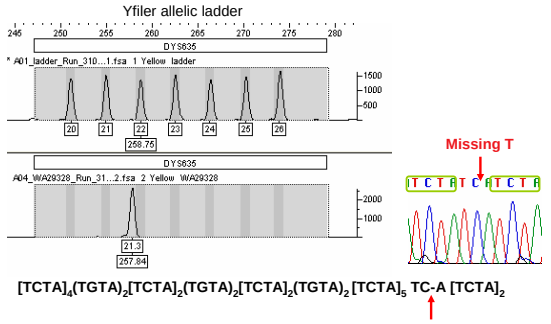


Schoske *et al.* (2004) High-throughput Y-STR typing of U.S. populations…, *Forensic Sci. Int.*, 139:107-121

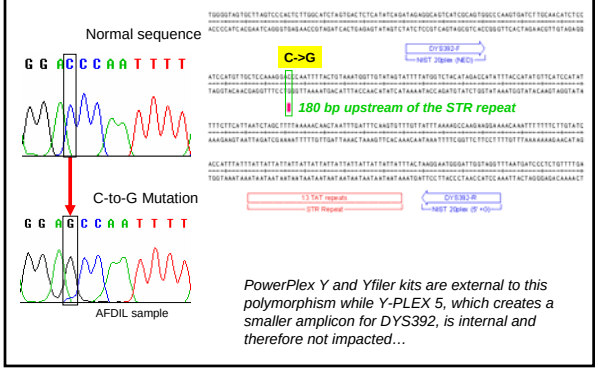


DYS635 Variant Allele 21.3

(Y-GATA-C4)



DYS392 Flanking Region Polymorphism



PowerPlex Y and Yfiler kits are external to this polymorphism while Y-PLEX 5, which creates a smaller amplicon for DYS392, is internal and therefore not impacted...

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 generated 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

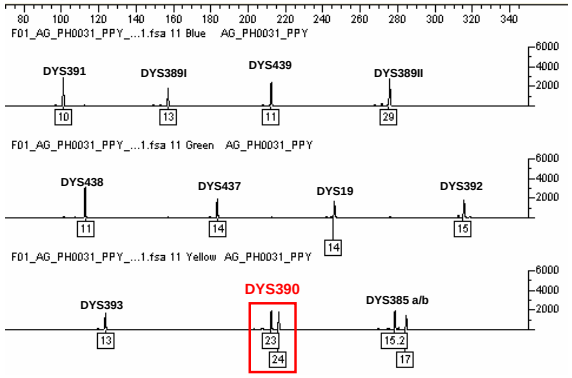
Ethnicity	Sample	locus	Allele (father)	Allele (child)	Comments
African American	65B	Y GATA H4	11	9	**loss of 2 repeats**
African American	46B	DYS389I and DYS389II	14,30	13,29	loss of 1 repeat
African American	58B	DYS389I and DYS389II	14,32	15,33	gain of 1 repeat
African American	18B	DYS390	24	23	loss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

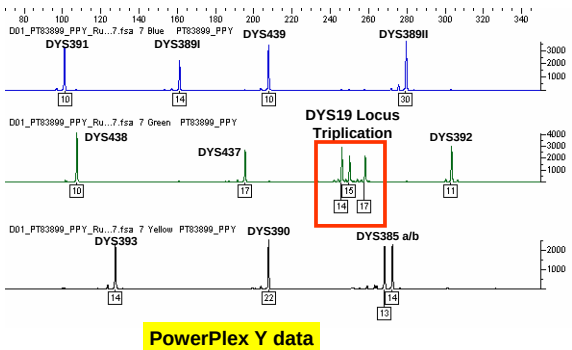
Events that impact Y-STR interpretation

Duplication Observed at DYS390

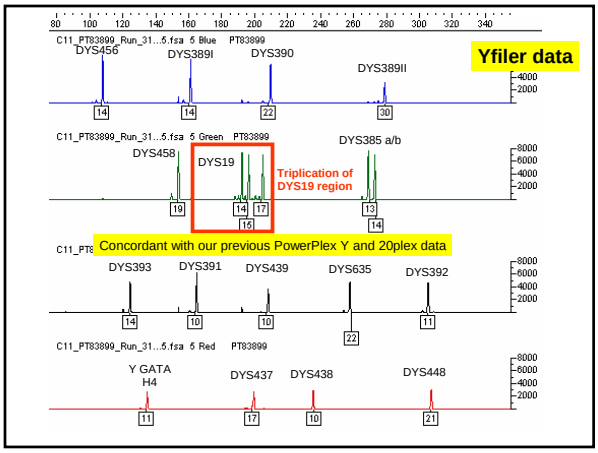


PowerPlex Y results on sample from Ann Marie Gross (MN BCA)

