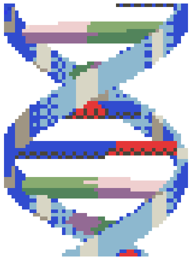
Y Chromosome DNA Variation Monitored by SNP and STR Analysis

John M. Butler

NIST Biotechnology Division

CHI 4th Annual DNA Forensics Meeting
June 1-2, 2000 (Springfield, VA)



NIST CSTL
BIOTECHNOLOGY

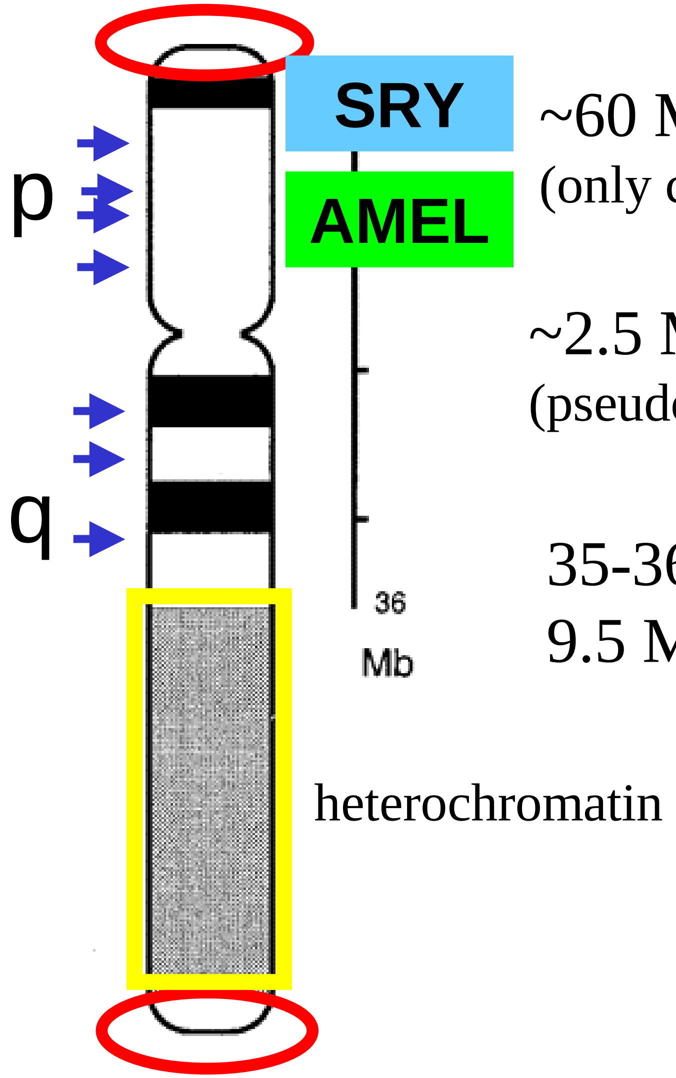
Presentation Outline

- Advantages of Y Chromosome Markers
- Y Chromosome Structure and Variation
- Commonly Used Y Chromosome Markers
- Y STR Multiplex Assays
- Population Studies
- Y SNP Markers
- Y Chromosome Work at NIST

Why the Y Chromosome?

- Applications
  - forensic investigations **(98% of violent crime by men)**
  - genealogical purposes
  - **evolutionary studies**
- Advantages to Human Identity Testing
  - male component isolated without differential extraction
  - paternal lineages
- Needs
  - population studies to evaluate diversity of haplotypes
  - robust assay for accurate characterization of Y markers

Y Chromosome Structure



Nucleic Acids Res. **28(2), e8 (2000)**

~60 Mb total DNA sequence
(only chromosome 22 is smaller)

~2.5 Mb on tips recombine with X
(pseudoautosomal regions)

35-36 Mb euchromatin
9.5 Mb sequenced (27%)

Genetic variation at multiple points along the Y chromosome is combined to form a Y haplotype for a sample

Y Chromosome DNA Markers

- STRs (microsatellites)
  - DYS19, DYS385, etc.
  - mostly tetranucleotide repeats
- Bi-allelic markers (unique event polymorphisms--UEP)
  - SNPs (single nucleotide polymorphisms)
  - Y *Alu* polymorphism (YAP) or other insertions/deletions *("indels")*
- Minisatellite
  - MSY1 (DYF155S1) composed of 48-114 copies of a 25 bp repeat unit with 5 sequence variant repeat types
  - typed by MVR-PCR (minisatellite variant repeat)

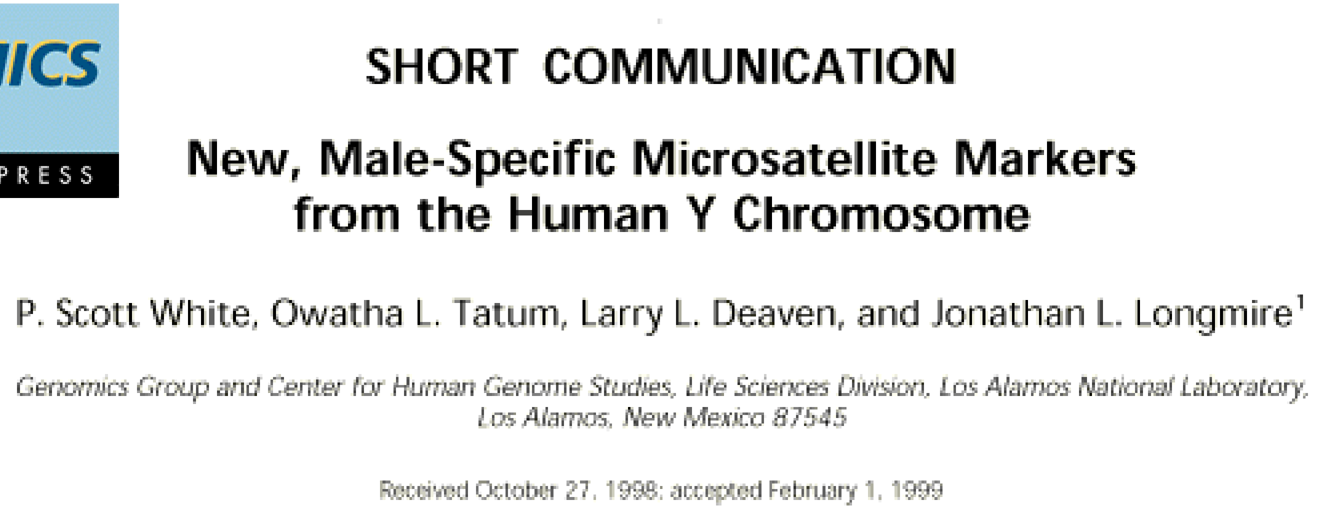
J.M. Butler, *Forensic DNA Typing*, Table 8.1

Y Chromosome STR Markers

Marker Name	GenBank Accession	Repeat Motif	Allele Range	PCR Product Sizes	Reference
DYS19	X77751	TAGA	8-16	178-210 bp	Roewer 1992
DYS385	Z93950	GAAA	10-22	252-300 bp	Schneider 1998
DYS388	G09695	ATT	12-17	128-143 bp	[illegible]
DYS389 I	G09600	(TCTG) (TCTA)	I: 7-13	239-263 bp	[illegible]
DYS389 II	G09600	(TCTG) (TCTA)	II:23-31	353-385 bp	[illegible]
DYS390	G09611	(TCTA) (TCTG)	18-27	191-227 bp	[illegible]
DYS391	G09613	TCTA	8-13	275-295 bp	[illegible]
DYS392	G09867	TAT	7-16	236-263 bp	[illegible]
DYS393	G09601	AGAT	9-15	108-132 bp	[illegible]
YCAIII	AC006370	CA	19-25	192-204 bp	Kayser 1997
DYS434	AC002992	ATCT	8-11	110-122 bp	Ayub 2000
DYS435	AC002992	TGGA	9-13	210-228 bp	Ayub 2000
DYS436	AC005820	GTT	10-15	128-143 bp	Ayub 2000
DYS437	AC002992	TCTA	8-11	186-202 bp	Ayub 2000
DYS438	AC002531	TTTTC	6-12	203-233 bp	Ayub 2000
DYS439	AC002992	AGAT	9-14	238-258 bp	Ayub 2000
Y-GATA-A4	G42670	AGAT	11-14	242-254 bp	White 1999
Y-GATA-A7.1	G42675	ATAG	7-12	161-181 bp	White 1999
Y-GATA-A7.2	G42671	TAGA	8-12	174-190 bp	White 1999
Y-GATA-A8	G42672	TCTA	8-14	219-244 bp	White 1999
Y-GATA-A10	G42674	TATC	11-14	160-172 bp	White 1999
Y-GATA-C4	G42673	TATC	11-16	251-271 bp	White 1999
Y-GATA-H4	G42676	TAGA	10-13	362-370 bp	White 1999

Most Commonly Used Markers

New Y STR Markers



SHORT COMMUNICATION

New, Male-Specific Microsatellite Markers from the Human Y Chromosome

P. Scott White, Owatha L. Tatum, Larry L. Deaven, and Jonathan L. Longmire[1]

Genomics Group and Center for Human Genome Studies, Life Sciences Division, Los Alamos National Laboratory,
Los Alamos, New Mexico 87545

Received October 27, 1998; accepted February 1, 1999

Genomics 57, 433-437 (1999)

Y-GATA-A4

Y-GATA-A7.1

Y-GATA-A7.2

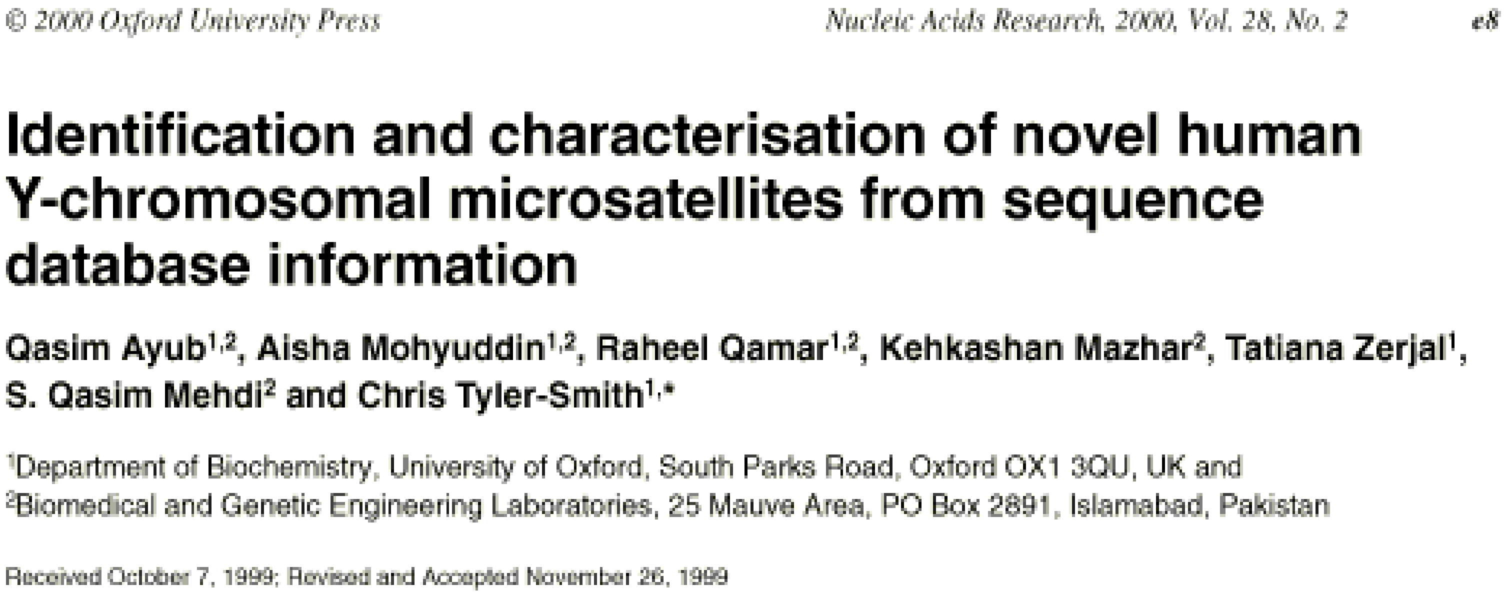
Y-GATA-A8

Y-GATA-A10

Y-GATA-C4

Y-GATA-H4

New Y STR Markers



Nucleic Acids Research, 2000, Vol. 28, No. 2 e8

Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information

Qasim Ayub[1,2], Aisha Mohyuddin[1,2], Raheel Qamar[1,2], Kehkashan Mazhar[2], Tatiana Zerjal[1], S. Qasim Mehdi[2] and Chris Tyler-Smith[1,*]

