

NIST
National
Institute
Standards
and
Technology

... working with industry to develop and apply technology, measurements and standards

Recent Developments in Y-Chromosome Analysis of Y-STRs and Y-SNPs

AAFS Y Chromosome Workshop
February 12, 2002
John M. Butler
National Institute of Standards and Technology

There is a growing interest in the Y-chromosome to aid forensic, paternity, and missing persons testing...

Human Genome

We need:

- Markers to type
- Population data
- Multiplex assays
- Standards

STR Markers

DYS19
DYS389/II
DYS390
DYS391
DYS392
DYS393
DYS385
YCAII
DYS437
DYS438
DYS439

Outline of Presentation

- Standardizing Information on Y Chromosome Markers
 - STRBase and Other Internet Resources
- Development of New Y STR Multiplexes
 - Primer Design Strategies
 - NIST Y STR 20plex Results
- Y SNP Typing Technologies
 - Y SNP Multiplex Assays
- NIST SRM 2395 with Y STR and Y SNP Data

Y STRs

Y Chromosome
Short Tandem Repeats

History of Y STR Marker Discovery

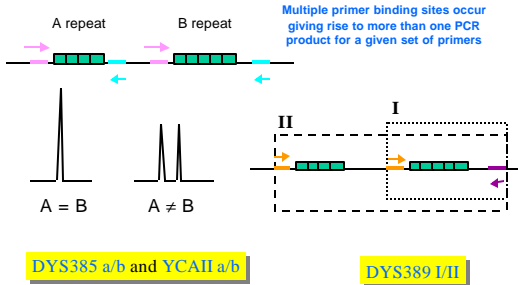
- **1992** - DYS19 (Roewer et al.)
- **1994** - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)
- **1996** - DYS389I/II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)
- **1996** - DYF371, DYS425, DYS426 (Jobling et al.)
- **1997** - DYS288, DYS388 (Kayser et al.)
- **1998** - DYS385 a/b (Schneider et al.)
- **1999** - A7.1, A7.2, A10, C4, H4 (White et al.)
- **2000** - DYS434, DYS435, DYS436, DYS437, DYS438, DYS439 (Ayub et al.)
- **2000** - G09411, G10123 (de Knijff unpublished)
- **2001** - DYS441, DYS442 (Iida et al.)
- **2002** - DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, DYS463, DYS464 (Redd et al.)

History of Y STR Marker Discovery

- **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.)
- **1999** - A7.1, A7.2, A10, C4, H4 (White et al.)
- **2000** - DYS434, DYS435, DYS436, DYS437, DYS438, DYS439 (Ayub et al.)
- **2000** - G09411, G10123 (de Knijff unpublished)
- **2001** - DYS441, DYS442 (Iida et al.)
- **2002** - DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, DYS463, DYS464 (Redd et al.)

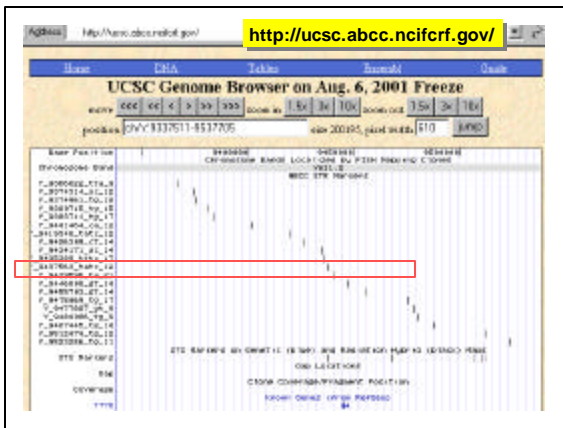
Blue labeled markers are in Y STR 20plex assay

Y STR Typing of Duplicated Regions

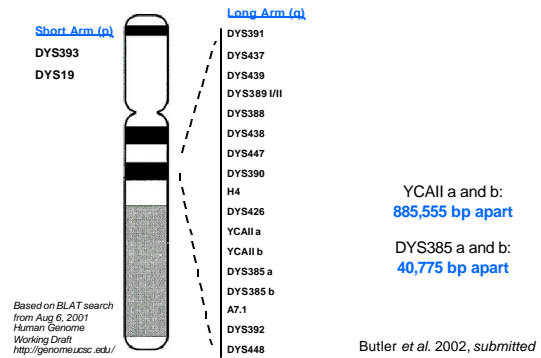


Human Genome Data

- Human Genome Project has enlarged knowledge of Y chromosome...
- Public human genome sequence available at <http://genome.ucsc.edu> ("Golden Path")
- Entire Y chromosome sequence can be downloaded -- **51 Mbytes (last half is all "Ns" from heterochromatin region)**
- ABCC STR sets – total number of Y STRs that might be polymorphic
- We can position all Y STR markers on current Y chromosome sequence