Locus Triplication at DYS19

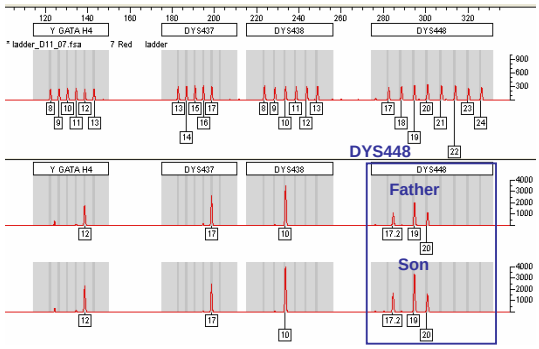


PowerPlex Y data



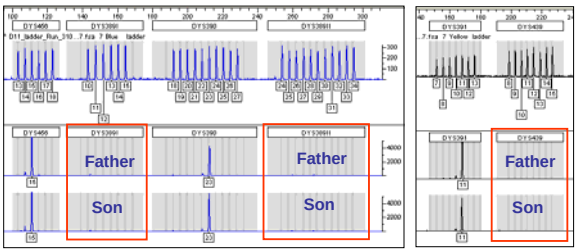
DYS448 Triplication

Seen in Both Father and Son

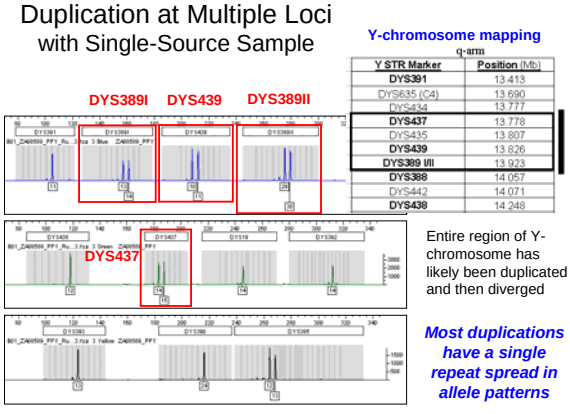


DYS389I, DYS389II, DYS439 Deletions

Seen in Both Father and Son



Duplication at Multiple Loci with Single-Source Sample



Y-chromosome mapping

q-arm

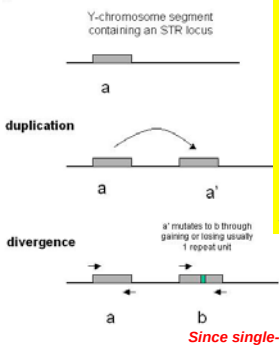
Y STR Marker	Position (Mb)
DYS391	13.413
DYS635 (C4)	13.690
DYS434	13.777
DYS437	13.778
DYS435	13.807
DYS439	13.826
DYS389 I/II	13.923
DYS388	14.057
DYS442	14.071
DYS438	14.248

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

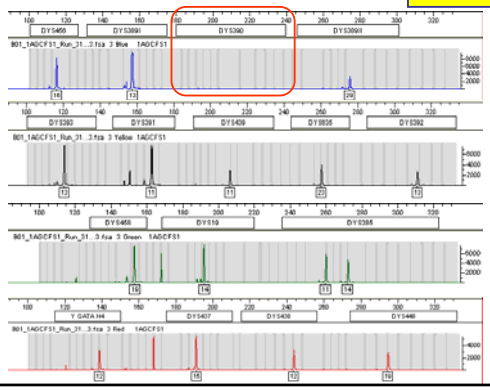
- ***Note the number of loci containing >1 allele (other than multi-copy DYS385)***
- ***Consider relative position on the Y-chromosome if multiple loci have two alleles***
- ***See if repeat spread is >1 repeat unit***
- ***Examine DYS385 for presence of >2 alleles***