[1]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK and
[2]Biomedical and Genetic Engineering Laboratories, 25 Mauve Area, PO Box 2891, Islamabad, Pakistan

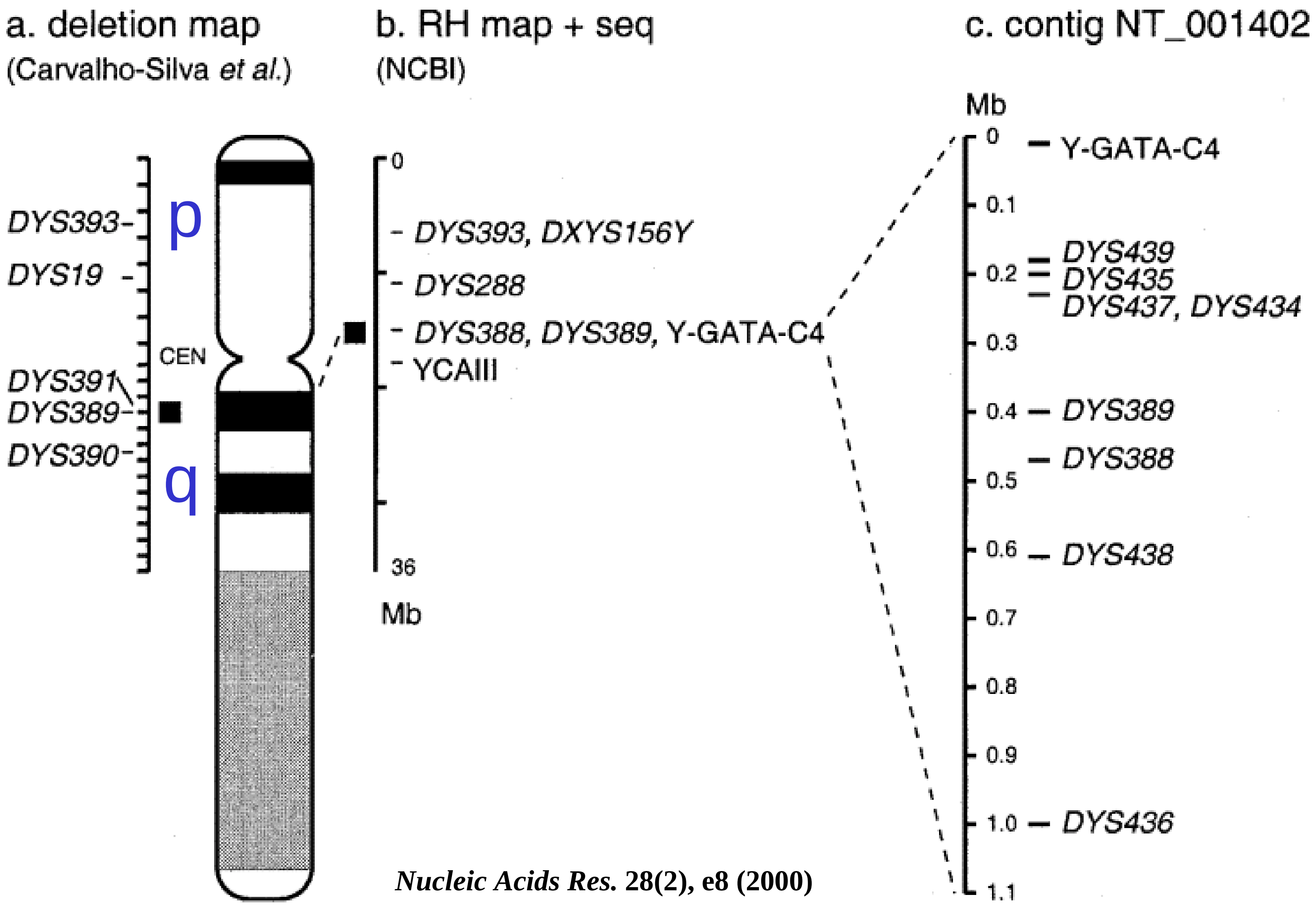
Received October 7, 1999; Revised and Accepted November 26, 1999

