


Going Beyond the U.S. Haplotype: A Look at Additional Y-STR and Y-SNP Loci in U.S. Populations

John Butler

AAFS Workshop:
Y-STR Analysis on Forensic Casework
February 17, 2004



History of Y STR Marker Discovery

1992 - **DYS19** (Roewer et al.) "Extended Haplotype"

1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)

1996 - **DYS389II**, **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2000 - G09411 (DYS462), G10123 (de Knijff unpublished) "U.S. Haplotype"

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)

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

Commercial Y-STR Kits Available

- ReliaGene Technologies (New Orleans, LA)
 - Y-PLEX™ 6**: DYS19, DYS389II, DYS390, DYS391, DYS393, DYS385 a/b
 - Y-PLEX™ 5**: DYS389II, DYS392, DYS438, DYS439
 - Y-PLEX™ 12**: DYS19, DYS385 a/b, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, amelogenin
- Promega Corporation (Madison, WI)
 - PowerPlex® Y**: DYS19, DYS385 a/b, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437
- Serac (Germany)
 - genRES® DYSpLex-1**: DYS19, DYS389II, DYS390, DYS391, DYS385 a/b, amelogenin
 - genRES® DYSpLex-2**: DYS19, DYS389II, DYS392, DYS393
- GKT Inc. (South Korea); *silver-stain kits*
 - GeneKin® Y-STR Systems I**: DYS388, DYS19, DYS392
 - GeneKin® Y-STR Systems II**: DYS393, DYS390, DYS391, DXY391X
 - GeneKin® Y-STR Systems III**: DXYS156X, DXYS156Y, DYS389II
 - GeneKin® Y-STR Systems IV**: DXYS156X, DXYS156Y, DYS385 a/b

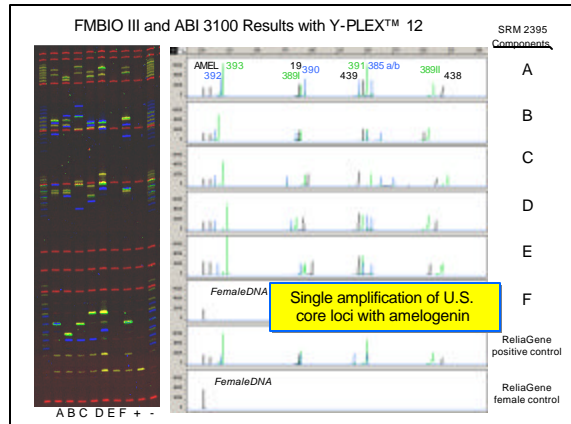
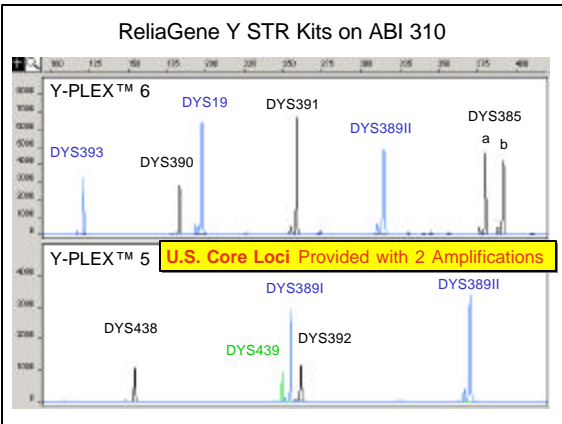
Commercial Y STR Kits