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

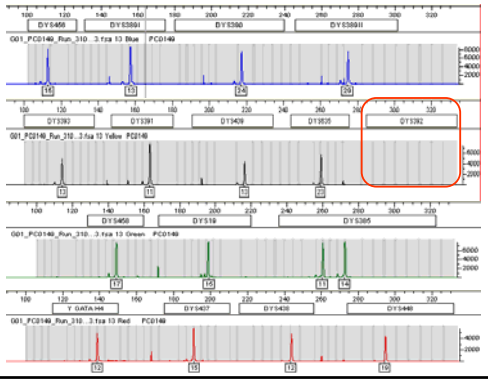
CFS-1 with Y filer

DYS390 is deleted



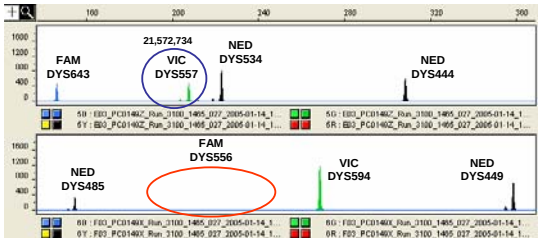
Sample PC0149 with Yfiler

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**

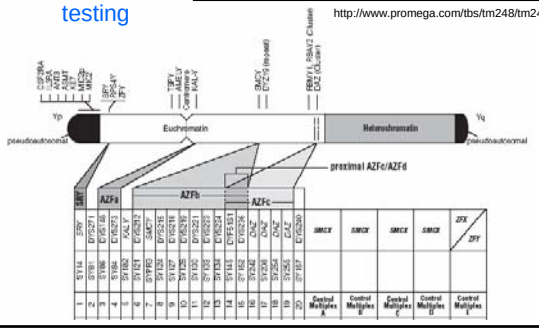
Promega sells a Y-deletion test for infertility testing

Y Chromosome Deletion Detection System, Version 2.0

Promega

Technical Manual No. 248

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



DYS19 Flanking Region Duplication

Int. J. Legal Med. (2004) 118: 178-183

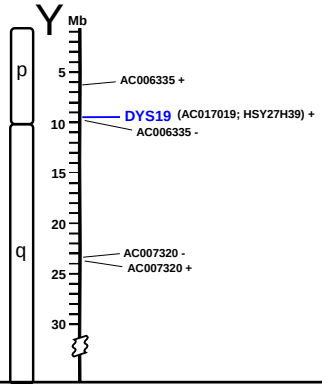
Int J Legal Med (2004) 118: 178–183
DOI 10.1007/s00414-004-0436-5

TECHNICAL NOTE

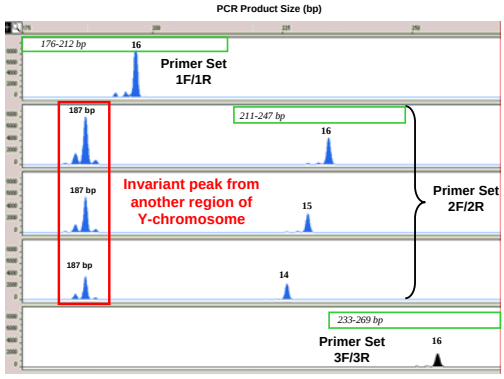
John M. Butler · Richard Schoske

Duplication of DYS19 flanking regions in other parts of the Y chromosome

Different primers around DYS19 repeat result in selection of different regions of the Y-chromosome

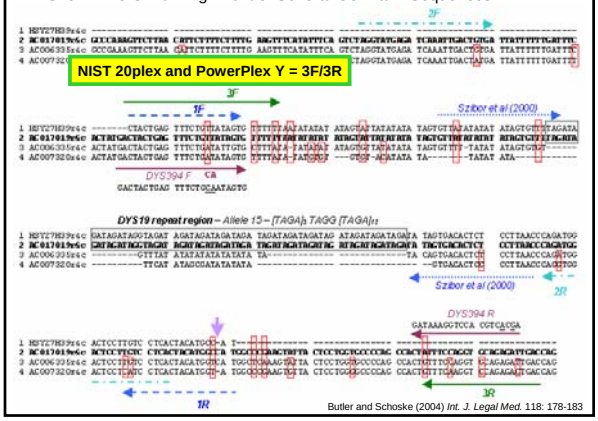


Impact of Different DYS19 Primers



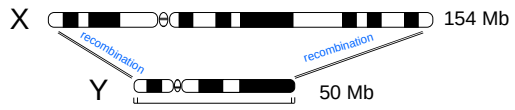
Butler and Schoske (2004) *Int. J. Legal Med.* 118: 178-183