***Nucleic Acids Res.* 28(2), e8 (2000)**

DYS434 DYS436 DYS438

DYS435 DYS437 DYS439

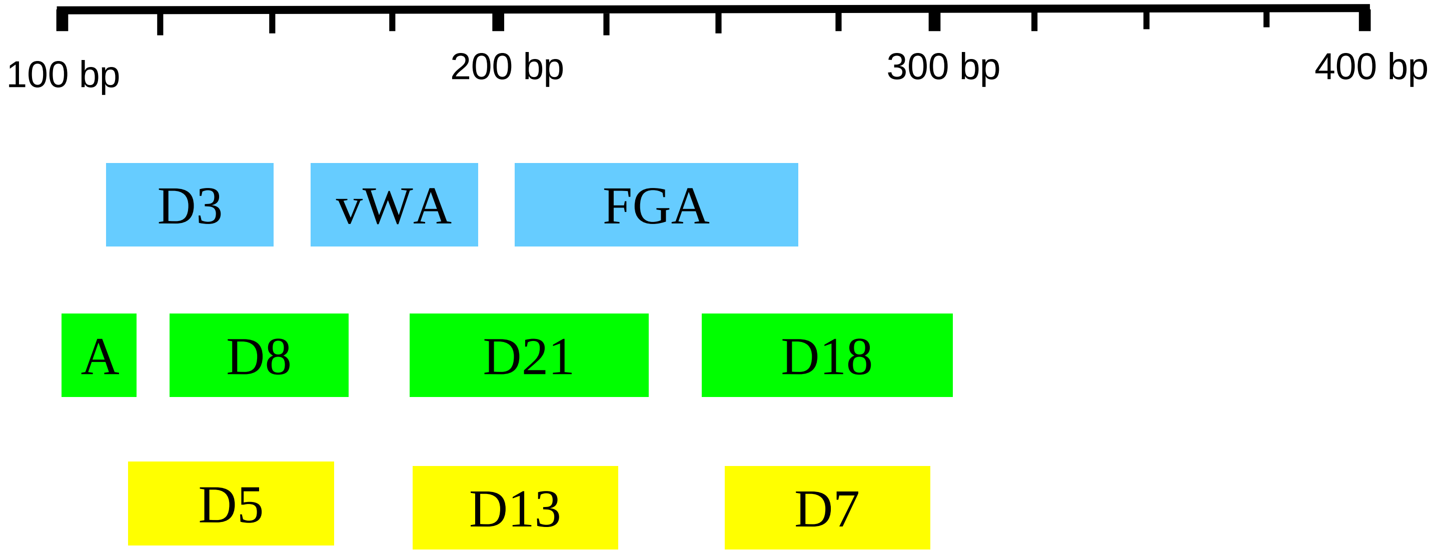
Map of Y Chromosome STR Markers



Nucleic Acids Res. **28(2), e8 (2000)**

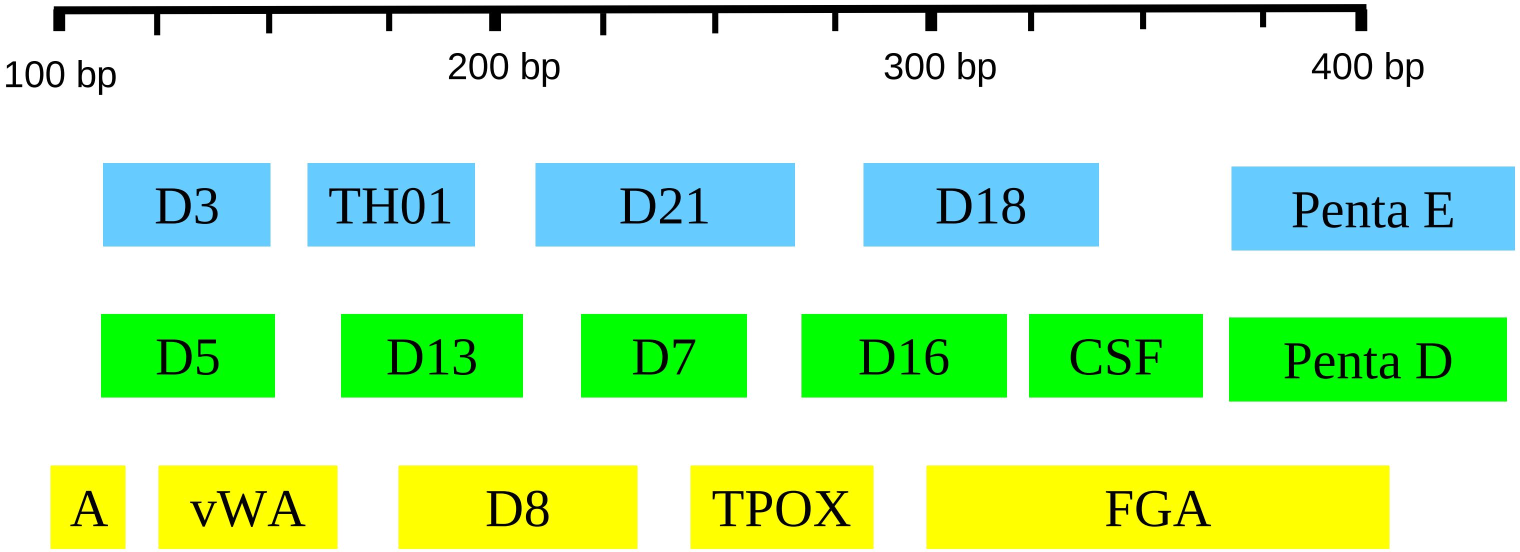
Current Forensic STR Multiplexes

Profiler Plus™



Current Forensic STR Multiplexes

PowerPlex™ 16



100 bp 200 bp 300 bp 400 bp

D3 TH01 D21 D18 Penta E

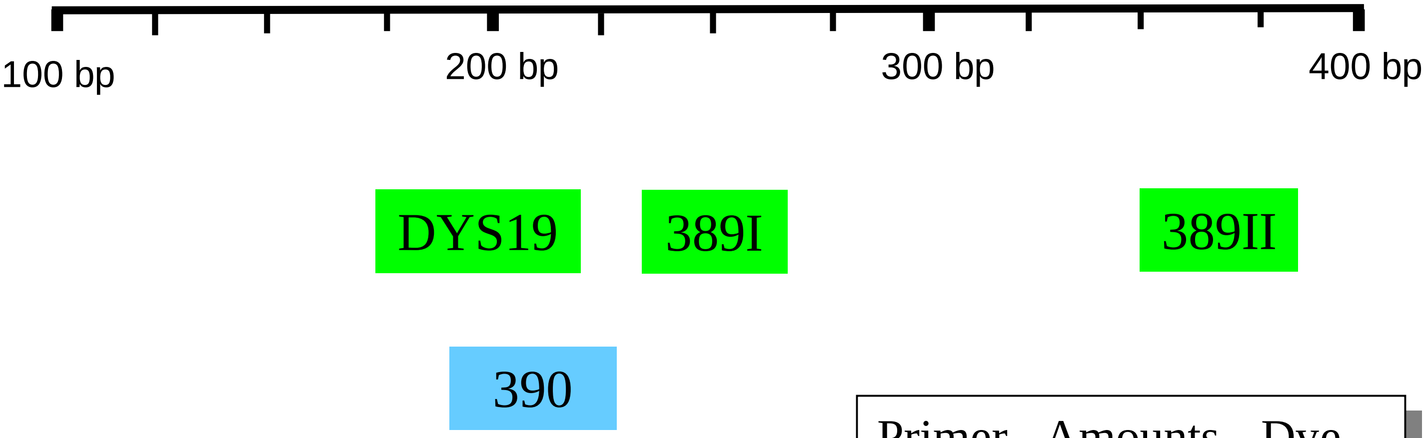
D5 D13 D7 D16 CSF Penta D

A vWA D8 TPOX FGA

Y STR Multiplex Assay

Prinz et al. 1997

(Forensic Sci Int, vol. 85, pp. 209-218)



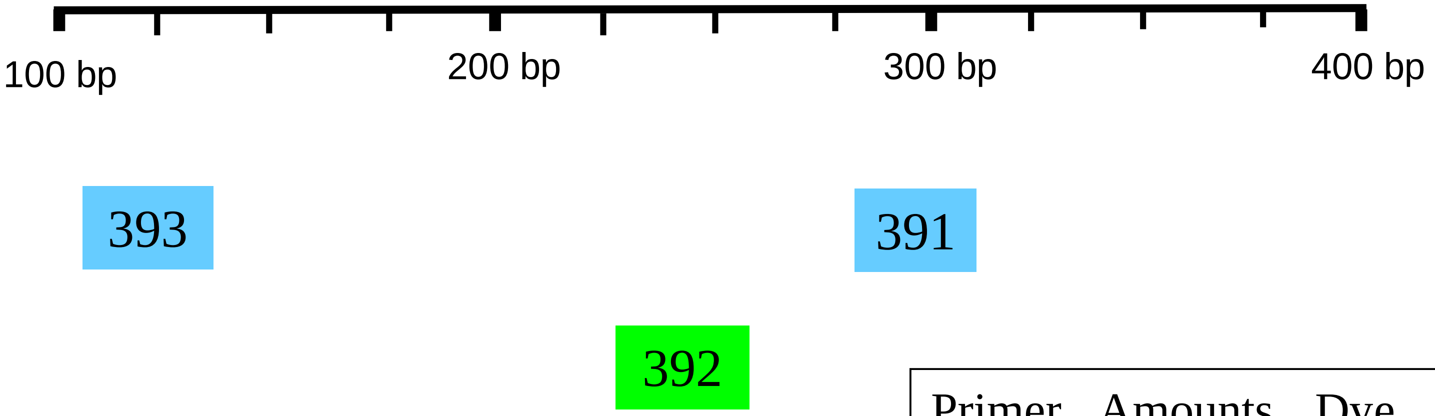
“Quadruplex I”

Primer	Amounts	Dye
Y19	0.25 µM	JOE
Y389	0.125 µM	FAM
Y390	0.25 µM	JOE

Y STR Multiplex Assay

Kayser et al. 1997

(Int J Legal Med, vol. 110, pp. 125-133)



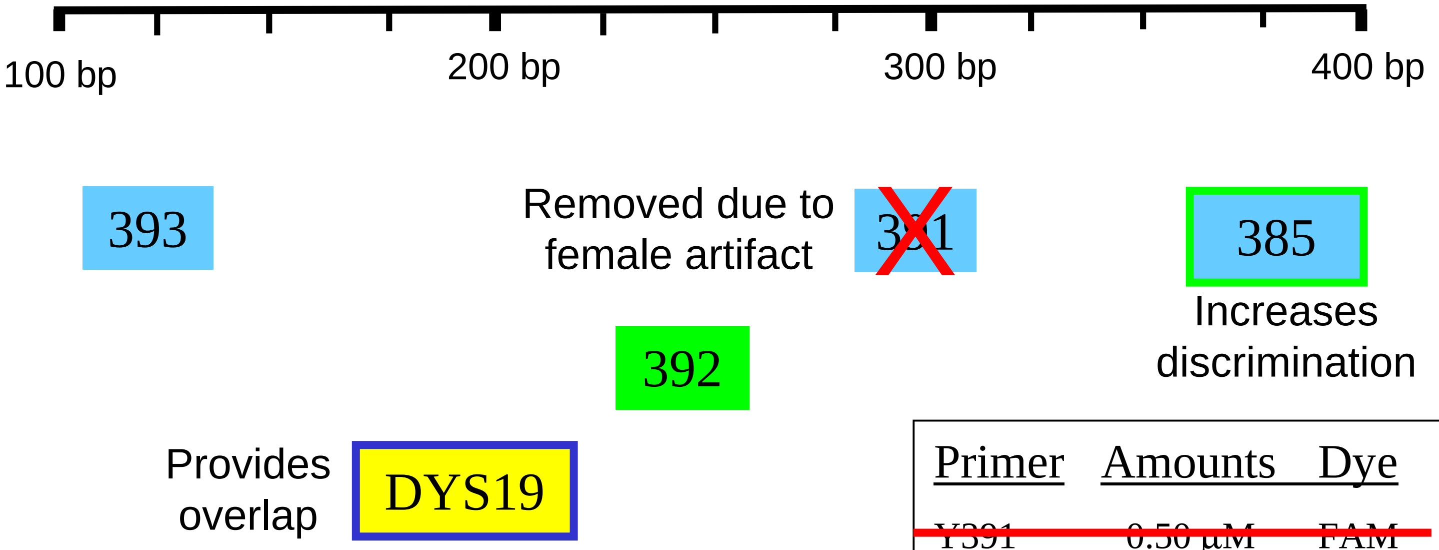
Primer	Amounts	Dye
Y391	0.3 µM	FAM
Y392	0.3 µM	JOE
Y393	0.06 µM	FAM

“Triplex I”

Y STR Multiplex Assay

Ishii et al. 1999

(Progress in Forensic Genetics 8, pp. 457-459)



Removed due to female artifact

Increases discrimination

Provides overlap

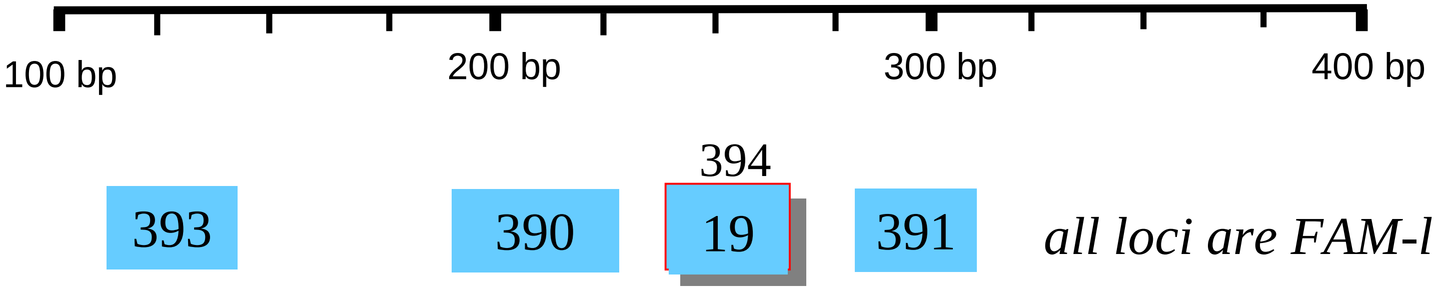
Modified “Triplex I”

Primer	Amounts	Dye
~~Y391~~	~~0.50 µM~~	~~FAM~~
Y392	0.30 µM	JOE
Y393	0.06 µM	FAM
Y19	0.20 µM	NED
Y385	0.05 µM	FAM

Y STR Multiplex Assay

Redd et al. 1997

(Biol. Chem. vol. 378, pp. 923-927)



all loci are FAM-labeled

4-plex

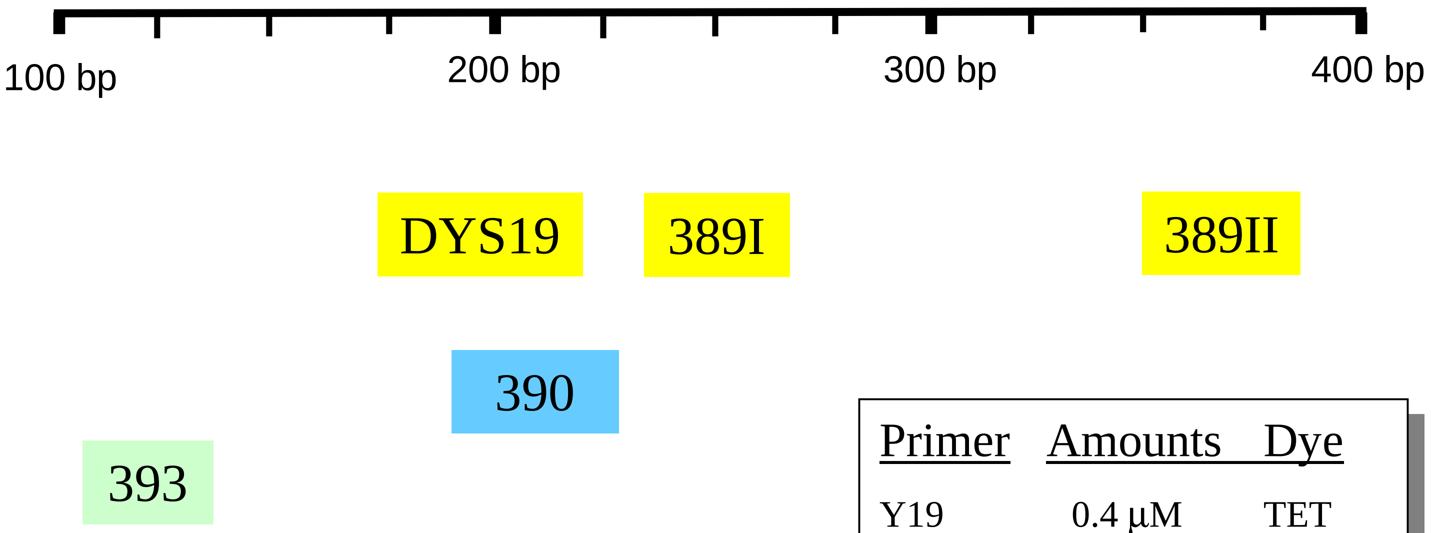
DYS394 primers were used in place of regular DYS19 primers
(amplicon is 44 bp larger)

Primer	Amounts
Y19	0.32 µM
Y390	0.16 µM
Y391	0.16 µM
Y393	0.10 µM

Y STR Multiplex Assay

Gusmão et al. 1999

(Forensic Sci Int, vol. 106, pp. 163-172)