Chromosomal Positions of Y STRs



Internet Resources for Y Chromosome Information

<http://www.ystr.org/usa/> <http://www.ystr.org/europe/> <http://www.ystr.org/asia/>



<http://www.medfac.leidenuniv.nl/fldo/> <http://ycc.biosci.arizona.edu/>



<http://www.cstl.nist.gov/biotech/strbase/>



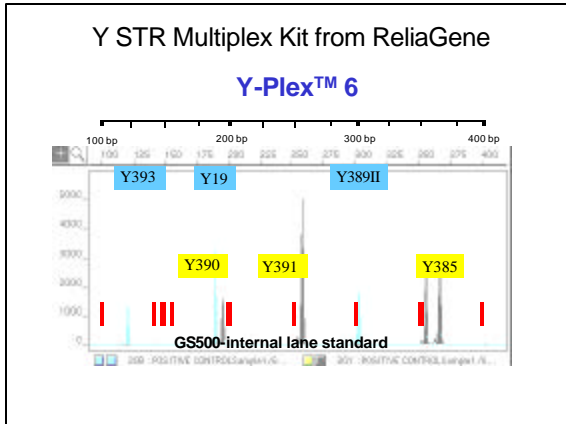
STRBase Short Tandem Repeat DNA Internet Database



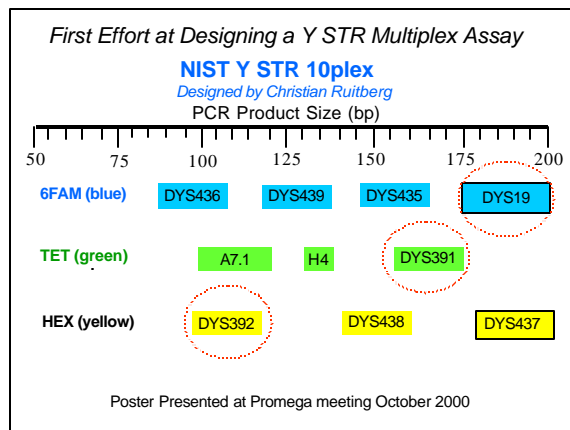
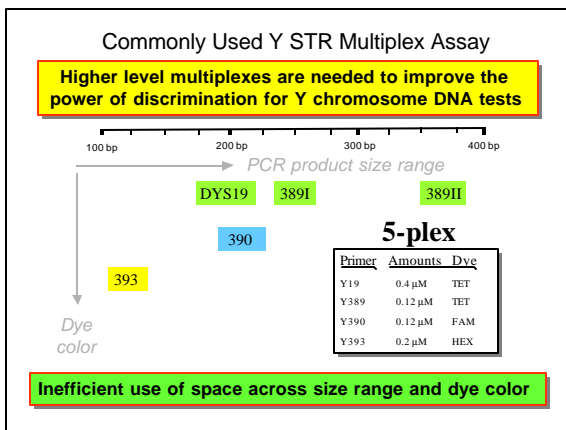
<http://www.cstl.nist.gov/biotech/strbase>

Y-Chromosome STR Information Available

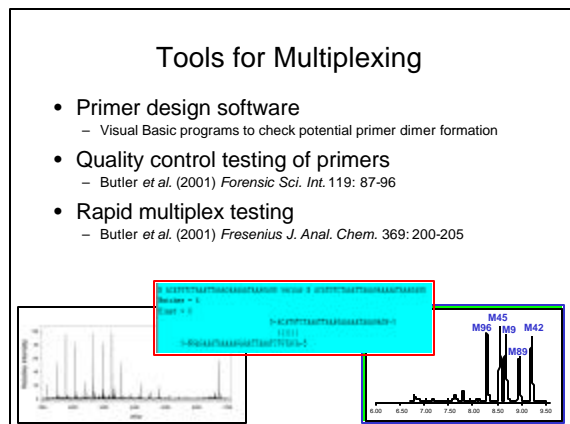
- ❖ Over 200 publications on Y STRs & SNPs cataloged
- ❖ Allele information on 16 Y STR loci
- ❖ Downloadable PowerPoint on Y STRs and Y SNPs
- ❖ Links to other Y-chromosome sites
- ❖ Information on new Y STR multiplexes developed at NIST