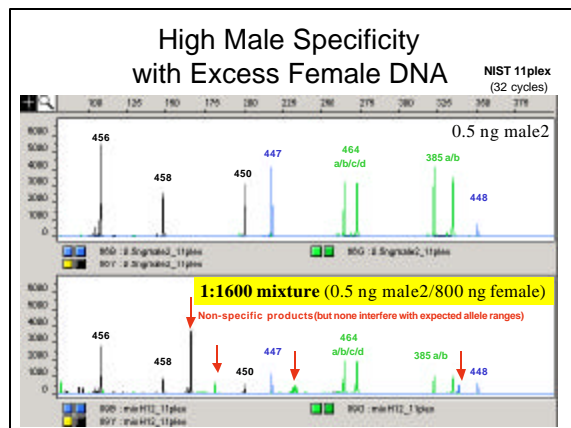
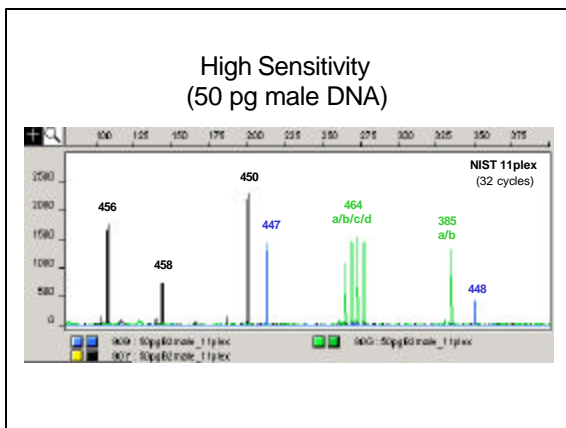
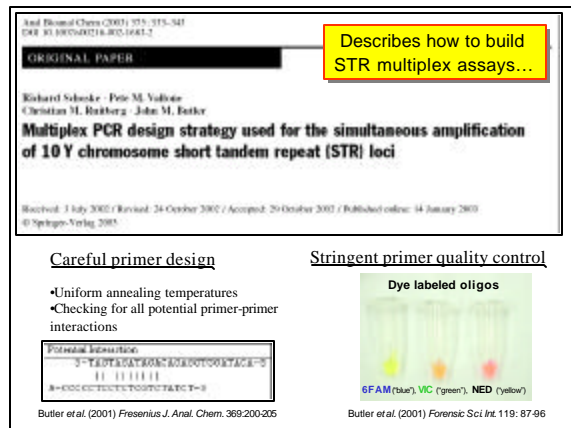
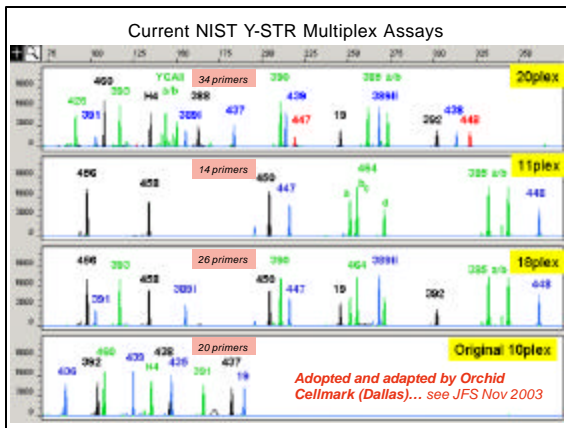
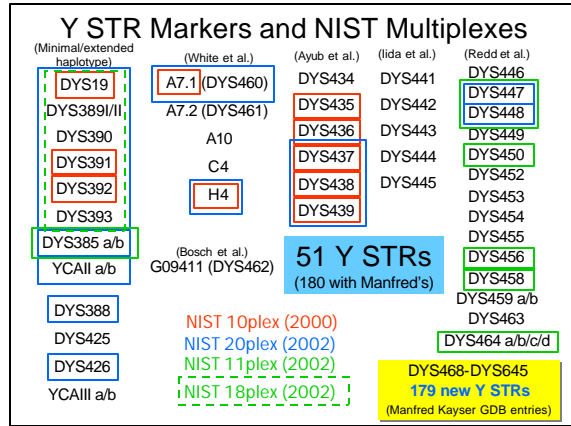
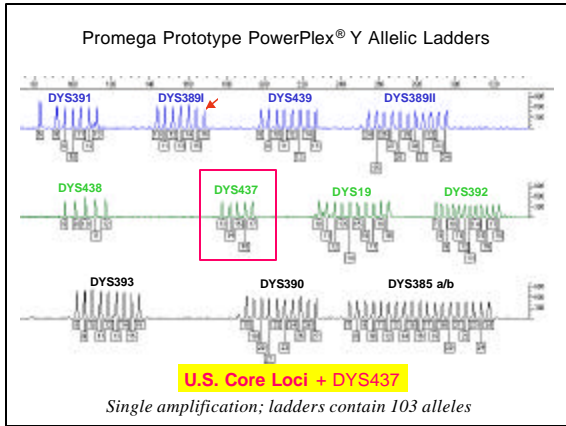
(Minimal/extended haplotype)	(White et al.)	(Ayub et al.)	(Iida et al.)	(Redd et al.)
DYS19	A7.1 (DYS460)	DYS434	DYS441	DYS446
DYS389II	A7.2 (DYS461)	DYS435	DYS442	DYS447
DYS390	A10	DYS436	DYS443	DYS448
DYS391	C4	DYS437	DYS444	DYS449
DYS392	H4	DYS438	DYS445	DYS450
DYS393		DYS439		DYS452
DYS385 a/b				DYS453
YCAII a/b	(Bosch et al.) G09411 (DYS462)			DYS454
				DYS455
				DYS456
				DYS458
				DYS459 a/b
				DYS463
				DYS464 a/b/c/d

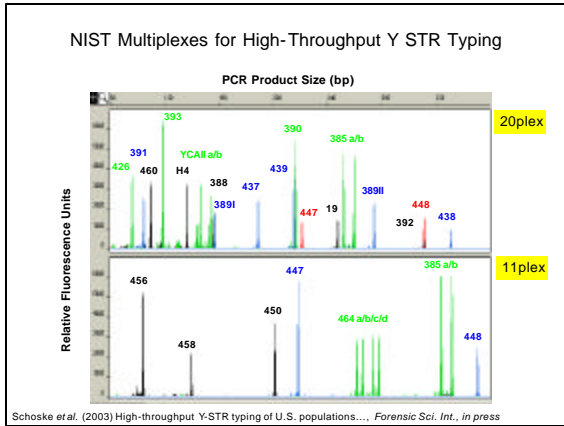
51 Y STRs
(180 with Manfred's)

Y-PLEX 6 (ReliaGene)
Y-PLEX 5 (ReliaGene)
Y-PLEX 12 (ReliaGene)
PowerPlex Y (Promega)

DYS468-DYS645
179 new Y STRs
(Manfred Kayser GDB entries)