DYS19 Primers with Alignment of Several GenBank Sequences



Butler and Schoske (2004) *Int. J. Legal Med.* 118: 178-183

X- and Y-Chromosomes Recombine at Their Tips



Pseudoautosomal Region 1 (PAR1)

Non-Recombining Portion of Y Chromosome (NRY)

Pseudoautosomal Region 2 (PAR2)

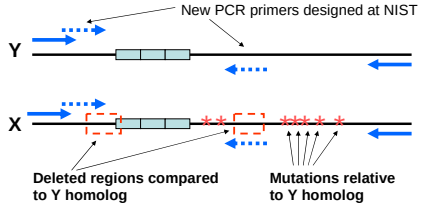
Male-specific region of the Y (MSY)

Focus in Forensics is on the NRY
PCR primers for Y-STRs need to be carefully designed to avoid X-chromosome homology

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.4(A), ©Elsevier Science/Academic Press

DYS391 Primer Improvements

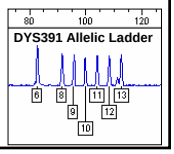
Female artifact problems seen with published and Y-PLEX 6 primers



Significant homology exists between X and Y sequences for the DYS391 locus

We have designed primers to anneal to regions that only appear on the Y chromosome (targeted X deletion regions)

These primers produce smaller PCR products and were adopted by Promega for their **PowerPlex Y kit**



Butler *et al.* (2002) *Forensic Sci. Int.* 129:10-24

Summary

- Mutation rates are similar to autosomal STRs (~0.2%) – based on father-son studies
- Variant alleles are observed as in autosomal STRs due to flanking region mutations, etc.
- Regions of the Y-chromosome can be duplicated or deleted causing Y-STRs to be duplicated or deleted
- Careful primer design is important to avoid X-chromosome homology or Y-chromosome duplications

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

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

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

7 matches in 27,773 individuals from 236 worldwide populations

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
Liguria, 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

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.

Difficult Questions…

- Which database(s) should be used for Y-STR profile frequency estimate determination?
- Are any of the current forensic Y-STR databases truly adequate for reliable estimations of Y-STR haplotype frequencies?
  - Some individuals share identical Y-STR haplotypes due to recurrent mutations, not relatedness…
  - Is the database a random collection reflecting Y-STR haplotype frequencies of the population?
  - Is the Y-STR haplotype frequency relevant for the population of the suspect?

Issues raised by Peter de Knijff at his Promega meeting presentation (Oct 2004)

Conclusions from Peter de Knijff

From his presentation at the Promega meeting (Oct 2004)

A haplotype frequency taken from any Y-STR database should not be reported or seen as a random match probability

- Because all male relatives have the same haplotype
- Males can share haplotypes without being related

Database estimates are at most qualitative…

What Peter de Knijff Reports with a Y-STR Match

From his presentation at the Promega meeting (Oct 2004)

- The Y-STR profile of the stain matches with the suspect.
- Therefore, the suspect cannot be excluded as the donor of the stain.
- On the basis of this DNA evidence, I **can also not exclude all paternally related male relatives of the suspect** as possible donors of this stain.
- In addition, **an unknown number of males from the same region cannot be excluded**. A more accurate answer can only be obtained if (1) we have detailed knowledge of the population structure of the region of interest, (2) the Y-STR frequencies therein are known, and (3) we have knowledge about the family structure of the suspect.

Can Y-STR results be combined with autosomal STR information?

- Still subject to some debate among experts (most say “yes”)
- Problem of different inheritance modes
- Multiply random match probability from the autosomal STR profile obtained with the upper bound confidence limit from the Y-STR haplotype frequency estimate