- ### Standard Way to Type "Extended Haplotype"
- DYS19, DYS388I, DYS388II, DYS385, DYS391, DYS392, DYS393, DYS394, YCAI
- GDB primers (original primers; not for multiplexing)
 - 2-3 multiplexes for minimal haplotype
 - DYS385 run separately sometimes
 - YCAI run separately to obtain extended haplotype
 - Different PCR conditions
 - Primers often require titrations (10 fold difference in amounts) in order to establish reasonable balance between Y STR PCR products
- Published primers were not designed to work together (originally used in singleplex PCRs)



- ### Multiplex PCR Requires Combination of Many Primers
-
- 22 different primers are combined for an 11-plex
- HPLC may be needed for improving purity following synthesis especially with dye labeled primers
- Quality control testing of oligonucleotides used in multiplex PCR reactions is essential
 - Reverse-phase HPLC
 - MALDI-TOF mass spectrometry
 - Y STR 20plex contains 34 different primers
- Butler et al. (2001) Forensic Sci. Int. 119: 87-96



Assay Development Goals

Discussion with Mecki Prinz December 2000

- Single tube amplification of all loci in minimal (9-loci) or extended (11-loci) European Y STR haplotypes
- Incorporation of additional polymorphic markers recently published
- No female amplification products (avoid X homology)
- Spacing between loci in same color to allow additional undiscovered alleles to be accommodated
- Similar concentration of primers to produce balanced amplification products
- Sensitivity to <500 pg male DNA with 28 cycle PCR (similar behavior to current STR kits)

The first demonstration of our Y STR 20plex was on July 5, 2001

Summary of Y STR Multiplex Assays

#	Locus	Ref. Allele	NIST 20plex	NIST 10plex	Y-PLEX 6	Gusmao 5plex
1	DYS19	15	253 bp	196 bp	193 bp	196 bp
2	DYS389I	12	155 bp			249 bp
3	DYS389II	29	275 bp		302 bp	369 bp
4	DYS390	24	217 bp		195 bp	216 bp
5	DYS391	11	109 bp	172 bp	253 bp	
6	DYS392	13	305 bp	114 bp		
7	DYS393	12	121 bp		116 bp	120 bp
8,9	DYS385	11	258 bp		356 bp	
10,11	YCAII	23	159 bp			
12	DYS388	12	157 bp			
13	DYS426	12	98 bp			
14	DYS437	16	194 bp	193 bp		
15	DYS438	10	320 bp	154 bp		
16	DYS439	20	226 bp	135 bp		
17	A7.1	10	113 bp	112 bp		
18	H4	12	138 bp	137 bp		
19	DYS447	23	212 bp			
20	DYS448	23	311 bp			
21	DYS435	9		143 bp		
22	DYS436	12		92 bp		

New Y STRs

New Y STRs	Core Repeat	Alleles
DYS446	CTCTT	9
DYS447	TAATA	8
DYS448	AGAGAT	7
DYS449	TTTC	10
DYS450	TTTTA	4
DYS452	TATAC	7
DYS453	AAAT	5
DYS454	AATA	3
DYS455	AATA	5
DYS456	GATA	6
DYS458	GAAA	8
DYS459	TAAA	9
DYS463	AARGG	9
DYS464	CCTT	41

Discovered by Mike Hammer's Group at the University of Arizona

14 Y STRs
8 tetras
5 pentas
1 hexa

Will be included in future multiplex assays

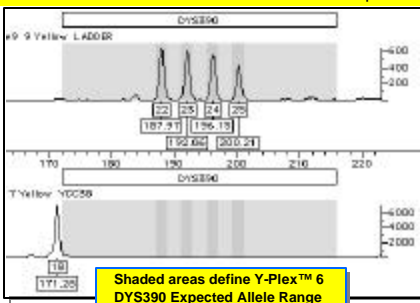
Redd et al. (submitted)

Development Strategy for Y STR Multiplexes

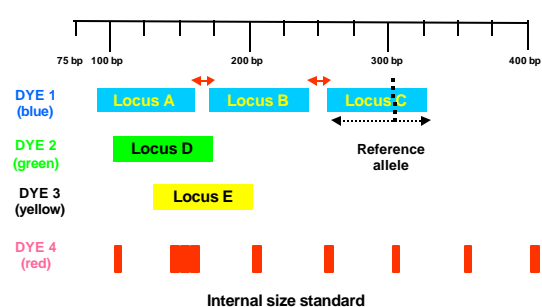
- Careful definition of allele ranges
 - Literature searches of over 200 papers to locate all known alleles
 - Evaluation of diverse population samples to search for rare alleles (M. Hammer cell lines)
 - Permits markers to be packed together more closely in a single dye color
- Efforts to avoid polymorphic nucleotides in primer binding sites
 - Alignment of multiple GenBank entries
 - Permits primers to be designed with less risk of null alleles
- Incorporation of newly developed NIST multiplex assay design and testing tools