Primer	Amounts	Dye
Y19	0.4 µM	TET
Y389	0.12 µM	TET
Y390	0.12 µM	FAM
Y393	0.2 µM	HEX

5-plex

Primer Sequences

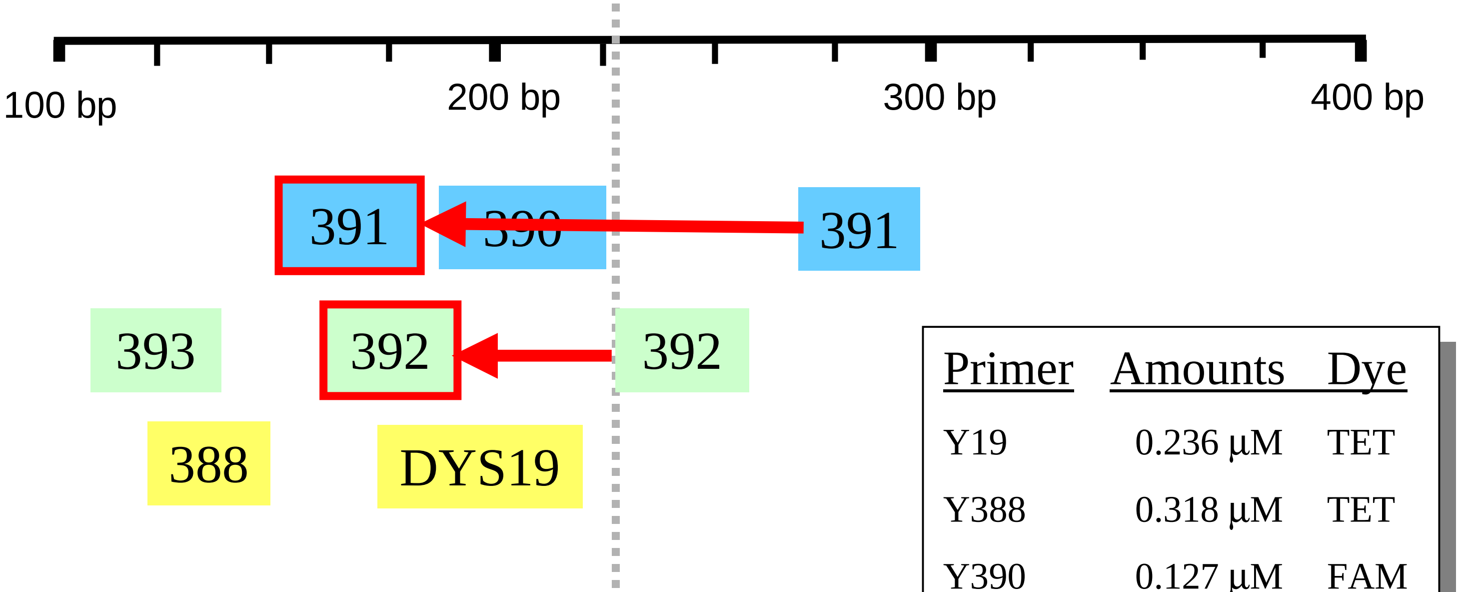
Locus	Repeat Motif	# Repeats	# Alleles	Allele size ranges (bp)	Gene Diversity (h)	PCR Primer Sequences
DYS19	(GATA)n	10-19	9	174-210	0.72	Primer A: CTACTG AGT TTC TGT TAT AGT Primer B: ATG GCA TGT AGT GAG GAC A Primer A2: GTT ATA TAT ATA TAG TGT TTAG 3) Primer B2: GTT AAG GAG AGT GTC ACT A 3)
DYS389I 1)	(GATA)n (GACA)n	9-16	7	235-263	0.61	Primer A:CCA ACT CTC ATC TGT ATT ATC TAT Primer B: TCT TAT CTC CAC CCA CCA GA Primer 2B: TTA TCC CTG AGT AGT AGA AGA AT 5)
DYS389II 1)	(GATA)n (GACA)n	26-33	8	355-383	0.75	see DYS389I
DYS390	(GATA)n (GACA)n	18-27	9	191-227	0.61	Primer A: TAT ATT TTA CAC ATT TTT GGG CC Primer B: TGA CAG TAA AAT GAA CAC ATT GC
DYS391	(GATA)n	7-14	8	271-299	0.49	Primer A: CTA TTC ATT CAA TCATAC ACC CA Primer B: GAT TCT TTG TGG TGG GTC TG
DYS392	(ATT)n	6-16	11	233-263	0.52	Primer A: TCA TTA ATC TAG CTT TTA AAA ACA A Primer B: AGA CCC AGT TGA TGC AAT GT
DYS393	(GATA)n	9-16	8	108-136	0.34	Primer A: GTG GTC TTC TAC TTG TGT CAA TAC Primer B: AAC TCA AGT CCA AAA AAT GAG G
DYS385	(GAAA)n	7-22	68	352-412	0.85	Primer A:AGC ATG GGT GAC AGA GCT A Primer B: GGG ATG CTA GGT AAA GCT G Primer 2B: CCA ATT ACA TAG TCC TCC TTC 4)
YCAII 2)	(CA)n	1-9	31	144 - 160	0.67	Primer A: TAT ATT AAA TAG AAG TAG TGA Primer B: TAT CGA TGT AAT GTT ATA TTA

*From **Y STR Database***
http://ystr.charite.de/

Y STR Multiplex Assay

Thomas et al. 1999

(Hum. Genet., vol. 105, pp. 577-581)



Primer	Amounts	Dye
Y19	0.236 μM	TET
Y388	0.318 μM	TET
Y390	0.127 μM	FAM
Y391	0.384 μM	FAM
Y392	0.155 μM	HEX
Y393	0.088 μM	HEX

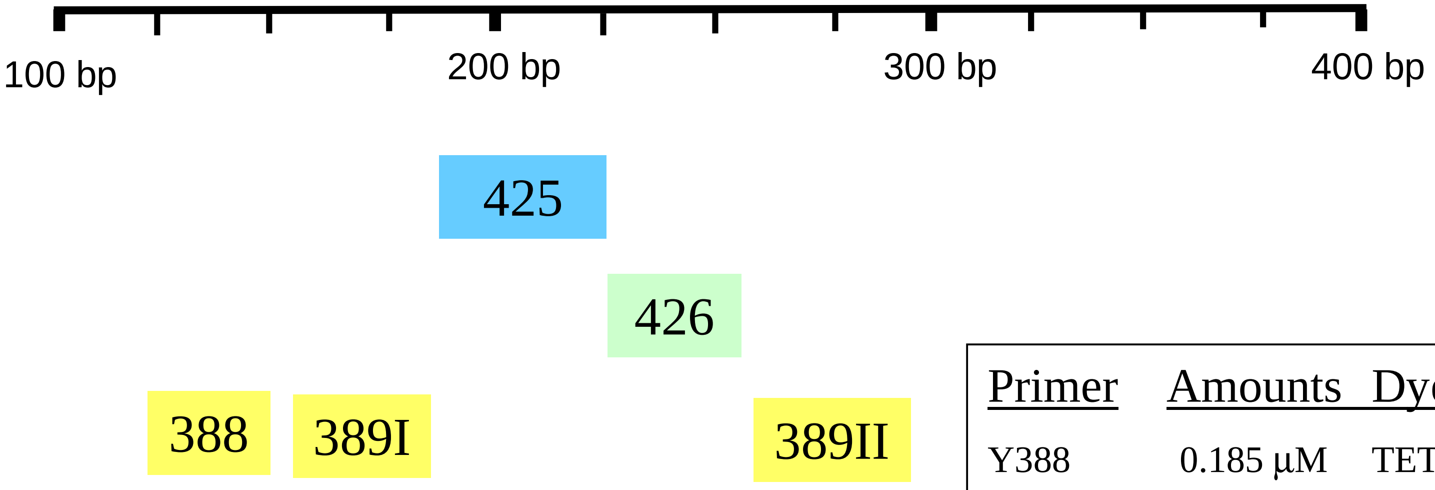
“MS1”

Microsatellite kit 1

Y STR Multiplex Assay

Thomas et al. 1999

(Hum. Genet., vol. 105, pp. 577-581)



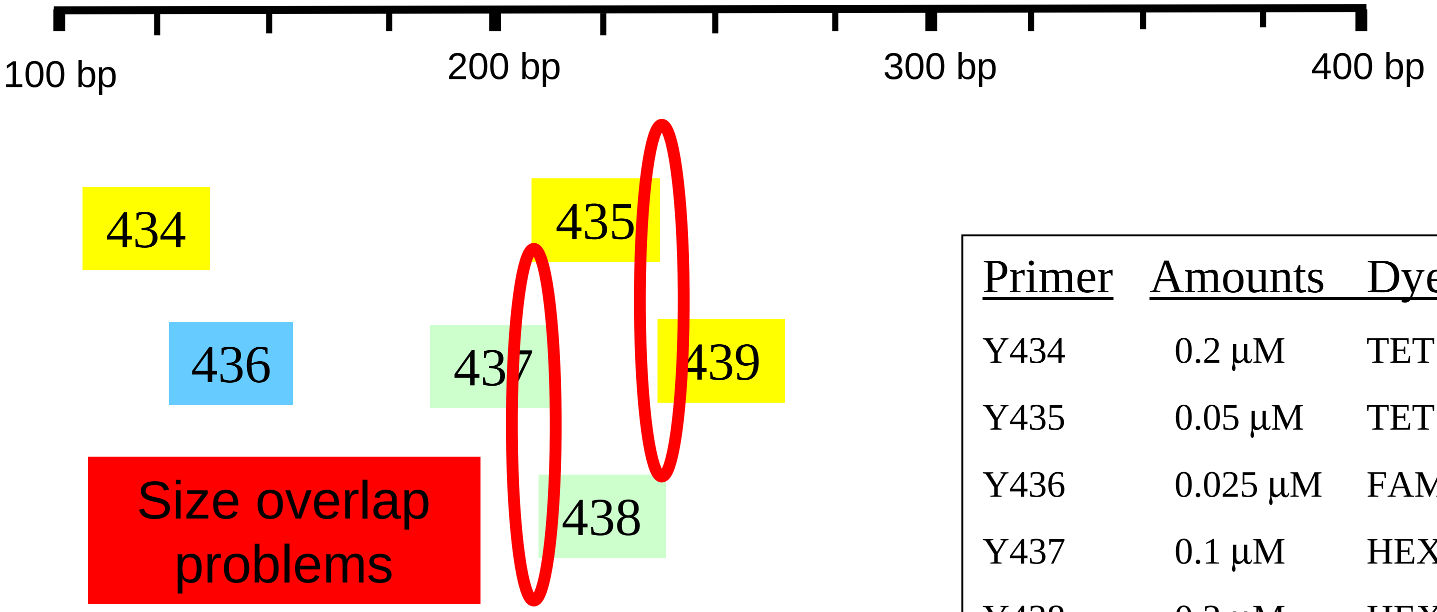
Primer	Amounts	Dye
Y388	0.185 µM	TET
Y389	0.350 µM	TET
Y425	0.150 µM	FAM
Y426	0.060 µM	HEX

“MS2”

Y STR Multiplex Assay

Ayub et al. 2000

(Nucleic Acids Res., vol. 28, e8)



Primer	Amounts	Dye
Y434	0.2 µM	TET
Y435	0.05 µM	TET
Y436	0.025 µM	FAM
Y437	0.1 µM	HEX
Y438	0.2 µM	HEX
Y439	0.2 µM	TET