Describes multiplex PCR primer sequences for NIST 11plex

ARTICLE IN PRESS

ELSEVIER

Forensic Science International

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

Richard Schoske^{a,*}, Peter M. Vallone^c, Margaret C. Kline^a, Janette W. Rodman^a, John M. Butler^b

^aTechnology Division, National Institute of Technology, 120 Pennington Drive, Mail Stop 5311, Gaithersburg, MD 20899, USA
^bDepartment of Chemistry, American University, Washington, DC 20054, USA
^cReceived 28 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 U.S. Population Samples

As of 06/2003 666 males (anonymous; self-identified ethnicities)

286 Caucasians
252 African Americans
128 Hispanics

Whole blood received from Interstate Blood Bank (Memphis, TN)

Working tubes/plates 1 ng/uL

To date: (~50,000 allele calls)
Identifier (15 autosomal markers + Amelogenin) (10,608)
Roche Linear Arrays (HV1/HV2 10 regions) (6,630)
Y STRs 22 loci—27 amplicons (17,388)
Y SNPs 50 markers on sub-set of samples (11,498)

On average ~80 ng total extracted genomic DNA

Stock tubes

Working tubes

Working plates

Samples supplied to OhioU for miniSTR typing and AFDIL for whole mtGenome sequencing

Typing Beyond the U.S. Core Y-STR Loci

- U.S. Core Y-STRs (selected Jan 2003 by SWGDAM Y-chromosome subcommittee): DYS19, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS385 a/b
- First examination of 11 U.S. haplotype loci vs. new ones: 647 U.S. males:
 - 260 African Americans (including 20 from Carl Ladd)
 - 244 Caucasians (including 20 from Carl Ladd)
 - 143 Hispanics (including 20 from John Hartmann)
- Types generated at NIST for 22 Y-STRs and 50 Y-SNPs
- Performance of U.S. haplotype vs. European "extended" haplotype (DYS438 and DYS439 vs. stutter-prone dinucleotide YCAII a/b)
- Resolution of most common types with additional markers

Schoske et al. (2003) High-throughput Y-STR typing of U.S. populations.... *Forensic Sci. Int., in press*

113 (179) different types in 679 males with DYS464

Y-STR	Population	African Americans	Caucasians	Hispanics
	STR diversity (N=647) Rank	STR diversity (N=260) Rank	STR diversity (N=244) Rank	STR diversity (N=143) Rank
DYS464 a/b/c/d	0.956 1	0.954 1	0.954 1	0.937 1
DYS385 a/b	0.912 2	0.942 3	0.838 2	0.903 2
YCAII a/b	0.790 3	0.797 3	0.701 5	0.773 4
DYS455	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	0.751 4	0.695 9
DYS433	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
DYS391	0.656 12	0.636 11	0.679 9	0.717 7
DYS437	0.637 13	0.699 17	0.582 15	0.624 14
H4	0.611 14	0.612 12	0.562 14	0.609 15
DYS392	0.609 15	0.494 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	0.467 19	0.552 16	0.577 17
DYS463	0.519 19	0.379 21	0.462 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
DYS385	0.365 22	0.546 22	0.501 18	0.312 22

Schoske et al. (2003) High-throughput Y-STR typing of U.S. populations.... *Forensic Sci. Int., in press*

Extended Haplotype vs. US Haplotype

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

Y-STR Marker Combination	260 African Americans	244 Caucasians	143 Hispanics			
	DC	UH	DC	UH	DC	UH
*extended haplotype	0.9488	0.0012	0.9971	0.0029	0.9975	0.0025
*U.S. haplotype	0.9293	0.0007	0.9574	0.0026	0.9986	0.0014