Need to Define Allele Range Well

An extreme "off-ladder" allele seen in YCC panel



Multiplex Design Schematic



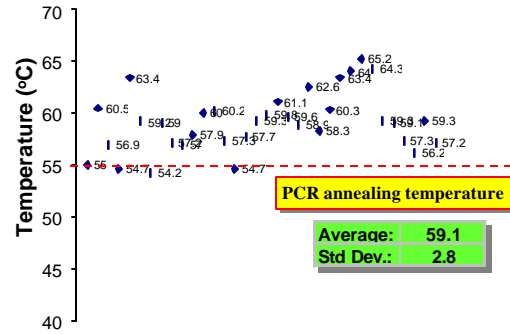
Intelligent Primer Design

- Overall balance in calculated primer T_m
- Check for primer-primer interaction

Examples

- YCAII
 - improved primer T_m compared to commonly used set
- DYS391
 - efforts to avoid X homology
- DYS19
 - moving primers in an effort to improve efficiency resulted in discovery of duplicated region of the Y

Y STR 20plex Primers (Predicted T_m)



YCAII: Improving Primer T_m

Mathias et al. (1994) Hum. Mol. Genet. 3: 115-123

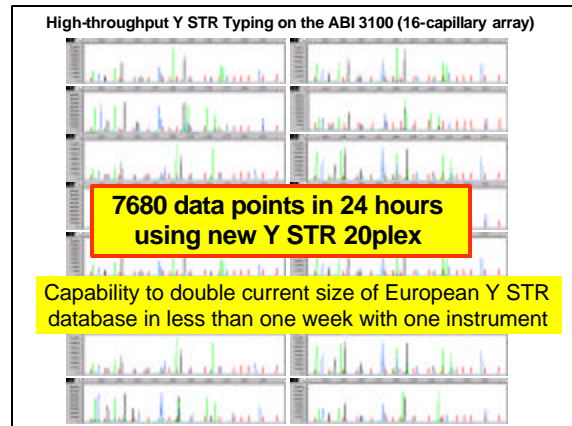
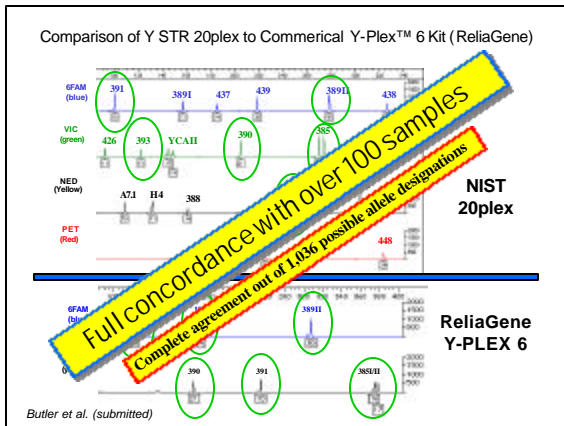
Butler et al. (submitted)

DYS391: Avoiding the X Homolog

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

Butler et al. (submitted)

Comparison of Various Primers on DYS19 and a Newly Discovered Duplicated Region of the Y Chromosome

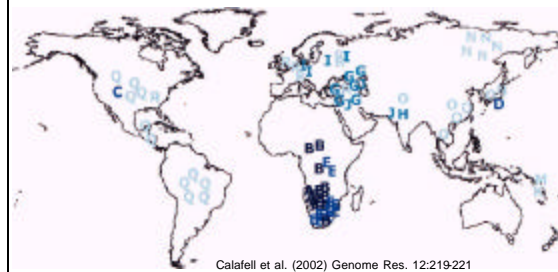


Advantages of Y STR 20plex

- Single amplification of “extended haplotype”
 - European loci: 19, 385 I/II, 389I, 389II, 390, 391, 392, 393, YCAII a/b
 - Additional loci: 426, A7.1, H4, 388, 437, 438, 439, 447, 448
- Sensitive to <250 pg with 28 cycle PCR
- Male-specific with >100X female DNA
- 10 loci with amplicons less than 200 bp in size to aid results with degraded DNA