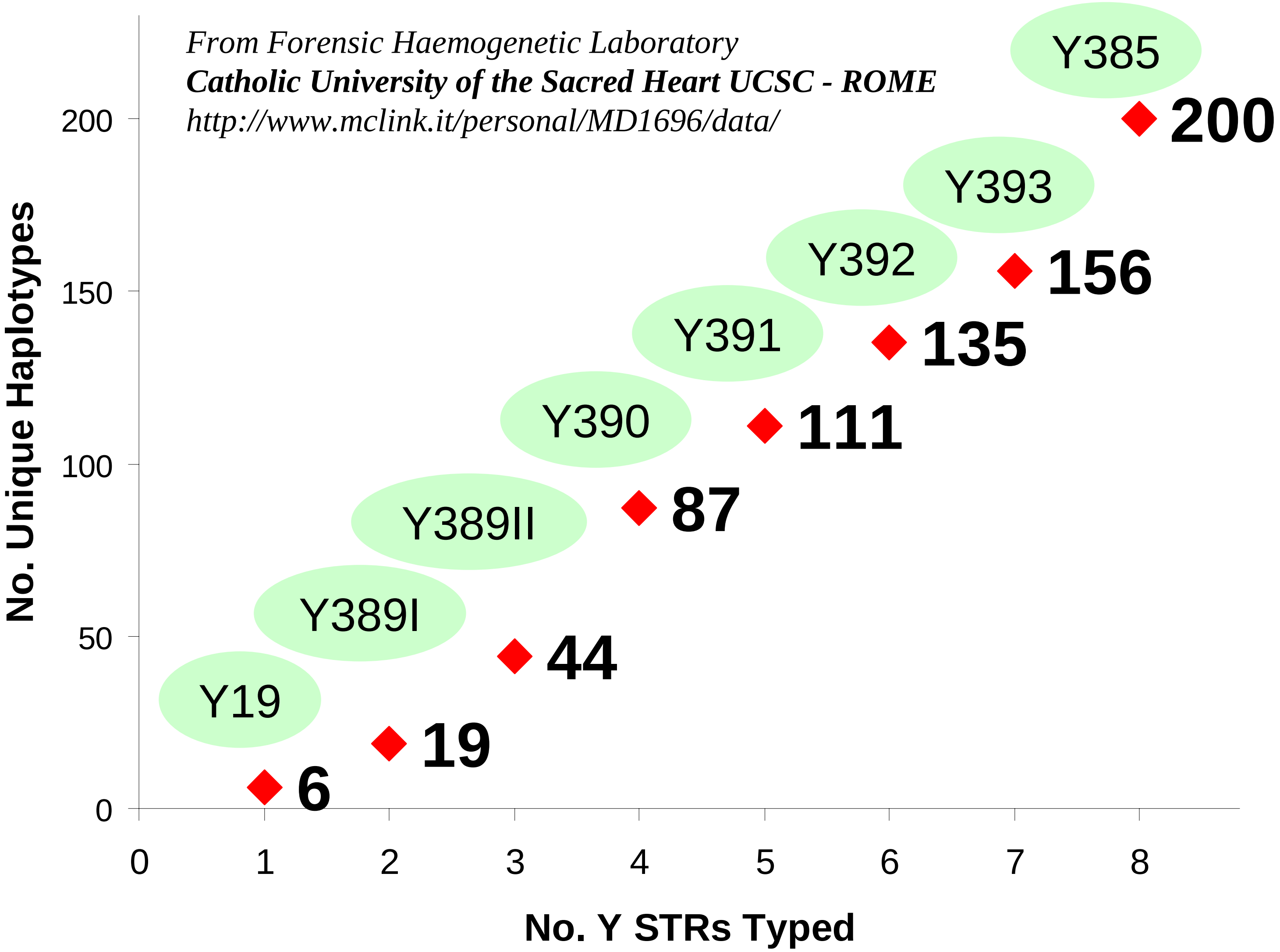
6-plex

“State of the Y STR Assay”

- A number of multiplex reactions have been reported in the literature but Y STR multiplexes have not reached their potential…
- Very little PCR optimization to-date (most work has been done with the original PCR primer sequences)
- No commercial Y STR kit exists yet (therefore these markers remain inaccessible to the general forensic DNA community)
- New Y STR markers are becoming available which will greatly improve the power of discrimination between unrelated individuals (e.g., DYS385) and these will need to be incorporated into future multiplex sets

Italian Population Haplotypes

216 Unrelated Males



From Forensic Haemogenetic Laboratory
Catholic University of the Sacred Heart UCSC - ROME
http://www.mclink.it/personal/MD1696/data/

Italian Population

216 Unrelated Males

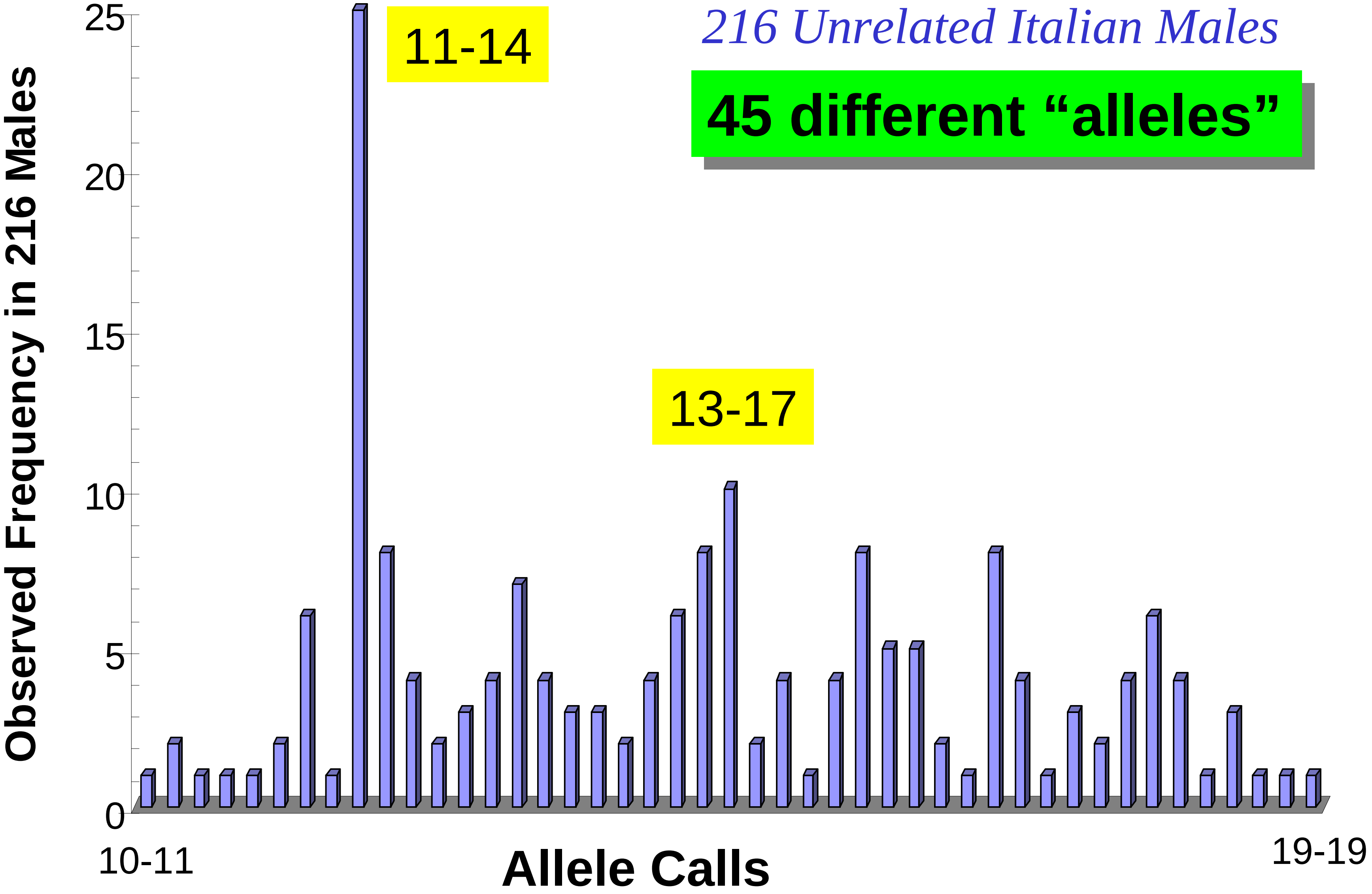
Y STR Locus	# Alleles Seen	# Repeat Units
DYS19	6	10, 13-17
DYS389I	4	9-12
DYS389II	6	24-29
DYS390	6	21-26
DYS391	4	9-12
DYS392	7	10-16
DYS393	5	11-15
DYS385	45	10-11—19-19

Dual product likely from a gene duplication

From Fo
Catholic ***ME***
http://ww

One primer pair produces two amplicons (behaves like a heterozygous locus)

DYS385 Allele Frequencies



216 Unrelated Italian Males

45 different “alleles”

11-14

13-17

Haplotypes vs. # Males Tested

Population Examined	# Males Tested	DNA Markers Used	# Observed Haplotypes	Discrimination Capacity	Reference (author, year)
Norwegian	300	9 STRs + Tat	225	0.75	Dupuy 2000
Italian	100	9 STRs	99	0.99	Ricci 2000
German	154	8 STRs	124	0.81	Henke 2000
Polish	107	9 STRs	88	0.82	Kupiec 2000
Japanese	154	5 STRs	95	0.62	Mukoyama 2000
SE China	51	7 STRs	48	0.94	Gusmao 2000
Portuguese, Central	162	8 STRs	142	0.88	Carvalho 2000
Portuguese, Azores	63	8 STRs	51	0.81	Carvalho 2000
Valencia (East Spain)	140	8 STRs	94	0.67	Aler 2000

Dupuy, B.M., et al. (2000) Progress in Forensic Genetics 8, pp. 260-262
Ricci, U., et al. (2000) Progress in Forensic Genetics 8, pp. 275-277
Henke, L., et al. (2000) Progress in Forensic Genetics 8, pp. 296-298
Kupiec, T., et al. (2000) Progress in Forensic Genetics 8, pp. 312-314
Mukoyama, R., et al. (2000) Progress in Forensic Genetics 8, pp. 318-320
Gusmao, L., et al. (2000) Progress in Forensic Genetics 8, pp. 324-326
Carvalho, M., et al. (2000) Progress in Forensic Genetics 8, pp. 302-304
Aler, M., et al. (2000) Progress in Forensic Genetics 8, pp. 305-308

18th Congress of the International Society of Forensic Haemogenetics (ISFH), San Francisco 1999