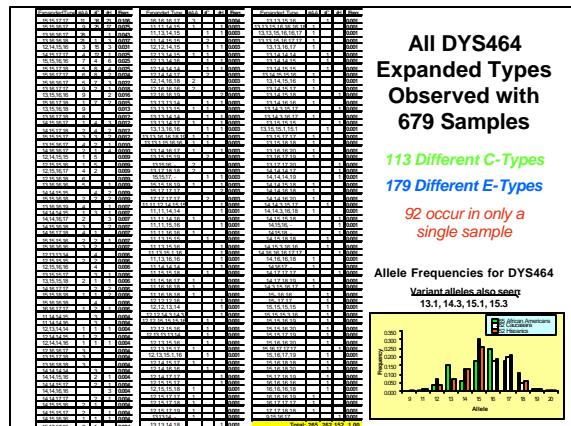
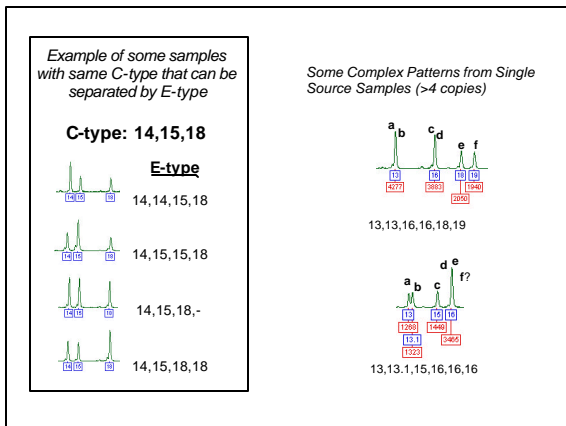
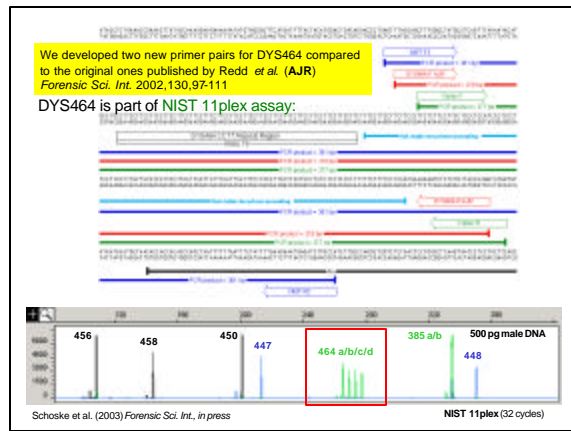
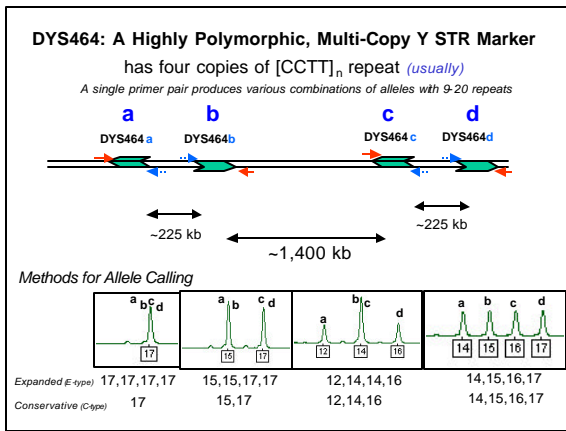
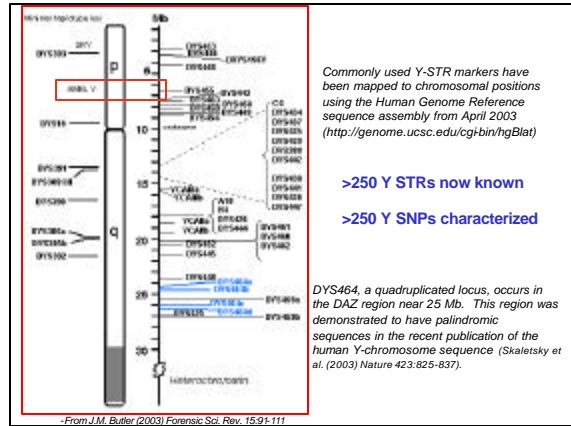
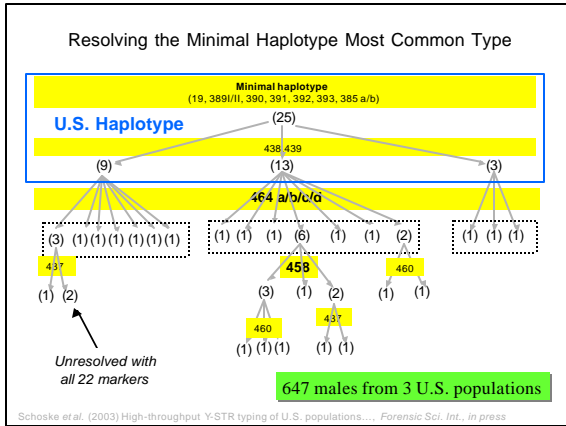
HD = haplotype diversity
RMP = random match probability (1-HD)
HD = (n-1)(1 - Σp_i²)

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

Y-STR Marker Combination	260 African Americans	244 Caucasians	143 Hispanics			
	DC	UH	DC	UH	DC	UH
*extended haplotype	91.9%	227	93.6%	394	93.9%	120
*U.S. haplotype	91.9%	222	93.3%	376	93.3%	121

DC = discriminative capacity (number of haplotypes/size of sample)
UH = unique haplotype (occurs only once in a given population)

Schoske et al. (2003) High-throughput Y-STR typing of U.S. populations.... *in press*



Poster at ISFG 2003
Progress in Forensic Genetics 10, in press

ARTICLE IN PRESS

Forensic value of the multicopy Y-STR marker DYS464[†]

John M. Butler^a, Richard Schokke^b

^aWest Virginia University, National Institute of Standards and Technology, Gaithersburg, MD 20899
^bReceived 1 September 2003; received in revised form 19 September 2003; accepted 11 October 2003

DYS464 Allelic Ladder

NIST allelic ladder for DYS464 produced from 8 different DNA samples. The variant alleles 14.3, 15.1, and 15.3 are helpful as a tool for measuring single base resolution in electrophoretic systems.

ELSEVIER

Forensic Science International 137 (2003) 231–236

Announcement of population data

Molecular characterization and Austrian Caucasian population data of the multi-copy Y-chromosomal STR DYS464

Burkhard Berger, Harald Niederstätter, Anita Brandstätter, Wálther Parson^a

^aInstitute of Legal Medicine, University of Innsbruck, Mikroskopische Anatomic, A-6020 Innsbruck, Austria
 Received 5 May 2003; received in revised form 20 July 2003; accepted 30 July 2003

They advocate using DYS464 as a screening tool

The addition of DYS464 to the *maHT* offers a potentially increased power of discrimination. In our population study the number of different haplotypes increased from 110 (*maHT*) to 122 (*maHT* + DYS464). But even when analysed alone, DYS464 allowed us to distinguish between 54 different genotypes. Therefore DYS464 is practical for high-throughput screenings and could be successfully applied for mass screenings, e.g. in connection with rape cases or mass disasters.