Y Chromosome Consortium (YCC) Cell Line Samples

74 males and 2 females from diverse world populations

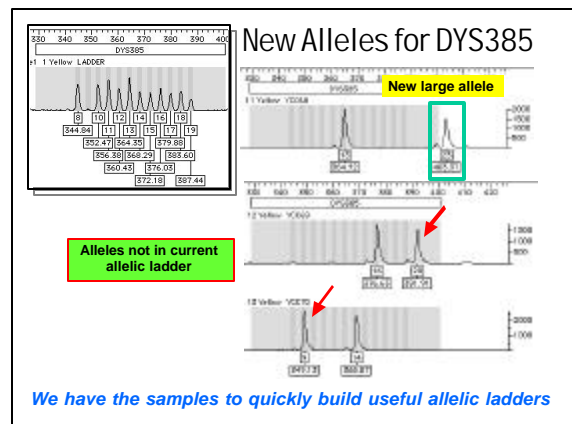


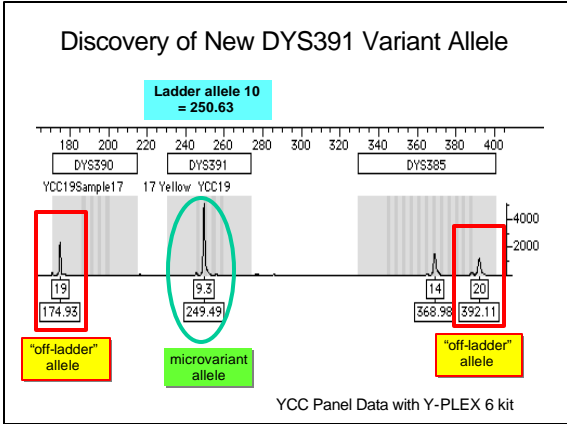
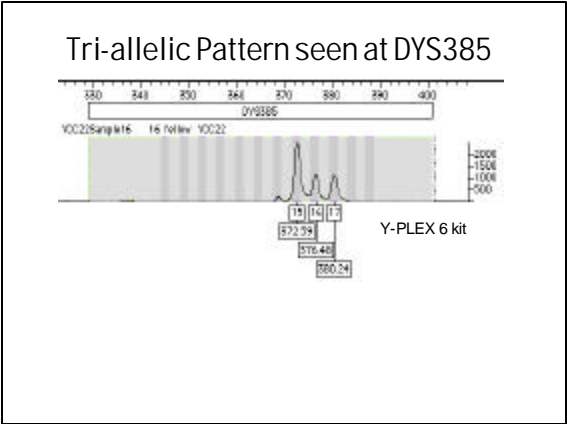
Source of YCC Samples

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 North Asians |
| • 5 Pygmy | • 1 Cambodian |
| • 10 Bantu speakers | • 2 Melanesians |
| • 14 Europeans | • 12 Native Americans |
| • 5 Middle Easterns | • 74 male cell lines |
| • 3 Pakistanis | • 2 females (YCC1 and 54) |
| • 3 Chinese | • established in 1991 by Mike Hammer and Nathan Ellis |
| • 3 Japanese | |

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





- ### Plans with our Y STR Megaplex
- Currently being tested in 2 other labs
 - Mecki Prinz -- forensic samples with 4 dye 16plex
 - Mike Hammer -- population studies with 5 dye 20plex
 - Concordance studies with YCC sample panel
 - Allele sequencing...developmental validation
 - Determine Y STR haplotypes for Y SRM 2395
 - **First publication submitted with details on primers**
 - Construct Allelic Ladders/LSBs
 - We are committed to seeing this assay commercialized into a kit so that it can be widely available
- We plan to develop additional Y STR multiplexes using newly discovered markers**

- ### Allelic Ladders vs. LSBs
- **Locus-Specific Brackets (LSBs)** are artificial alleles that serve as internal sizing standards, one smaller and one larger in size than the common set of STR alleles
 - The flanking region of the LSB is identical to the STR alleles giving them the same electrophoretic properties
 - LSBs are used instead of allelic ladders to type samples
 - Collaboration has been established with OligoTrail to produce LSBs for Y STR multiplexes (NIJ-funded project)
-