Y STR DATABASE

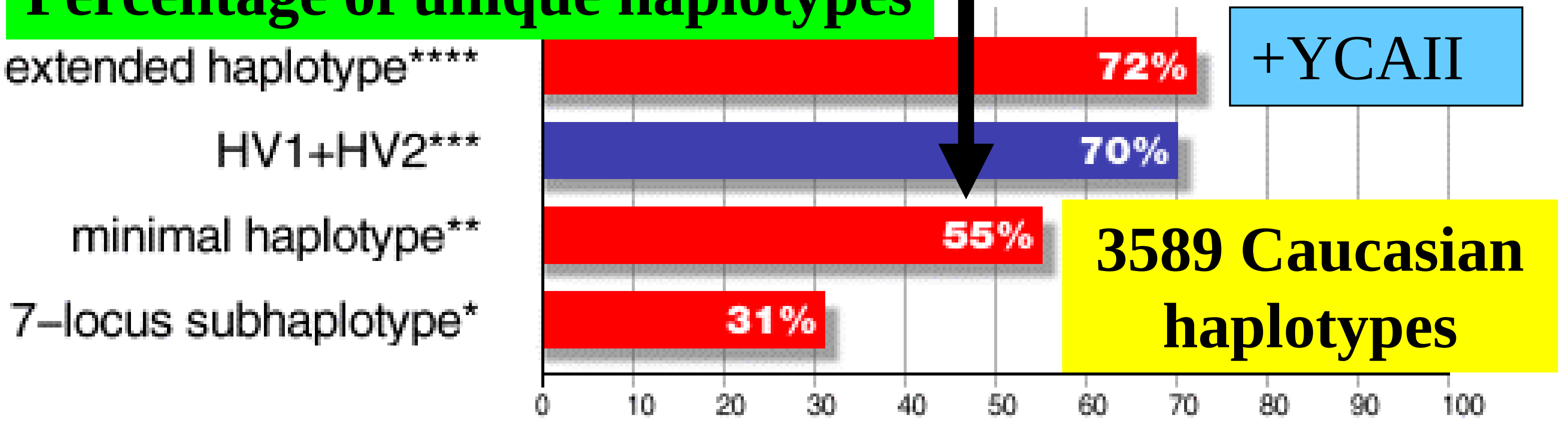
Y-STR Haplotype Reference Database

http://ystr.charite.de/

8 best studied Y STRs

DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385

Percentage of unique haplotypes



+YCAII

3589 Caucasian haplotypes

n = 3589 7-locus Caucasian haplotypes, DYS385 excluded, logged in the database
n = 3589 minimal Caucasian haplotypes logged in the database
n = 850 mt-DNA D-Loop sequences (from Institute of Legal Medicine Magdeburg, Germany)
n = 1650 extended Caucasian haplotypes logged in the database

Summary of Y DNA Population Variation

- Fairly significant discrimination powers can be achieved when using many Y STR markers…very dependent on the population samples selected
- Population sub-structure exists and is more significant for Y SNPs
- We will need larger databases of Y STRs and Y SNPs for calculating powers of discrimination for Y haplotypes (for the same reasons as mtDNA)

Y Bi-allelic Markers

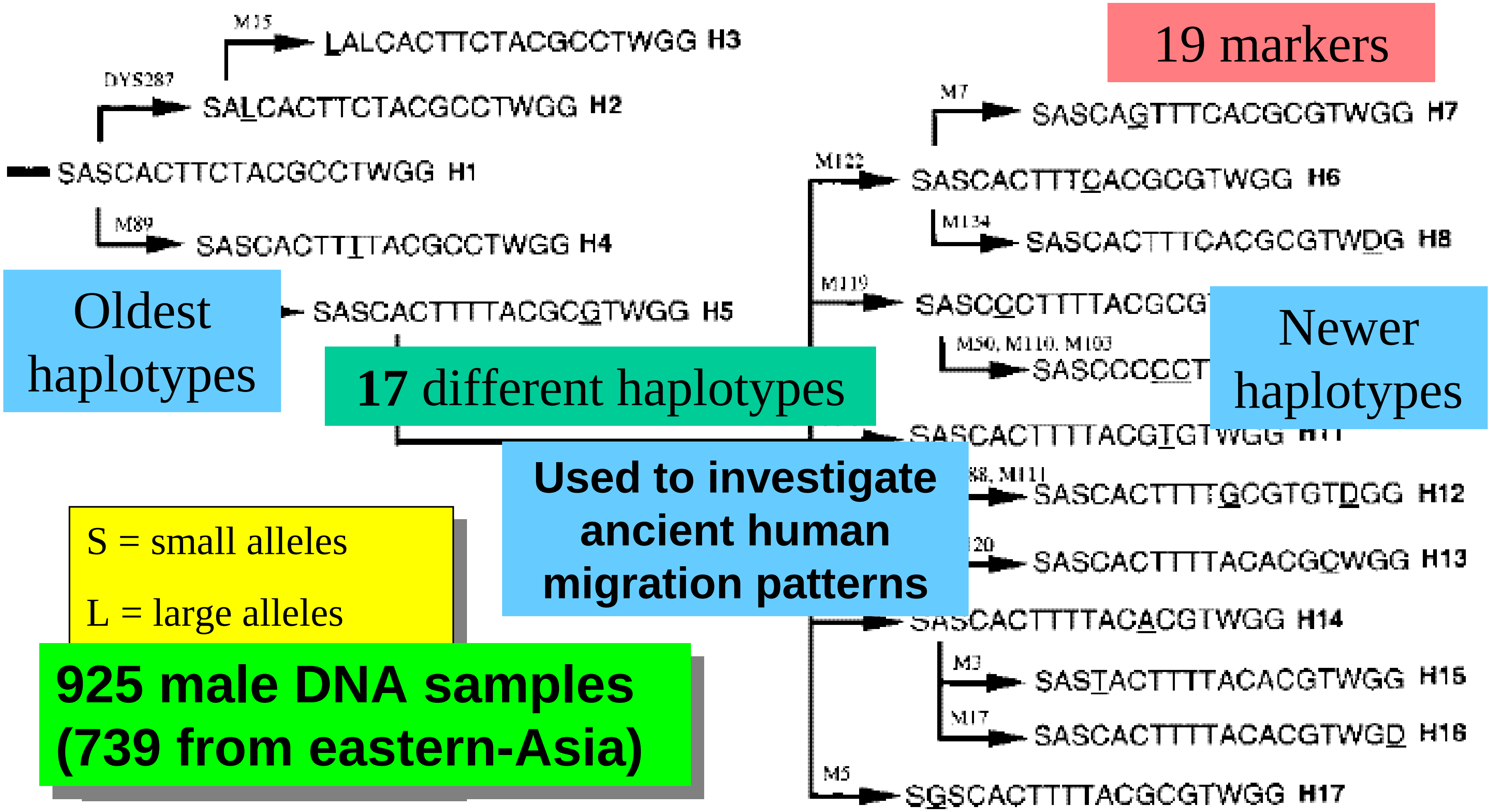
Marker Name	Polymorphism	PCR Product Size (bp)	Restriction Enzyme (PCR-RFLP assay)	Reference
YAP (DYS287)	Alu insertion (305 bp)	150 (YAP-)/ 455 (YAP+)		Hammer 1995
SRY-8299	G→A	509	*Bsr*BI	Whitfield 1995
SRY-1532	G→A	167	*Dra*III	Kwok 1996
SRY-2627	C→T	391	*Ban*I	Veitia 1997
sY81 (DYS271)	A→G	209	*Hsp*92II	Seielstad 1994
M9	C→G	340	*Hinf*I	Underhill 1997
92R7	C→T	55	*Hind*III	Mathias 1994
LLY22g	C→A	210	*Hind*III	Zerjal 1997
Tat	T→C	112	*Hsp*92II , *Mae*II	Zerjal 1997
DYS199	C→T	201	*Mfe*I	Santos 1999
SRY-465	C→T	123	*Fnu*HI	Yuasa 2000

154 Y SNPs from Peter Underhill (Stanford) discovered by DHPLC

Y Haplotypes with Bi-allelic Markers

M15-M5-DYS287-M3-M119-M7-M50-M110-M89-M122-M88-M103-M45-M95-M9-M120-M111-M134-M17

19 markers



Oldest haplotypes

17 different haplotypes

Newer haplotypes

Used to investigate ancient human migration patterns

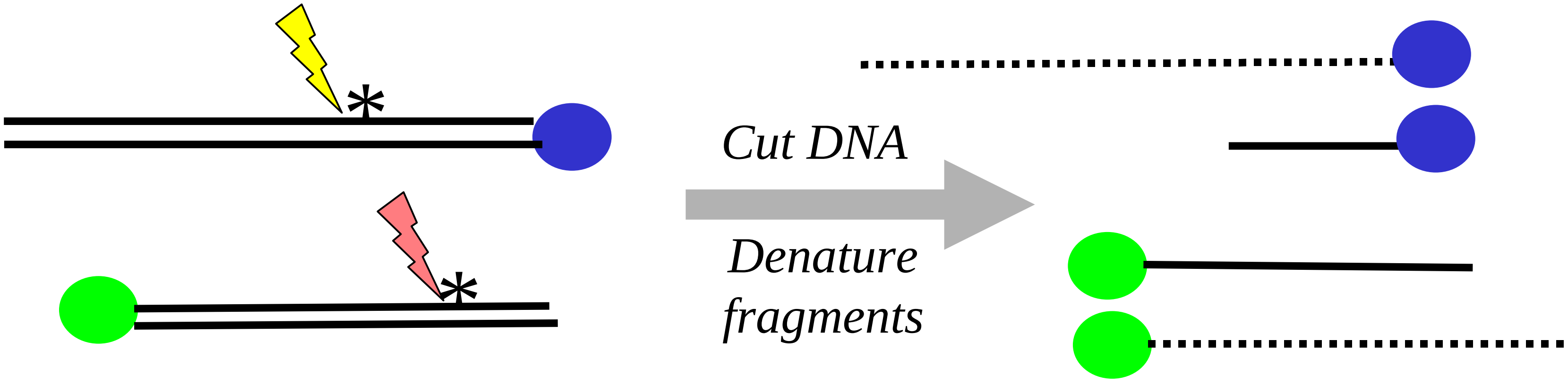
S = small alleles

L = large alleles

925 male DNA samples (739 from eastern-Asia)

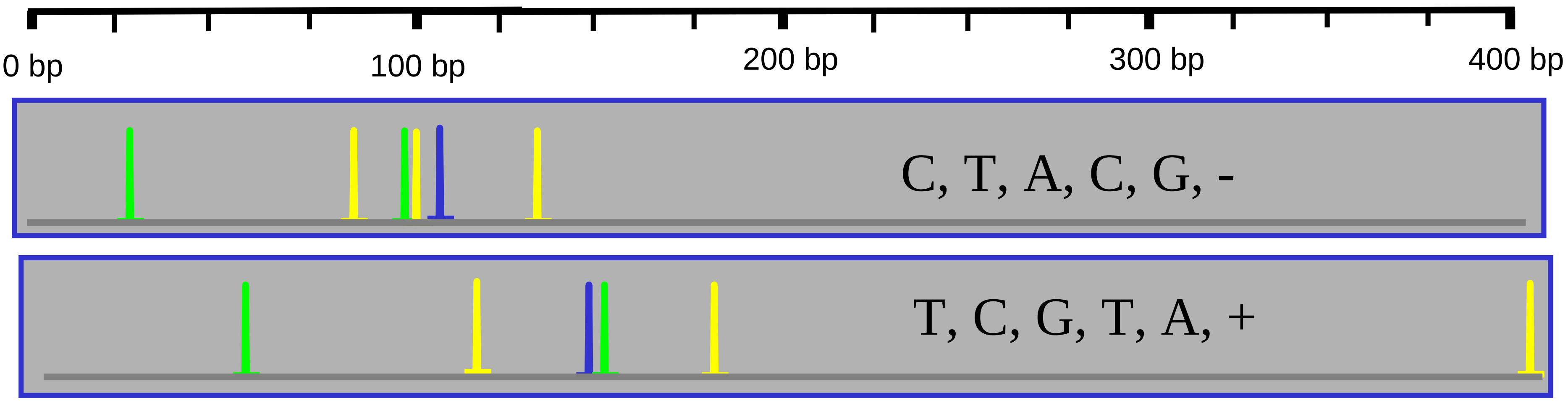
Su, B., *et al*. (1999) *Am. J. Hum. Genet*. 65: 1718-1724