Y-STR Conclusions

- Additional markers beyond the European minimal haplotype and SWGDAM-recommended U.S. haplotype do help resolve common lineages
- The multi-copy marker DYS464 is extremely polymorphic and would be beneficial to use

Y-SNPs

SNP Typing Instrumentation at NIST

PCR & primer extension
 Multi-Color Capillary Electrophoresis (ABI 310 or 3100)

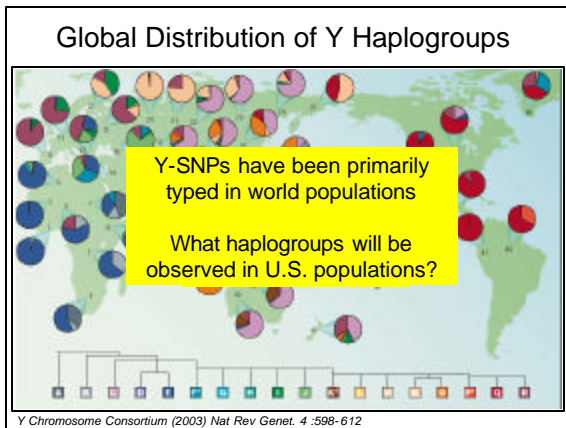
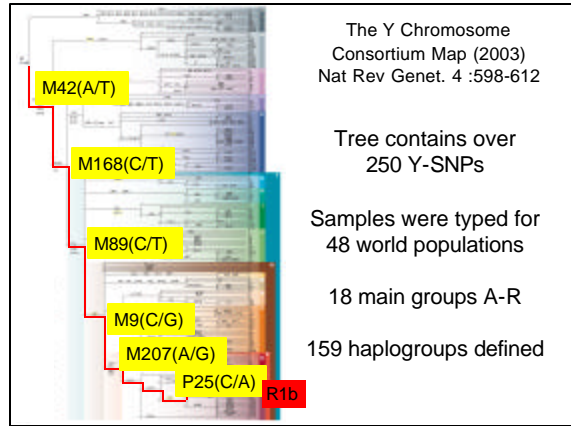
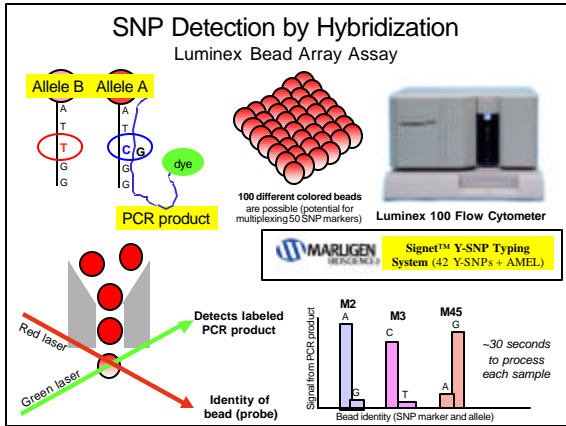
Luminex Beads hybridization
 Luminex 100 Flow Cytometer

Primer Extension
 Time-of-Flight Mass Spectrometer

TaqMan
 ABI 7000 SDS

Allele-Specific Primer Extension (ASPE)

Electrophoretic Run Time (sec)



Y-SNPs in U.S. populations

What haplogroups will be observed?

How specific will certain Y-SNPs be for a U.S. population group?

Forensic utility in comparison/addition to Y-STRs

Commercial kit (Marligen) 42 Y-SNPs

Medium sized multiplexes developed in-house (CE or MS)

Y-SNPs Typed at NIST

42 SNPs + Amelogenin present in 5 multiplexes (commercially available kit from Marligen)

18 SNPs in 3 NIST-designed 6plexes (8 unique)

10 SNPs in 2 NIST-designed 5plexes (1 unique)

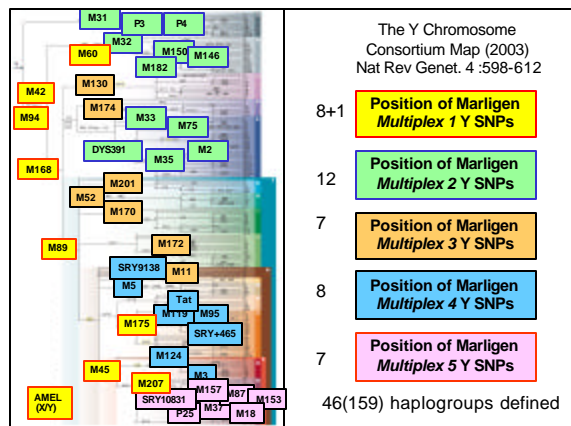
19 of the SNP sites overlapped...