Y SNPs

Single Nucleotide Polymorphisms and Other Biallelic Markers

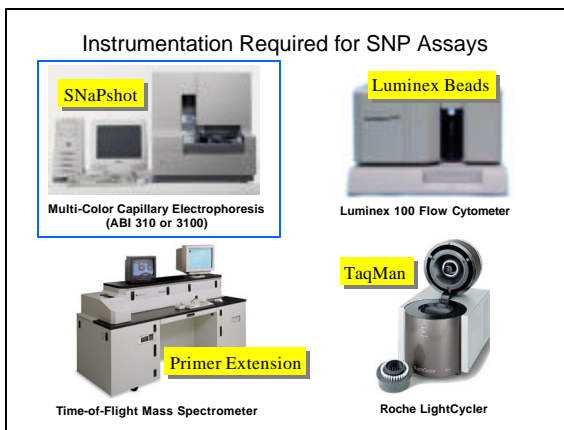
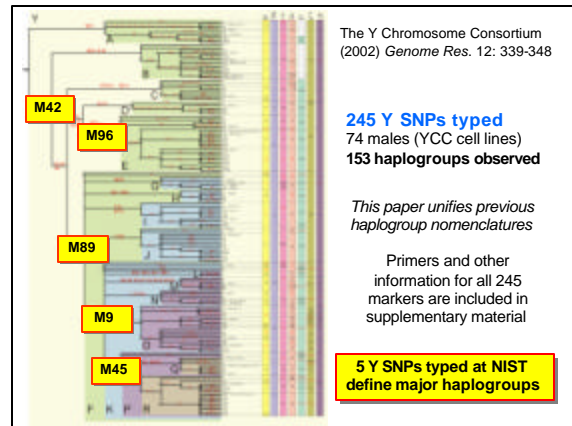
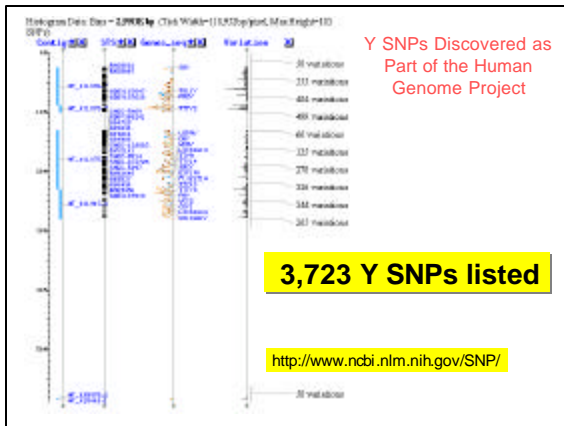
Y Bi-allelic Markers

Marker Name	Polymorphism	PCR Product Size (bp)	Restriction Enzyme (PCR-RFLP assay)	Reference
YAP (DYS287)	Alu insertion (305 bp)	150 (YAP-) / 455 (YAP+)		Hammer 1994
SRY-8299	G→A	509	<i>Bvu</i> BI	Whitfield 1995
SRY-1532	G→A	167	<i>Dra</i> III	Kwok 1996
SRY-2627	C→T	391	<i>Ban</i> I	Ventia 1997
sY81 (DYS271)	A→G	209	<i>Hsp</i> 92II	Seielstad 1994

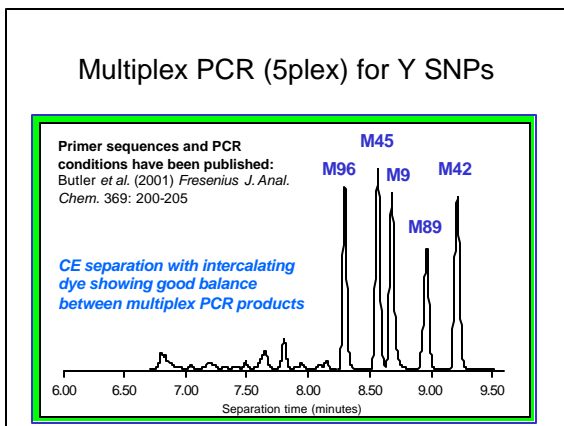
Until 1997, only 11 Y biallelic markers were known...