Multiplex Restriction Enzyme Digestion



Cut DNA

Denature fragments



C, T, A, C, G, -

T, C, G, T, A, +

92R7, Tat, sY81, SRY+465, SRY4064, YAP

Multiplex Restriction Digestions

M.G. Thomas et al. 1999

(Hum. Genet., vol. 105, pp. 577-581)

“UEP1” (6 Y SNPs assayed)

Locus	Enzyme		PCR Size		Labeled size (cut)	
92R7	*Hind*III	**HEX**	55 bp	**R**	28 (C)	53 (T)
Tat	*Nla*III	**TET**	112 bp	**R**	83 (T)	112 (C)
sY81	*Nla*III	**FAM**	142 bp	**F**	105 (A)	142 (G)
SRY+465	*Fnu*4HI	**HEX**	148 bp	**R**	98 (C)	148 (T)
SRY4064	*Bsr*BI	**TET**	225 bp	**F**	135 (G)	174 (A)
YAP	n/a	**TET**	99/413 bp	**F**	99 (-)	413 (+)

Multiplex Restriction Digestions

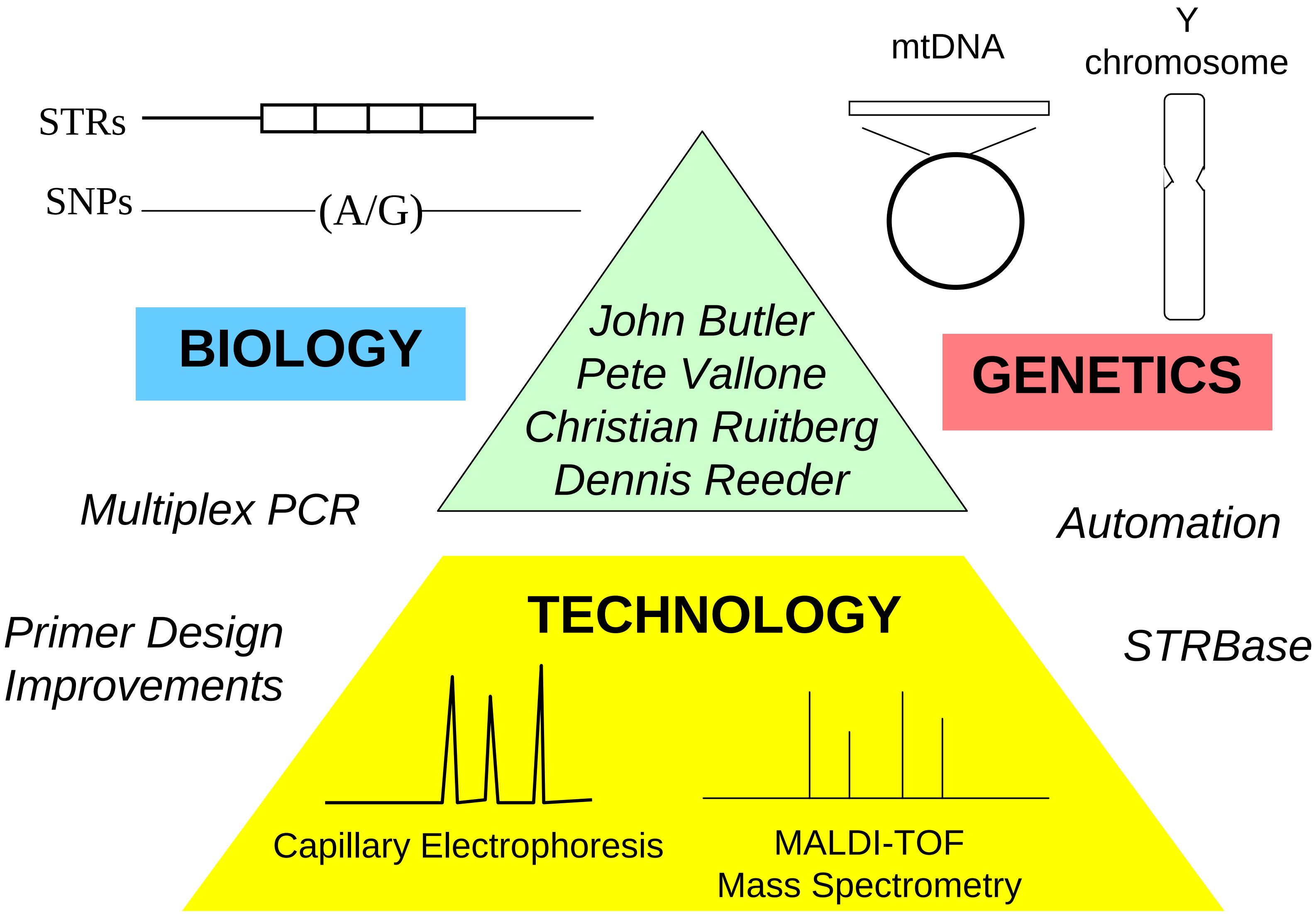
M.G. Thomas et al. 1999

(Hum. Genet., vol. 105, pp. 577-581)

“UEP2” (5 Y SNPs assayed)

Locus	**Enzyme**		**PCR Size**		**Labeled size (cut)**	
M9	*Hin*fI	**TET**	214 bp	**F**	48 (C)	76 (G)
M13	*Bsp*143I	**HEX**	119 bp	**F**	56 (G)	119 (C)
M17	*Afl*III	**TET**	124 bp	**R**	101 (.G)	124 (+G)
M20	*Sxp*I	**FAM**	106 bp	**F**	62 (A)	102 (G)
SRY10831	*Dra*III	**FAM**	73 bp	**F**	41 (G)	73 (A)

NIJ-Funded RESEARCH EFFORT at NIST



STRs

SNPs (A/G)

mtDNA

Y chromosome

BIOLOGY

John Butler
Pete Vallone
Christian Ruitberg
Dennis Reeder

GENETICS

Multiplex PCR

Automation

TECHNOLOGY

Primer Design Improvements

STRBase

Capillary Electrophoresis

MALDI-TOF Mass Spectrometry

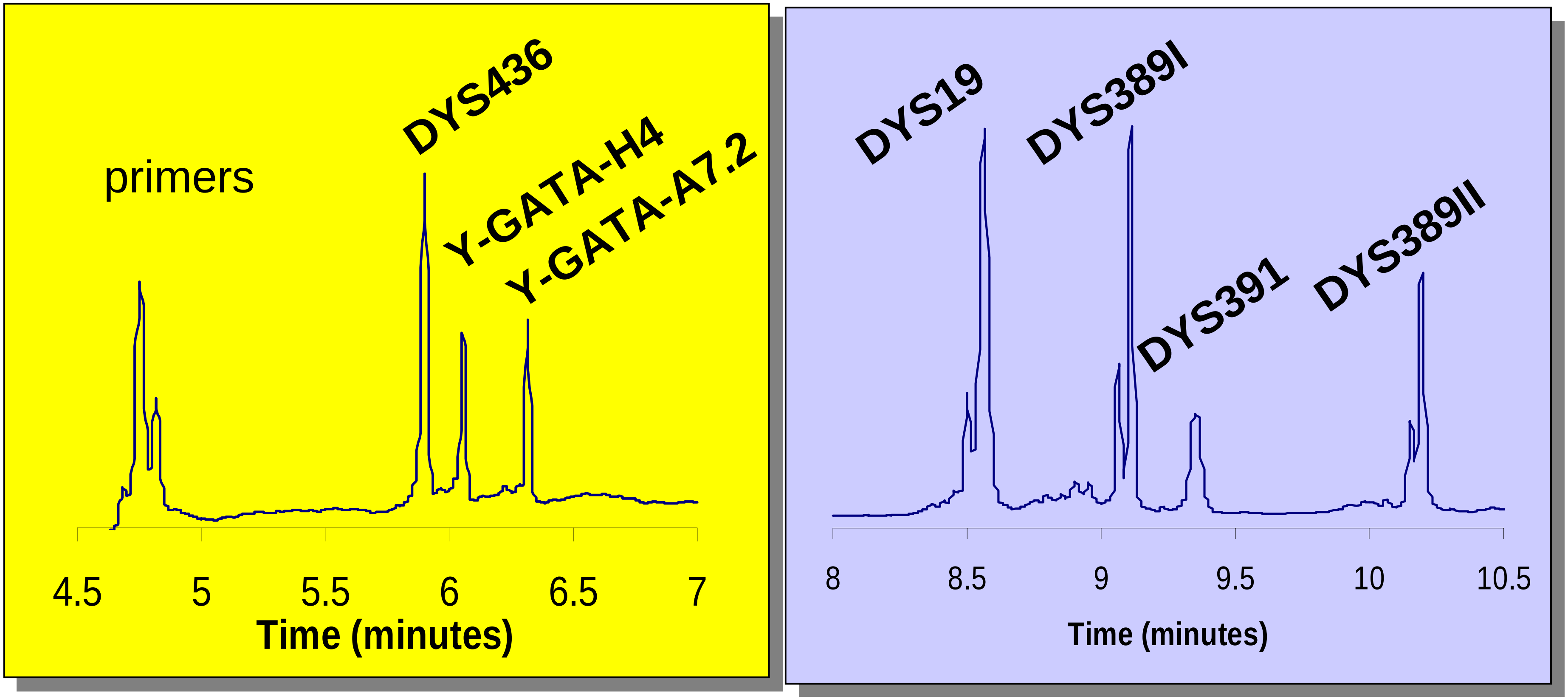
NIST Work with Y Markers

- Improvement in Y STR Multiplex Assays
- Development of Y STR Standard Reference Material (haplotypes for SRM2391a components)
- Multiplex Y SNP Assay Development and Marker Evaluation
- Rapid Construction of Y Haplotype Databases Using High-throughput Capabilities of Time-of-Flight Mass Spectrometry

Y STR Multiplex Assays

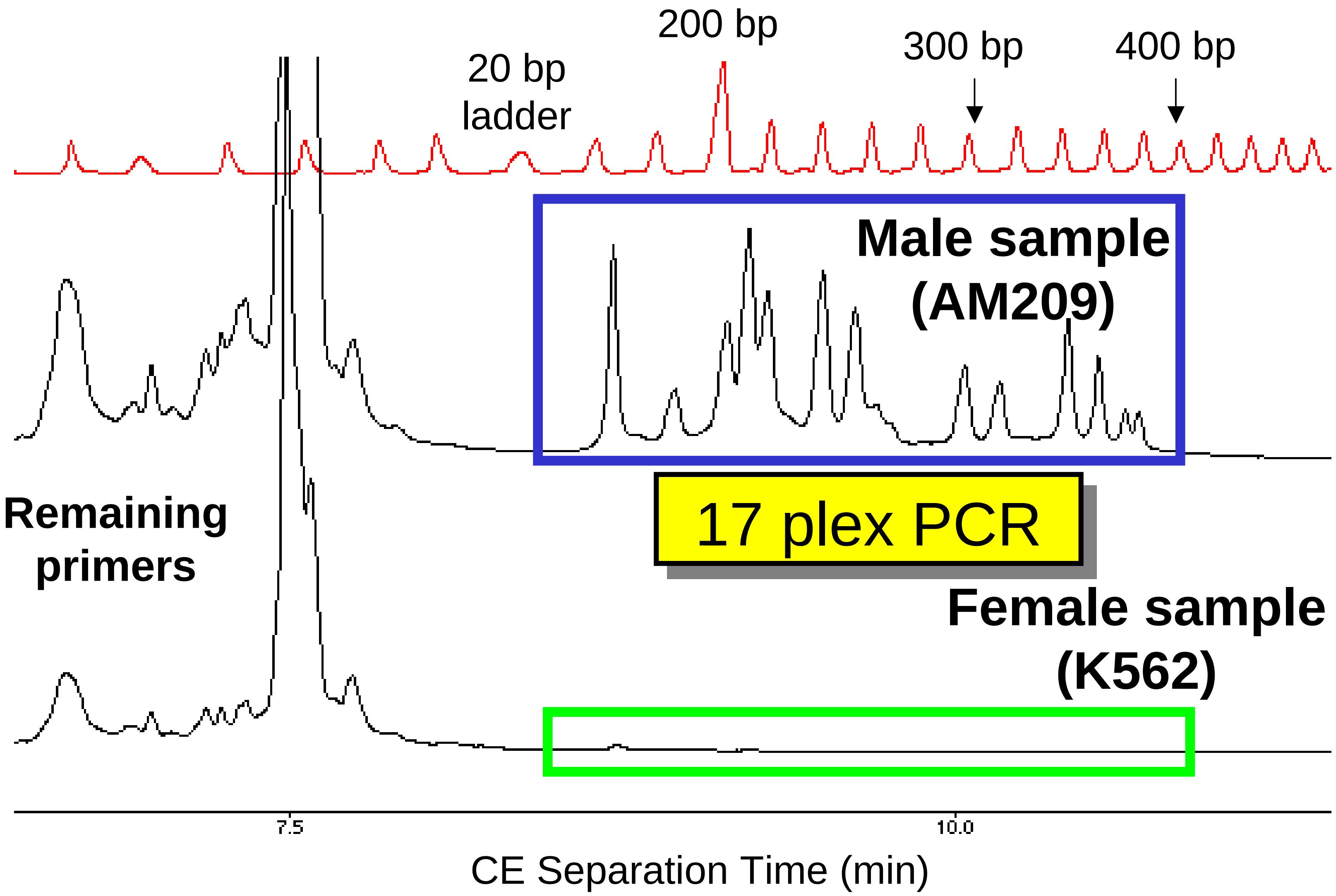
Capillary electrophoresis used for rapid separation and quantitation to adjust balance between loci and achieve robust multiplexes

Unlabeled amplicons with intercalating dyes (single-color)

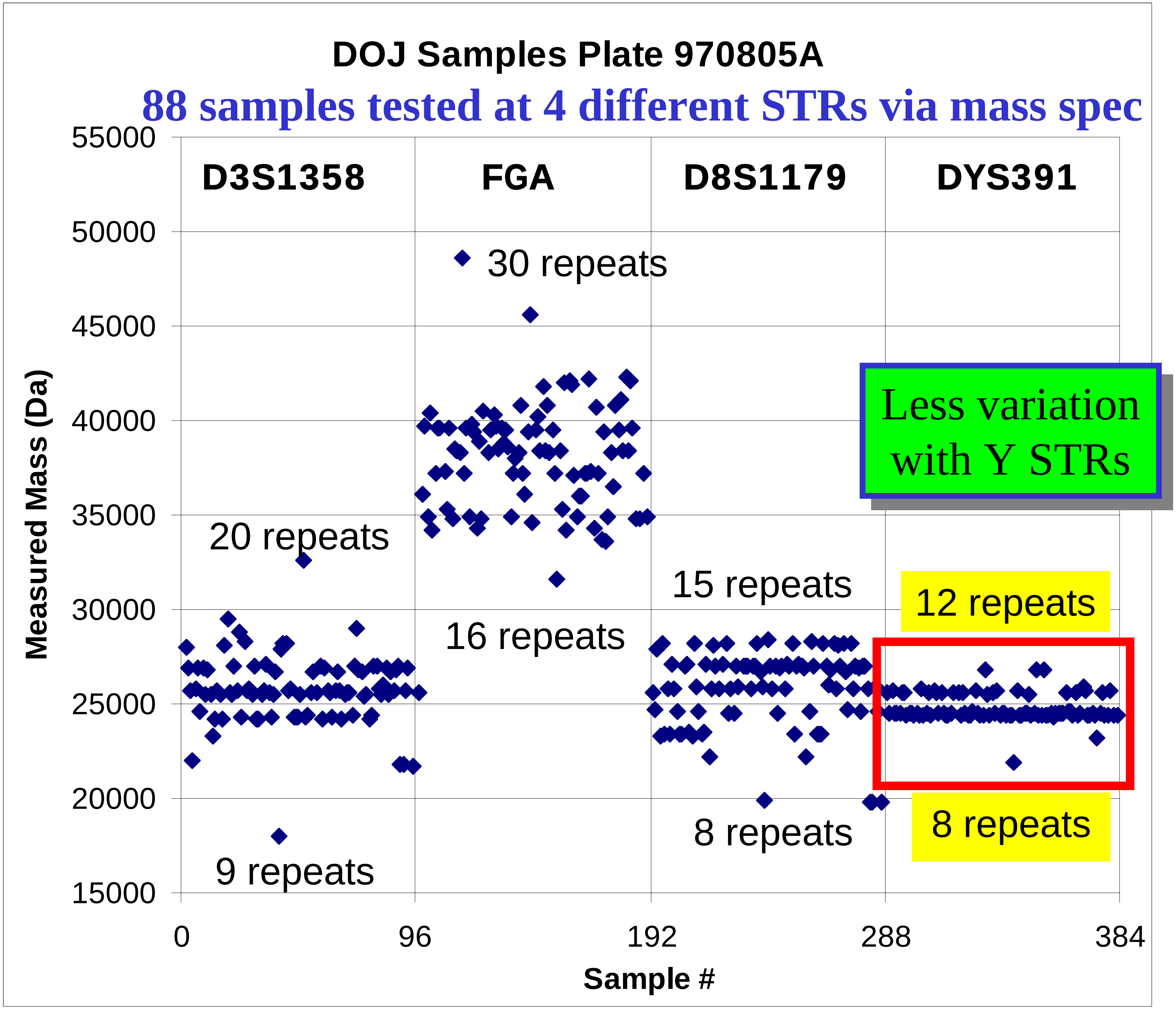


Rapid CE Separation

Male-Specific Multiplex PCR at 17 Y SNP Loci



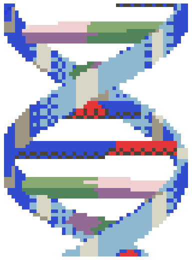
88 samples tested at 4 different STRs via mass spec



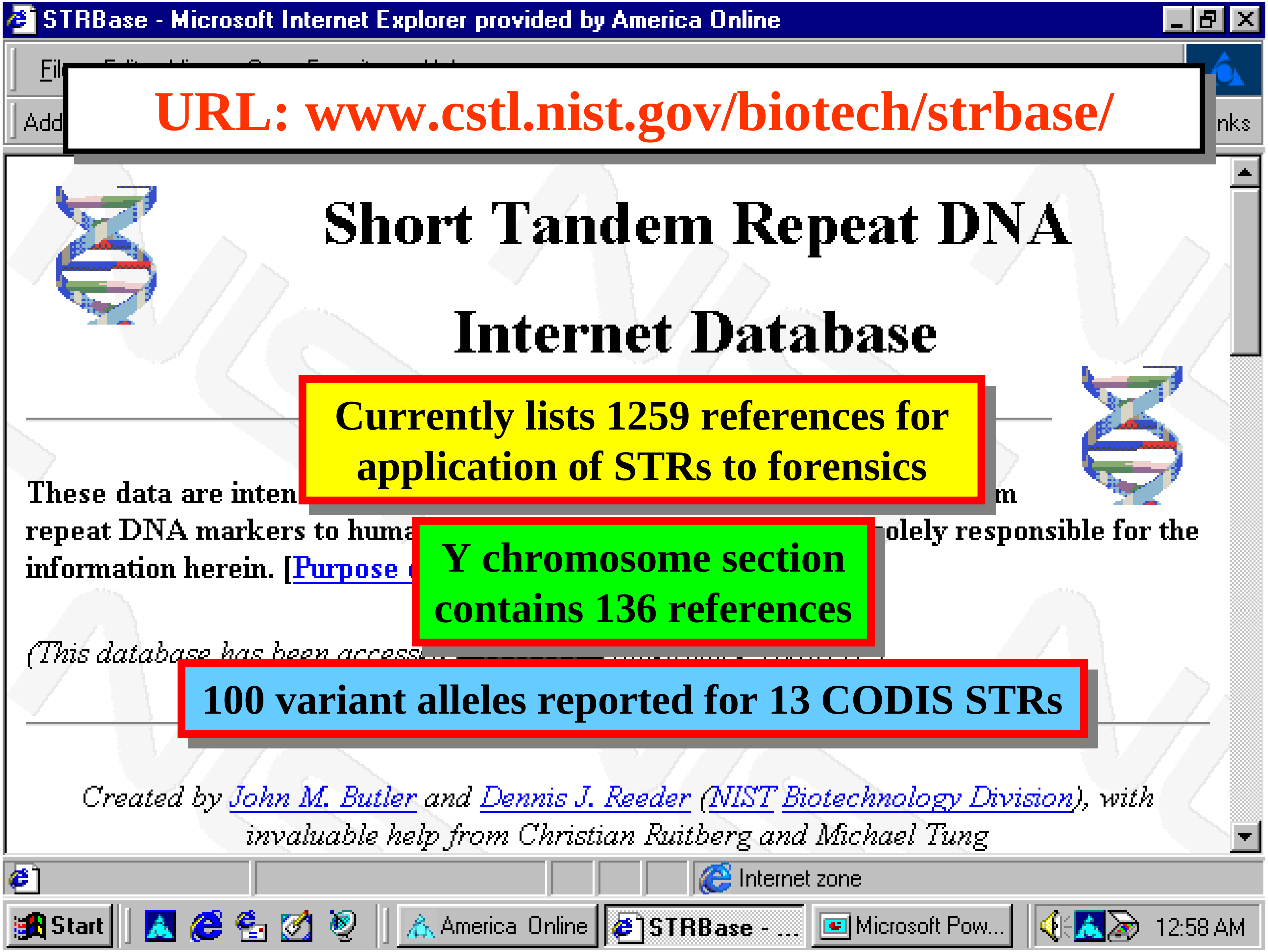
Less variation with Y STRs

For More Information...

- Peter de Knijff’s Y STR web page:
  - http://ruly70.medfac.leidenuniv.nl/~fldo/hptekst.html
- Y STR Haplotype database:
  - **http://ystr.charite.de**
- STRBase
  - **http://www.cstl.nist.gov/biotech/strbase**



URL: www.cstl.nist.gov/biotech/strbase/



Short Tandem Repeat DNA Internet Database

Currently lists 1259 references for application of STRs to forensics

These data are inten … m repeat DNA markers to huma … olely responsible for the information herein. [Purpose …

Y chromosome section contains 136 references

(This database has been access…

100 variant alleles reported for 13 CODIS STRs

Created by John M. Butler and Dennis J. Reeder (NIST Biotechnology Division), with invaluable help from Christian Ruitberg and Michael Tung

Acknowledgments

Research partially funded by NIJ Grant 97-VX-LB-0003

- Christian Ruitberg
- Peter Vallone
- Dennis Reeder
- Margaret Kline
- Lisa Forman (NIJ)

- Peter Underhill (Stanford Univ.)