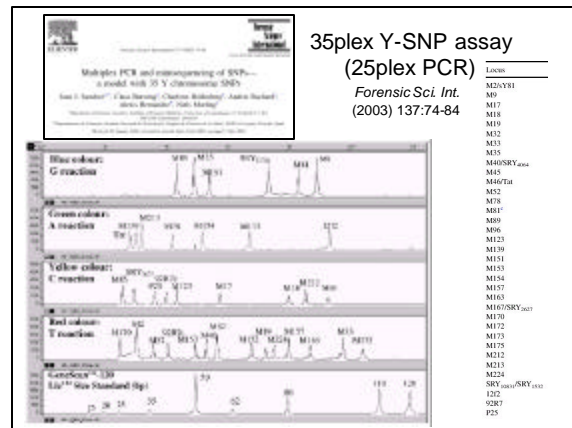
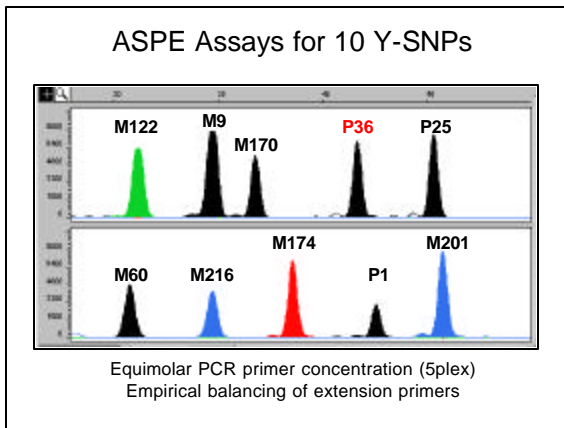
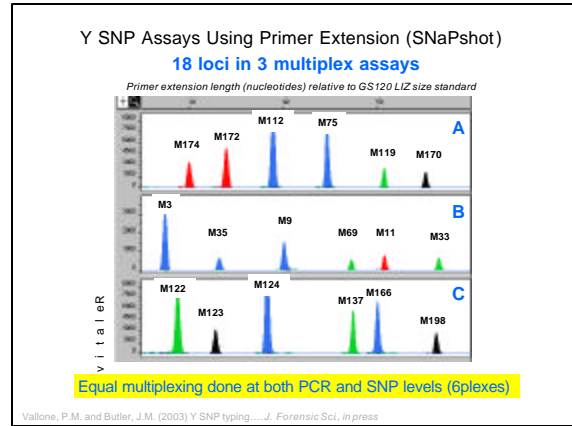
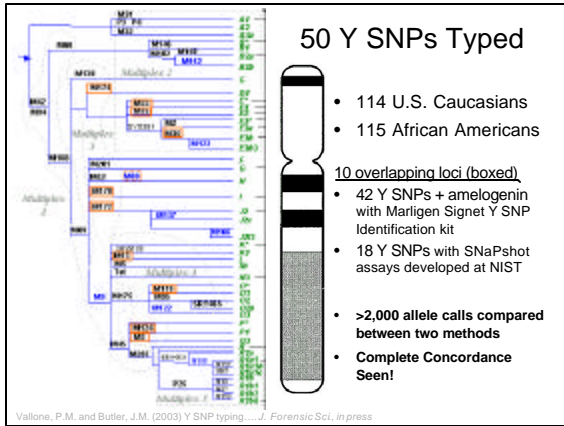
Resulting in a total of 51 Y-SNPs

115 African Americans

114 Caucasians

95 Hispanics (presently typed for 10 Y-SNPs)





Summary of Y-SNP Data

(115 African Americans and 114 Caucasians)

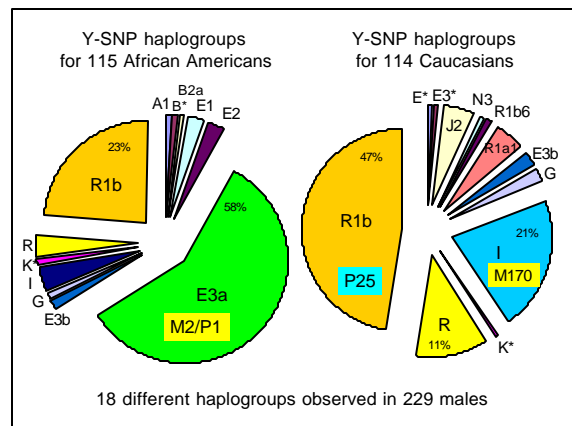
A total of 20 ng of genomic DNA was consumed for the 10 multiplexes

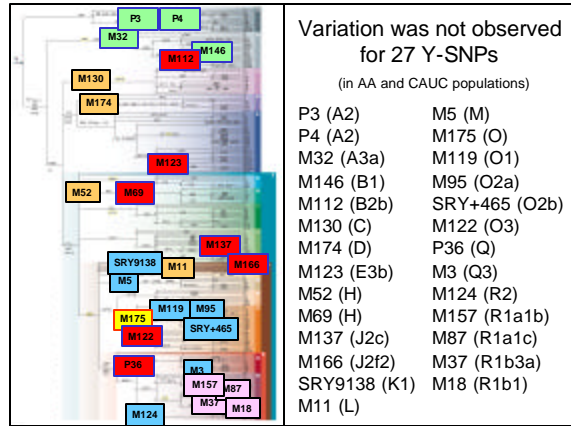
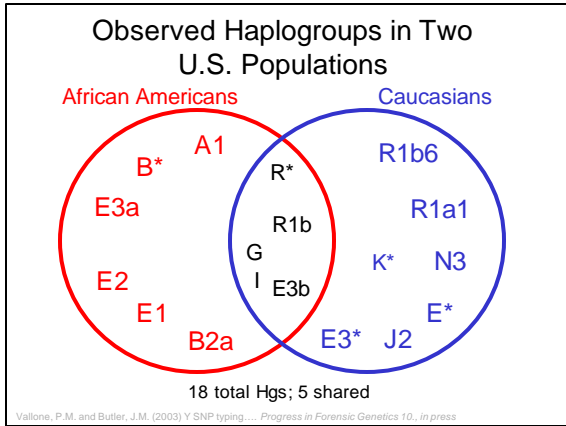
18 out of 46 haplogroups observed