>200 Y SNPs discovered by Peter Underhill (Stanford) using DHPLC

Underhill *et al.* (2000) *Nature Genetics* 26: 358-361

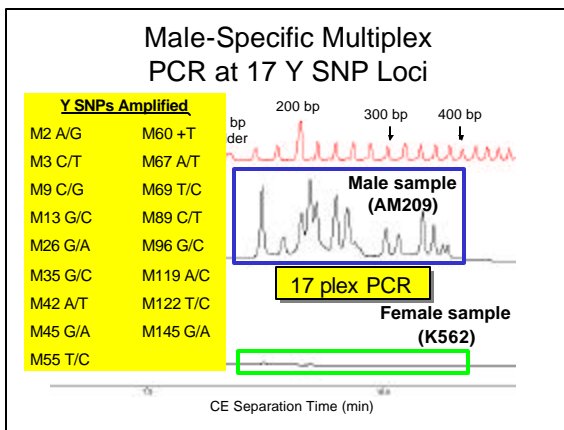
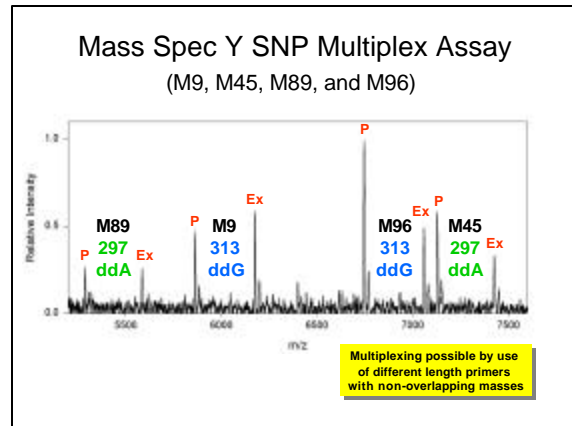
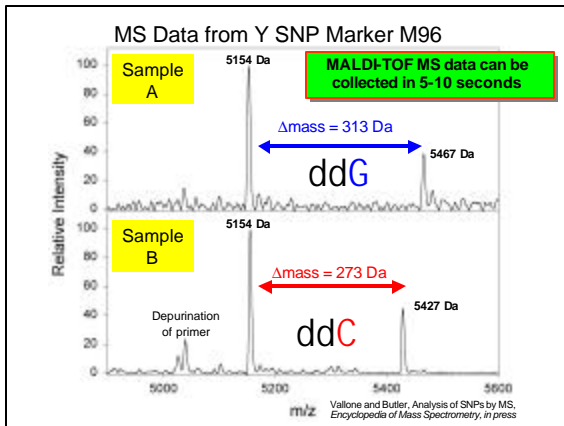


- ### Approach to Y SNP Assays
- Use branch-defining Y SNPs
 - Develop multiplex PCR with optimized markers
 - M96, M42, M45, M9, M89
 - Recent *Genome Research* paper spells out best markers against the YCC panel
 - Develop 10-20plex Y SNP panel with SNaPshot and mass spec approaches
 - 9plex PCR and 10plex SNaPshot assay already demonstrated with coding region mtDNA markers
 - 10plex mass spec assay from control region mtDNA following amplification of entire D-loop



Y SNP Results for 16 Standard Samples

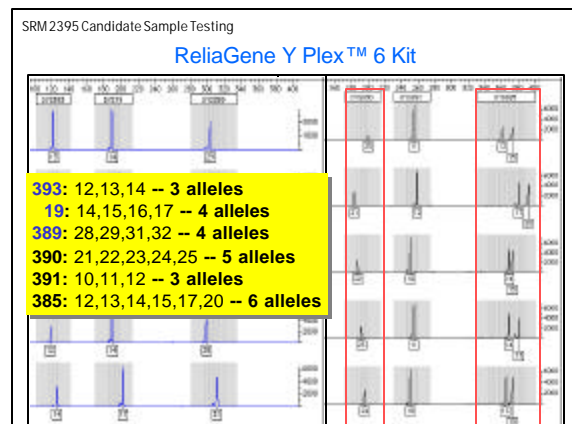
SampleID	M9 (C/G)	M42 (A/T)	M45 (G/A)	M89 (C/T)	M96 (G/C)
Male 1	G	T	A	T	C
Male 2	G	T	A	T	C
Male 3	G	T	A	T	C
Male 4	G	T	A	T	C
Male 5	G	T	A	T	C
Male 6	G	T	A	T	C
Male 7	G	T	A	T	C
Male 8	C	T	G	C	G
Male 9	C	T	G	C	G
Male 10	C	T	G	C	G
Male 11	C	T	G	C	G
Male 12	C	T	G	C	G
Male 13	C	T	G	C	G
Male 14	G	T	A	T	C
Male 15	C	T	G	C	G
Female	-	-	-	-	-

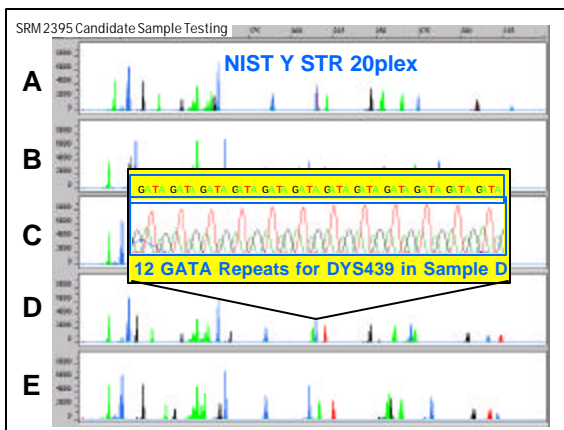
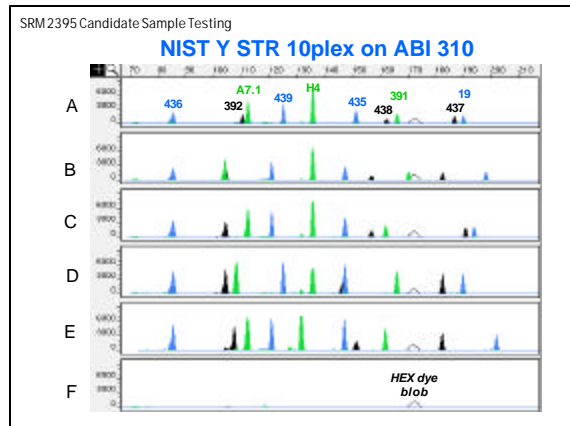
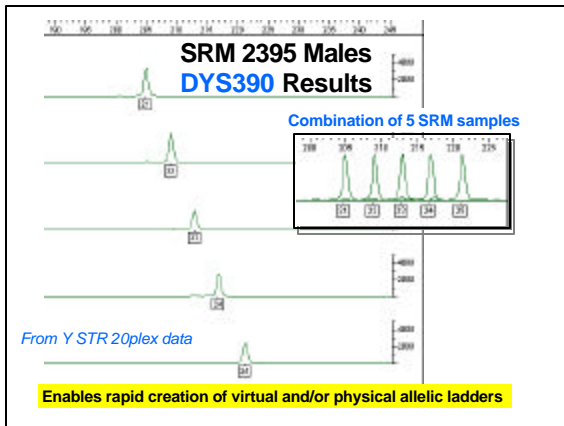


SRM 2395

NIST Standard Reference Material for Human Y Chromosome Markers

- SRM 2395: A Y Chromosome Standard Reference Material
- Evaluation of 144 different male DNA samples
 - 66 genomic DNAs (anonymous from blood bank)
 - 78 cell lines
 - Selection of **6 candidate samples**
 - milligram quantities of genomic DNA
 - 5 males (labeled A-E)
 - 1 female (labeled F)
 - Plan to supply **100 ng** of each DNA sample
 - Chosen to exhibit wide range of alleles for all common Y STRs and Y SNPs
 - Samples will be sequenced at 20 Y STRs and ~10 Y SNPs
 - Standards will be useful for any primer sets





SRM 2395 Candidate Sample Testing

Y SNP Results with Primer Extension and MALDI-TOF MS

	M9(C/G)	M42(A/T)	M45(G/A)	M89(C/T)	M96(G/C)
A	G	T	A	T	C
B	C	T	G	T	C
C	C	T	G	T	C
D	C	T	G	C	G
E	C	T	G	T	C
F	--	--	--	--	--

Summary of Y Chromosome Work at NIST

- We are standardizing information on Y chromosome markers and making it available through [STRBase](#)
- We are developing new Y STR multiplex assays ([Y STR 20plex](#) and [Y STR 10plex](#))
- We are evaluating SNP typing methodologies and developing [Y SNP assays](#) involving primer extension
- We are developing a Y Chromosome Standard Reference Material ([SRM 2395](#))

Acknowledgments

Funding:
 National Institute of Justice Grant #97-LB-VX-0003
 Interagency Agreement between NIJ and NIST Office of Law Enforcement Standards

NIST Personnel:
 John Butler (Project Leader)
 Pete Vallone
 Margaret Kline
 Jan Redman
 Rich Schoske (AU)
 Gordon Spangler (AU)
 Christian Ruitberg (RPI)
 Dave Duewer (Anal. Chem.)

Collaborators:
 Mike Hammer and Alan Redd (U. Arizona) for Y STR sequences and samples
 Mecki Prinz (NYC OCME)
 Dave Carlson (Marligen) on Y SNP work with Luminex beads

Human Genome Project participants (Jim Kent-UCSC)