Over 99 % success rate for allele calls (both methods)

Variation was observed in 24 of the 51 Y-SNPs

100% concordance for the 19 overlapping markers (>3,800 allele calls)

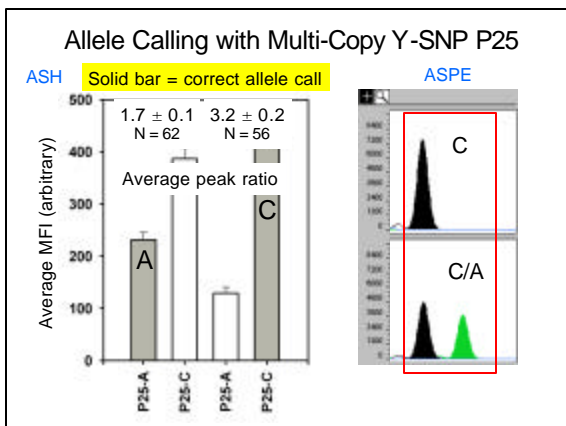
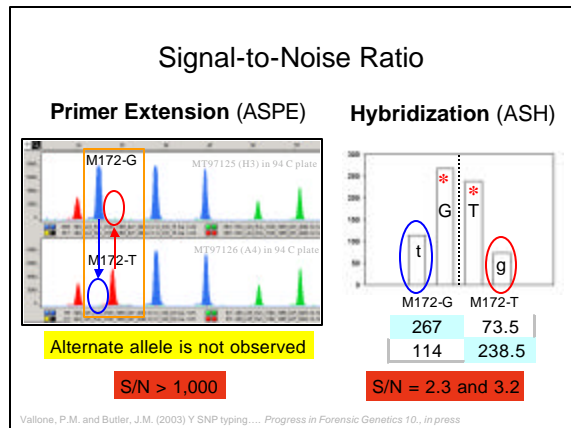




Derived in more than one population

Locus	All	AA	Cauc	Hisp	Hap
M2 A/G	0.23	0.58	not obs	0.08	E3a
DYS391 C/G	0.31	0.60	0.04	na	E3
M170 A/C	0.10	0.04	0.21	0.04	I
M35 G/C	0.02	0.02	0.03	na	E3b
M201 G/T	0.03	0.01	0.03	0.05	G
SRY10831 A/G	0.03	0.01	0.05	na	R1a

M2 is primarily derived in the African American pop



Forensic Utility

51 Y-SNPs versus 1 Y-STR

For N = 211 male samples

	51Y-SNPs	Y-STR DYS464
Amount of sample consumed	20ng	<1ng
Number for types observed	18	62
Analysis	Multiple	1 reaction
Degraded samples	+	?

Y-SNP Conclusions

- Full concordance was observed between hybridization and primer extension technologies on 18 different Y-SNPs (>3,800 allele calls)
- Caucasian admixture was observed with our African American population (Hg R and R1b in ~30%)—agrees with Kayser *et al.* (2003) *Genome Res.* 13:624-634 done with 9 Y-STRs
- Y-SNPs may have limited value for ethnic differentiation in U.S. populations
 - One exception: M2 not in Caucasians
- Y-SNPs are not a useful stand-alone assay for forensic purposes, but may be helpful in combination with Y-STRs

NIST SRM 2395 Human Y-Chromosome DNA Profiling Standard

NIST SRM 2395 Human Y Chromosome DNA Profiling Standard

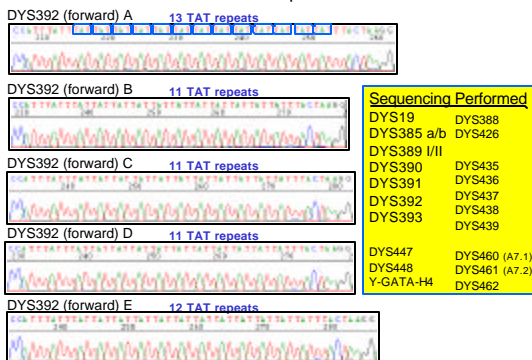
- **5 male samples** + 1 female sample (neg. control)
 - **100 ng** of each component (**50 µL** at ~2 ng/µL)
 - **22 Y STR markers sequenced** to provide certified values (number of repeats)
 - **9 additional Y STR markers typed**
 - **42 Y SNPs typed** with Marligen kit
- Now available from NIST Standard Reference Material office (www.nist.gov/srm)
 - Can be used to verify results with *any primer sets*
 - Will help U.S. labs meet DAB/FBI Standards

SRM 2395 Now Available...



<http://www.nist.gov/srm>

Sequence Determination of Y STR Repeat Region for Each Component




Y SNP Results on SRM 2395 from Marligen Signet™ Multiplexes (Luminex bead assay)

SRM 2395	AMEL	M207	M45	M89	DYS391	M2	M170	M172	M201
		(A/G)	(A/G)	(C/T)	(C/G)	(A/G)	(A/C)	(G/T)	(G/T)
Component A	XY	G	A	T	C	A	A	T	G
Component B	XY	A	G	T	C	A	A	G	G
Component C	XY	A	G	C	G	G	A	T	G
Component D	XY	A	G	T	C	A	A	T	T
Component E	XY	A	G	T	C	A	C	T	G
Component F	XX								

SRM components are all distinguishable from one another with these Y SNPs

42 Y SNPs measured across all samples




STRBase
Short Tandem Repeat DNA
Internet Database

working with industry to develop and supply us with technology, measurements and standards

<p>General Information</p> <ul style="list-style-type: none"> •Intro to STRs (downloadable PowerPoint) •STR Fact Sheets •Sequence Information •Multiplex STR Kits •Variant Allele Reports 	<p>Forensic Interest Data</p> <ul style="list-style-type: none"> •FBI CODIS Core Loci •DAB Standards •NIST SRM 2391 •Published PCR Primers •Y-Chromosome STRs •Population Data •Validation Studies 	<p>Supplemental Info</p> <ul style="list-style-type: none"> •Reference List •Technology Review •Addresses for Scientists •Links to Other Web Sites
---	--	---

Standardized information formats

<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 over 20 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 (published 20plex primers)
- ❖ Y STR mapped positions along chromosome

Example Y STR Fact Sheet from STRBase

www.cstl.nist.gov/biotech/strbase

We would like to collect variant alleles for Y STRs as they are discovered...

Allele	Set 1	Set 2	Set 3	Repeat Structure	Ref.
7	257	95	283	(TAT) ₇	368
8	349	98	286	(TAT) ₈	368
9	244	104	302	(TAT) ₉	368
10	349	107	305	(TAT) ₁₀	368
11	252	118	300	(TAT) ₁₁	368
12	255	113	311	(TAT) ₁₂	368
13	258	118	314	(TAT) ₁₃	368
14	365	118	317	(TAT) ₁₄	368
15	364	122	320	(TAT) ₁₅	368
16	367	118	323	(TAT) ₁₆	368

Summary of NIST Y Chromosome Work

- Development of new Y STR multiplex assays (**Y STR 20plex, etc.**)
- Evaluation of SNP typing methodologies and development of **Y SNP assays** involving primer extension and the SNaPshot kit
- Creation of a Y Chromosome Standard Reference Material (**SRM 2395**)
- Standardization of information on Y chromosome markers with internet accessibility (**STRBase**)

>30,000 Y chromosome allele calls generated to aid studies on optimal markers for U.S. populations

Publications from Our Group on Y Chromosome Assays

- Schoske, R., Butler, J.M., Vallone, P.M., Kline, M.C., Prinz, M., Redd, A.J., Hammer, M.F. (2001) Development of Y STR megaplex assays. *Proceedings of the Twelfth International Symposium on Human Identification 2001*, Promega Corporation. <http://www.promega.com/genetics/brochures/ym12proccontentsbutler.pdf>
- Butler, J.M., Schoske, R., Vallone, P.M., Kline, M.C., Redd, A.J., Hammer, M.F. (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int* 129: 10-24.
- Schoske, R., Vallone, P.M., Ruitberg, C.M., Butler, J.M. (2003) General strategy for the design of multiplex PCR assays involving Y STR markers. (2003)*Anal. Bioanal. Chem.*, 375: 333-343.
- Butler, J.M., Schoske, R., Vallone, P.M. (2003) Highly multiplexed assays for measuring polymorphisms on the Y-chromosome. *Progress in Forensic Genetics 9* (Brinkmann, B. and Carracedo, A., eds.), Elsevier Science: Amsterdam, The Netherlands, International Congress Series 1239, pp. 301-305.
- Schoske, R. (2003) The design, optimization and testing of Y chromosome short tandem repeat megaplexes. Ph.D. dissertation, American University, 270 pp.
- Butler, J.M. (2003) Recent developments in Y STR and Y SNP analysis. *Forensic Sci Rev.*, 15:91-111.
- Schoske R., Vallone, P.M., Kline, M.C., Redman, J.W., Butler, J.M. (2003) High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Sci Int.* in press.
- Vallone, P.M. and Butler, J.M. (2003) Y SNP typing of two U.S. populations using allele-specific hybridization and primer extension. *J. Forensic Sci.* in press.
- Butler, J.M. and Schoske, R. (2003) Forensic value of the multicopy Y-STR marker DYS464. *Progress in Forensic Genetics 10*, in press.
- Vallone, P.M. and Butler, J.M. (2003) Multiplexed assays for evaluation of Y-SNP markers in U.S. populations. *Progress in Forensic Genetics 10*, in press.

Acknowledgments

Funding:

Interagency Agreement between National Institute of Justice and NIST Office of Law Enforcement Standards

NIST Project Team:

Pete Vallone
Margaret Kline
Jan Redman
Rich Schoske (American U)
Dave Duewer

Collaborators:

Mike Hammer and Alan Redd (U. AZ)
for Y STR sequences

This presentation available as pdf file from
<http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm>