

NIST Human Identity Project Team – Leading the Way in Forensic DNA...

**Application of New Technologies
in Forensic Genetics**
John M. Butler, Ph.D.
 U.S. National Institute of Standards and Technology

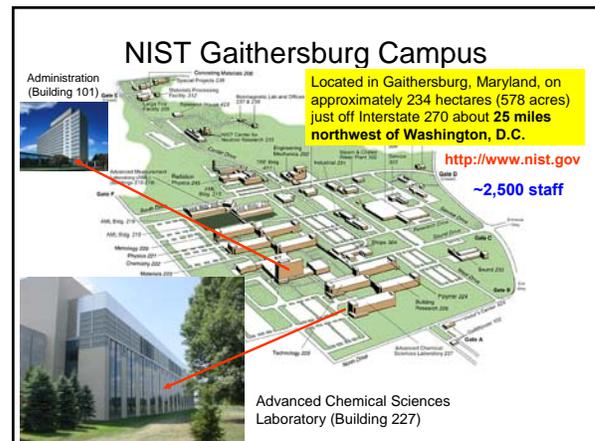
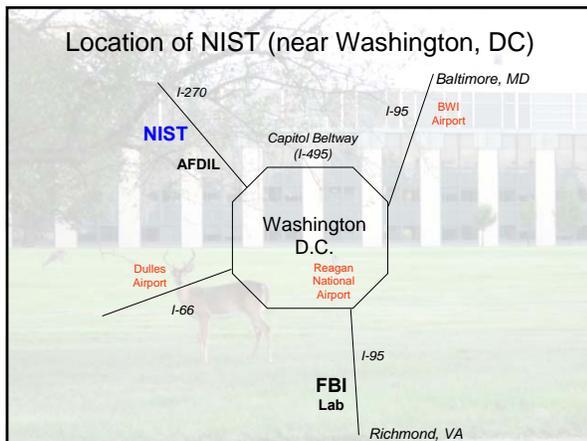
**Seminarium genetyków sądowych
(Seminar of Forensic Genetics)**
 April 20, 2007
 Krakow, Poland

NIST History and Mission

- National Institute of Standards and Technology (NIST) was created in 1901 as the National Bureau of Standards (NBS). The name was changed to NIST in 1988.
- NIST is part of the U.S. Department of Commerce with a mission to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life.
- NIST supplies over 1,300 Standard Reference Materials (SRMs) for industry, academia, and government use in calibration of measurements.

• NIST defines time for the U.S.

DNA typing standard



NIST Human Identity Project Team Laboratory Space (within Building 227)

Equipment List

- GeneAmp 9700 (3)
- GeneAmp 9600
- ABI 310
- ABI 3100 (now 3130xl)
- ABI 3130xl
- FM BIO III+
- ABI 7500 real-time PCR
- ABI 7000
- Agilent Bioanalyzer 2100
- Varian UV spec
- Varian HPLC
- Bruker TOF-MS
- Corbet robot
- 80 °C freezers

Room B257 (PCR Setup) Room B223 (Setup/Analysis)
 Room B261 (Post-PCR) Room B219 (DNA extraction)
 Room B219 (DNA extraction) Room A230 (ABI 3130xl)

National Institute of Justice

The Research, Development, and Evaluation Agency of the U.S. Department of Justice

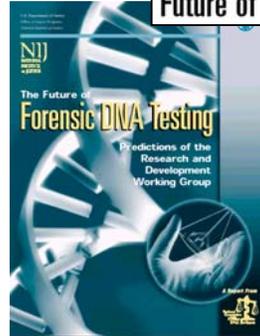
Current Areas of NIST Effort with Forensic DNA

- **Standards** <http://www.cstl.nist.gov/biotech/strbase>
 - Standard Reference Materials
 - Standard Information Resources (STRBase website)
 - Interlaboratory Studies
- **Technology**
 - Research programs in SNPs, miniSTRs, Y-STRs, mtDNA, qPCR
 - Assay and software development, expert system review
- **Training Materials**
 - Review articles and workshops on STRs, CE, validation
 - PowerPoint and pdf files available for download

Where Is the Future Going for DNA Technology That Can Be Applied to Forensic DNA Typing?

Constant state of evolution (like computers)

- Higher levels of multiplexes
- More rapid DNA separations
- Better data analysis software
- New DNA Markers



National Commission on the Future of DNA Evidence



• Report published in Nov 2000

• Asked to estimate where DNA testing would be 2, 5, and 10 years into the future

Conclusions

STR typing is here to stay for a few years because of DNA databases that have grown to contain millions of profiles

http://www.ojp.usdoj.gov/nij/pubs-sum/183697.htm

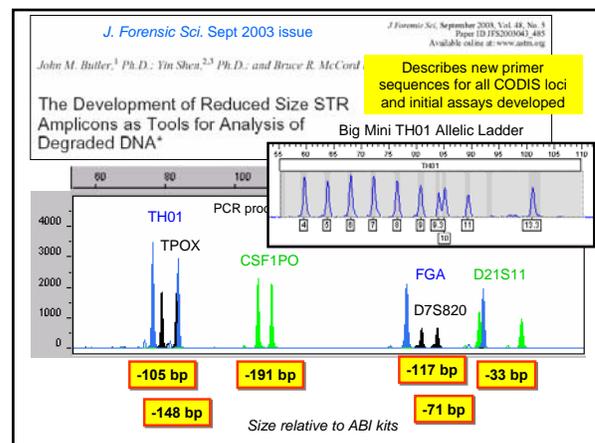
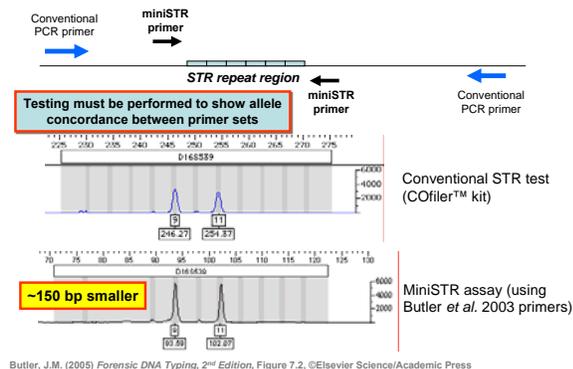
Status of Genetic Marker Systems Used in Forensic DNA Testing

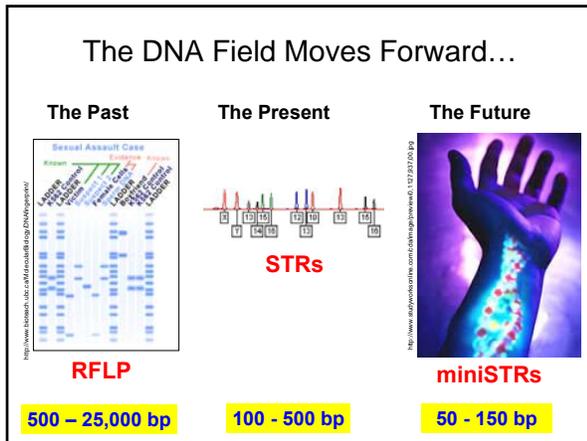
- **STRs** – widely used in national databases today
- **miniSTRs** – used in research and WTC; new MiniFiler kit just being released
- **mtDNA** – used in specialty labs for highly degraded specimens
- **Y-STRs** – growing use due to kits now available
- **SNPs** – research; likely to be limited in use

Technology: NIST Research Programs

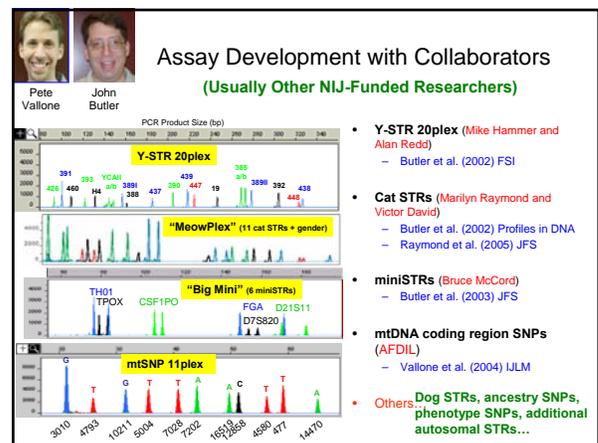
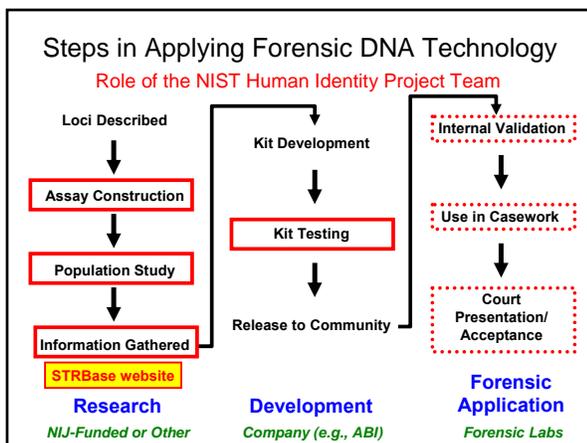
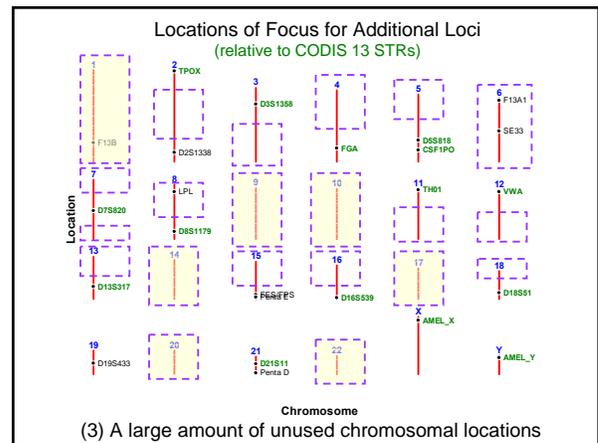
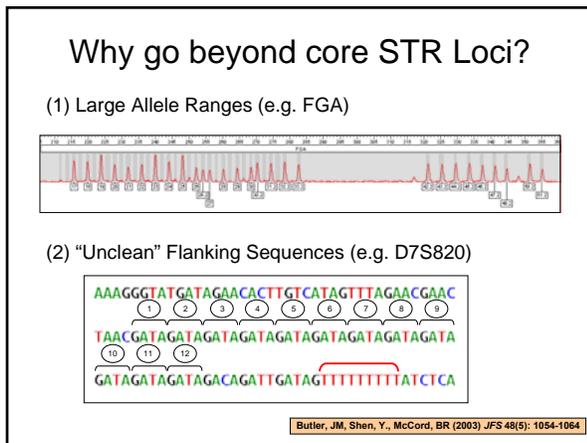
- miniSTRs
- Y-chromosome STRs
- mtDNA
- SNPs
- qPCR for DNA quantitation
- DNA stability studies
- Variant allele characterization and sequencing
- Software tools
- Expert System review
- Assay development with collaborators

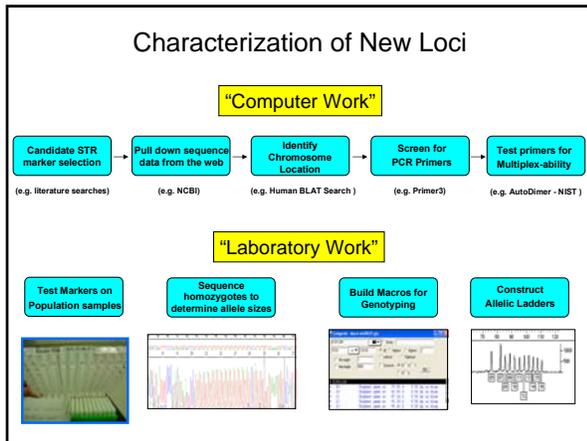
A miniSTR is a reduced size STR amplicon that enables higher recovery of information from degraded DNA samples





- ### How would additional loci be useful?
- Obtaining additional information with degraded DNA samples (miniSTRs)
 - Resolving common genotypes within populations
 - Kinship analysis





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**

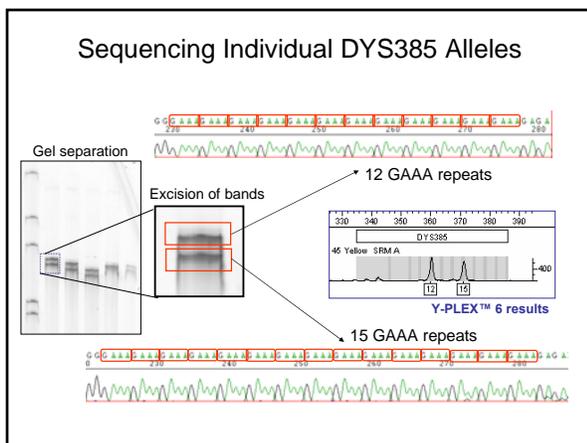
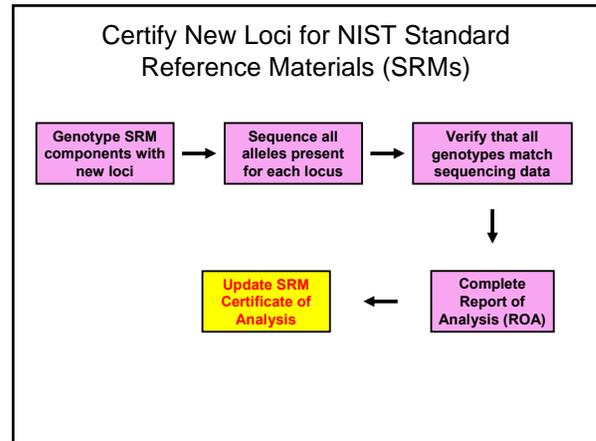
DNA extracted from whole blood (anonymous; self-identified ethnicities) received from Interstate Blood Bank (Memphis, TN) and Millennium Biotech Inc. (Ft. Lauderdale, FL)

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

- ### 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**



Standard Reference Materials

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

Traceable standards to ensure accurate measurements in our nation's crime laboratories

- SRM 2391b – CODIS STRs
- SRM 2392-I – mtDNA
- SRM 2395 – Y-STRs
- SRM 2372 – DNA quantitation

Helps meet DAB Std. 9.5 and ISO 17025

Working to update 2391b with new miniSTRs and 2395 with new Y-STRs

Calibration with SRMs enables confidence in comparisons of results between laboratories

Y-Chromosome Standard NIST SRM 2395



Human Y-Chromosome DNA Profiling Standard

- 5 male samples + 1 female sample (neg. control)
- 100 ng of each (50 µL at ~2 ng/µL)
- 22 Y STR markers sequenced
- 9 additional Y STR markers typed
- 42 Y SNPs typed with Marigen kit

Certified for all loci in commercial Y-STR kits:

Y-PLEX 6	SWGAM recommended loci
Y-PLEX 5	DYS19, DYS385 sub, DYS389/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439
Y-PLEX 12	
PowerPlex Y	

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

New STR Loci Information on STRBase

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



Value of Additional MiniSTR Loci

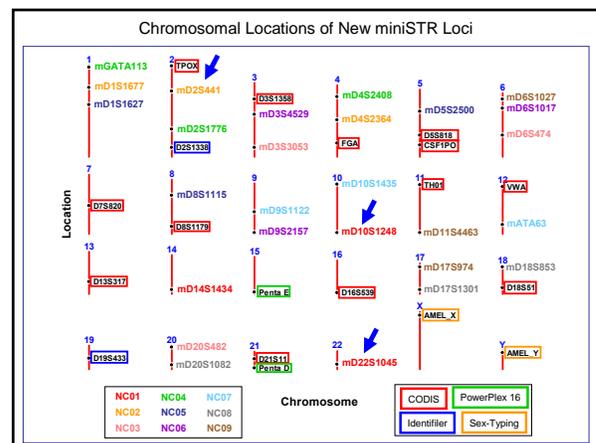
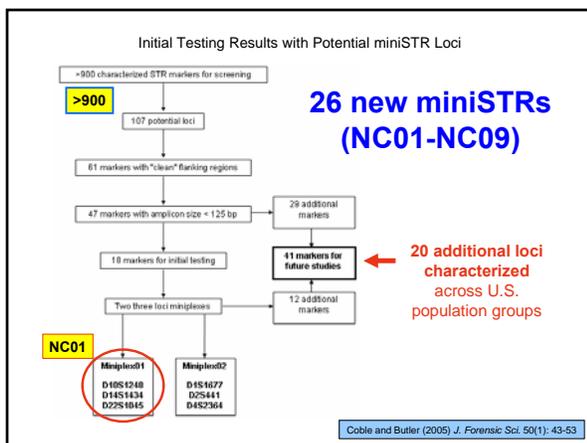
miniSTRs, MiniFiler, and Non-CODIS (NC) Loci



Becky Hill, Mike Coble, John Butler
 No longer at NIST (AFDIL Research Section Chief since April 2006)

- **MiniFiler:** concordance study with ABI (Hill *et al. J. Forensic Sci., in press*)
- **New Non-CODIS (NC) Loci:** 26 STR loci with allele sizes below 140 bp and good heterozygosities (above TPOX level) tested on NIST 665 U.S. population samples; physically unlinked to the 13 CODIS core loci; article submitted regarding primer sequences and locus characterization including population statistics
- **SRM 2391b components are being certified** through sequencing for D10S1248, D2S441, D22S1045; for reference purposes, genotypes for standard samples (9947A, 9948, 007, K562) are available on STRBase

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



Characterization of miniSTR D12ATA63

GenBank accession AC009771; positions 55,349..55,437

[FAM] – GAGCGAGACCCTGTCTCAAG
 GAAAGACATAGGATAGCAATT

Chr 12 106.825 Mb (12q23.3)

Trinucleotide [TAA][CAA] repeat
 76 - 106 bp
 Alleles 9 - 19

Heterozygosity Values
 U.S. Caucasian **0.842**
 African American **0.788**
 U.S. Hispanic **0.879**

D12ATA63 Allelic Ladder

Allele	Caucasian (N = 260)	African Am (N = 259)	Hispanic (N = 140)
9	--	--	0.0036
10	0.0019	0.0154	0.0036
11	0.1385	0.1525	0.1500
12	0.2154	0.1004	0.1786
13	0.0173	0.1564	0.0286
14	0.1615	0.3340	0.2214
15	0.0577	0.0772	0.0714
16	0.2981	0.1004	0.2643
17	0.0981	0.0521	0.0679
18	0.0096	0.0058	0.0071
19	0.0019	0.0058	0.0036

Comparison of heterozygosity values on 26 non-CODIS loci across the U.S. samples examined in this study.

Locus	N	Heterozygosity (Overall)	Rank	Ethnicity		
				African American	Caucasian	Hispanic
D9S2157	661	0.844	1	0.864	0.840	0.779
ATA63 (D12)	659	0.829	2	0.788	0.842	0.879
D10S1248 (NC01)	663	0.792	3	0.825	0.785	0.743
D22S1045 (NC01)	663	0.784	4	0.817	0.785	0.721
D2S441 (NC02)	660	0.774	5	0.798	0.780	0.721
D10S1435	663	0.766	6	0.798	0.770	0.700
D2S1776	654	0.763	7	0.740	0.801	0.734
D3S4529	660	0.761	8	0.752	0.723	0.829
D5S474	648	0.761	9	0.765	0.802	0.879
D5S2500	664	0.747	10	0.757	0.747	0.729
D1S1627	660	0.746	11	0.783	0.737	0.693
D1S1677 (NC02)	660	0.746	12	0.743	0.749	0.743
D6S1017	664	0.740	13	0.807	0.698	0.693
D3S3053	648	0.739	14	0.713	0.724	0.814
D9S1122	659	0.734	15	0.753	0.742	0.686
D17S974	664	0.732	16	0.757	0.702	0.743
D11S4463	664	0.730	17	0.780	0.676	0.743
D4S2408	654	0.722	18	0.752	0.709	0.691
D18S853	664	0.711	19	0.772	0.645	0.721
D20S1082	664	0.696	20	0.792	0.653	0.600
D14S1434 (NC01)	663	0.696	21	0.685	0.721	0.650
D20S482	648	0.691	22	0.673	0.689	0.729
GATA113 (D1)	654	0.668	23	0.673	0.632	0.727
D8S1115	664	0.663	24	0.629	0.660	0.729
D17S1301	664	0.649	25	0.626	0.717	0.554
D4S2364 (NC02)	660	0.511	26	0.385	0.551	0.664

European Labs Have Adopted the NIST-Developed NC miniSTRs

FSI (2006) 156(2): 242-244

Short communication

The evolution of DNA databases—Recommendations for new European STR loci

Peter Gill^{a,b}, Lyn Feraday^b, Niels Morling^c, Peter M. Schneider^d

^aForensic Science Service, Birmingham, UK
^bForensic Science Service, London, UK
^cDepartment of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Denmark
^dInstitute of Legal Medicine, University of Cologne, Germany

Received 25 May 2005; accepted 26 May 2005

...recommended that existing multiplexes are re-engineered to enable small amplicon detection, and that **three new mini-STR loci with alleles <130 bp (D10S1248, D14S1434 and D22S1045) are adopted as universal. This will increase the number of European standard Interpol loci from 7 to 10.**

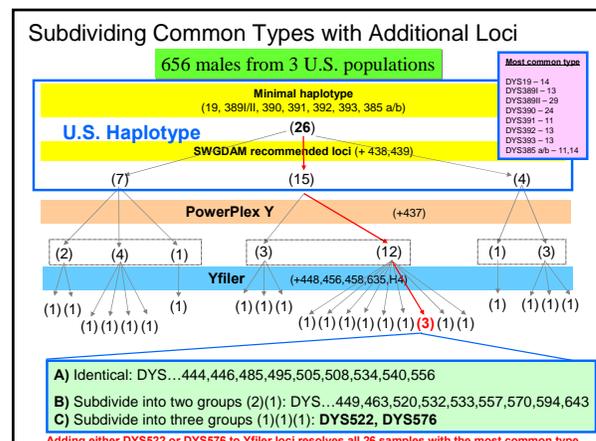
(D14 has been replaced with D2S441 from NC02)

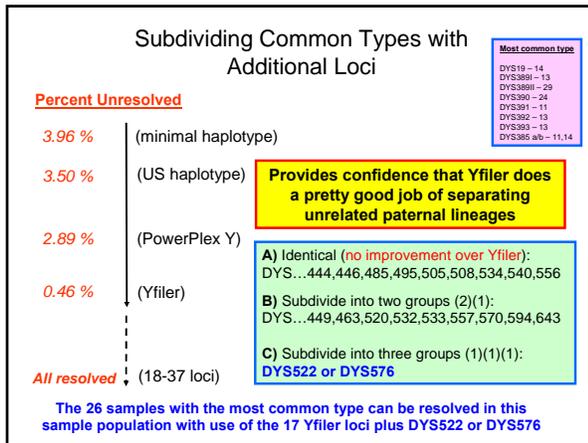
Going Beyond Commercial Y-STR Kits

- Most forensic DNA laboratories (certainly in the U.S.) will **only use commercially available kits** due to quality control issues
- Using these kits as a starting point, **are there additional loci that would be beneficial in separating samples with common types**, which could be advocated to companies for possible future adoption in Y-STR kits?
- Is it possible to regularly **resolve individuals from the same paternal lineage** (e.g., fathers and sons) if enough Y-STRs are examined?

Data Set Used to Examine Common Types

- Yfiler kit (**17 Y-STR loci**) run on all NIST male U.S. population samples
 - makes up ~20% of Applied Biosystems database
 - submitted to the YHRD
- Additional **20 Y-STR loci** run on full set of NIST population samples (and several less polymorphic ones only on subset of samples)
 - Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR Loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250-260.





# times haplotype observed	9
1	429
2	34
3	13
4	4
5	3
6	1
7	1
8	1
9	2
10	.
11	1
12	.
13	1
15	.
26	1

HD 0.996644
 %DC 0.748476
 # HT 491

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

429 of the 656 had a unique haplotype with the MHL loci, 34 sample haplotypes were observed twice in the sample set, 13 sample haplotypes were observed three times, etc.

Total = 656 samples

# times haplotype observed	9	11
1	429	486
2	34	33
3	13	10
4	4	6
5	3	1
6	1	1
7	1	2
8	1	.
9	2	.
10	.	1
11	1	.
12	.	.
13	1	.
15	.	1
26	1	.

HD 0.996644 0.998529
 %DC 0.748476 0.824695
 # HT 491 541

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

Total = 656 samples

# times haplotype observed	9	11	12
1	429	486	505
2	34	33	34
3	13	10	14
4	4	6	3
5	3	1	2
6	1	1	.
7	1	2	1
8	1	.	.
9	2	.	.
10	.	1	.
11	1	.	.
12	.	.	1
13	1	.	.
15	.	1	.
26	1	.	.

HD 0.996644 0.998529 0.999064
 %DC 0.748476 0.824695 0.853659
 # HT 491 541 560

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

Total = 656 samples

# times haplotype observed	9	11	12	17
1	429	486	505	626
2	34	33	34	12
3	13	10	14	2
4	4	6	3	.
5	3	1	2	.
6	1	1	.	.
7	1	2	1	.
8	1	.	.	.
9	2	.	.	.
10	.	1	.	.
11	1	.	.	.
12	.	.	1	.
13	1	.	.	.
15	.	1	.	.
26	1	.	.	.

HD 0.996644 0.998529 0.999064 0.999916
 %DC 0.748476 0.824695 0.853659 0.97561
 # HT 491 541 560 640

With the 17 loci in Yfiler across the 656 samples, there are 626 unique haplotypes, 12 haplotypes that were observed twice and 2 haplotypes that were observed three times

Total = 656 samples

# times haplotype observed	9	11	12	17	ALL 37
1	429	486	505	626	652
2	34	33	34	12	2
3	13	10	14	2	.
4	4	6	3	.	.
5	3	1	2	.	.
6	1	1	.	.	.
7	1	2	1	.	.
8	1
9	2
10	.	1	.	.	.
11	1
12	.	.	1	.	.
13	1
15	.	1	.	.	.
26	1

HD 0.996644 0.998529 0.999064 0.999916 0.996951
 %DC 0.748476 0.824695 0.853659 0.97561 0.996951
 # HT 491 541 560 640 654

When all 37 loci (Yfiler + 20 new loci) are run on 656 samples, only two haplotypes are observed twice

Total = 656 samples

Challenges of Defining Nomenclature for New Loci

- ### STR Repeat Nomenclature
- International Society of Forensic Genetics (ISFG) – *Int. J. Legal Med.* (1997) 110:175-176
- For sequences within genes, use the coding strand
 - For other sequences, select the first GenBank database entry or original literature description
 - Define the repeat sequence which will provide the largest number of consecutive repeats
 - If two sequences are repeated, include both motifs in determining the repeat number
 - **Microvariants:** should be designated by the number of complete repeats and the number of base pairs of the partial repeat separated by a decimal point (Int. J. Legal Med. 1994, 107:159-160) e.g. *TH01 allele 9.3*

Nomenclature Example (TH01)

Edwards *et al.* (1991) used **AATG** (adopted early on by Promega)
 Kimpton *et al.* (1993) used **TCAT** (Forensic Science Service) – **most widely used now**

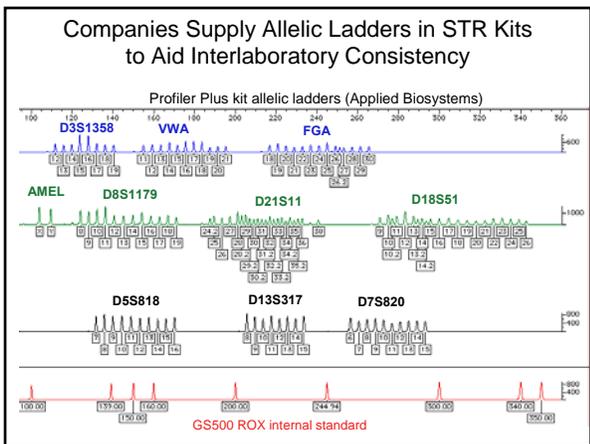
Differences in nomenclature can lead to confusion or even worse problems with database matches—standardization and consistency in use is essential...

Use of "TAGA" vs "GATA" results in a single repeat difference (Y-GATA-H4)

This simple difference has impacted the genetic genealogy community

Reference sequence: GenBank accession G42676 (submitted May 1999 by White et al.)

NIST SRM 2395 follows ISFG guidelines (for our primer pair): first adjacent repeat starting from 5' end is TAGA



D22S1045 (NC01)

Reverse strand called in Coble (2005) – TAA repeat CHANGE to ATT repeat in future work

GenBank Reference (AL022314) = 14 ATT repeats

If the T to C SNP is used, an extra 3 repeats could be added...

miniSTR Allele Nomenclature Changes
Due to further sequence analysis conducted in summer and fall of 2006

- **D10S1248 reduced by -1 repeat**
- **D22S1045 increased by +3 repeats**
- D2S441 kept the same
- **D14S1434 reduced by -4 repeats**
- **D1S1677 increased by +1 repeat**
- **D4S2364 reduced by -1 repeat**

Butler, J.M. and Coble, M.D. (2007) Authors' Response to Letter to Editor [regarding nomenclature for new miniSTR locus D10S1248]. *J. Forensic Sci.* 52(2): 494.

See also .../strbase/miniSTR.htm#Nomenclature_Errata

Y STR Allele Nomenclatures
SRM 2395 as an international standard should help resolve issues

DYS19

DYS439

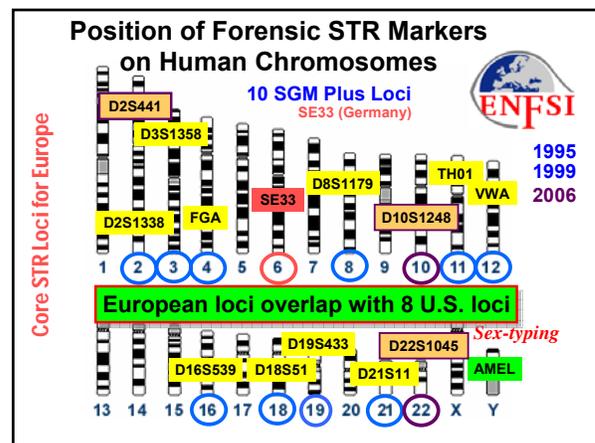
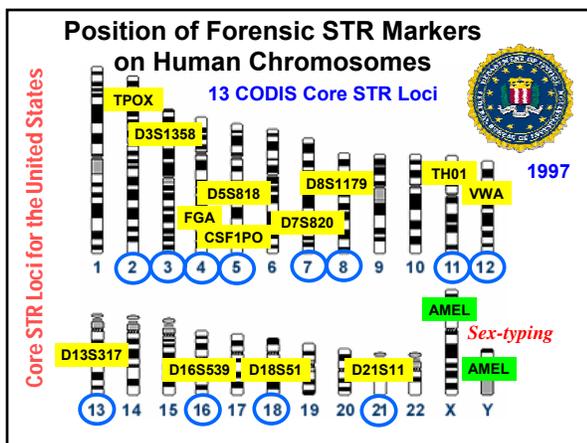
3 different nomenclatures have been published for DYS439

Uses of Additional Loci in the Forensic Community

- All STR loci
 - Missing persons investigations
 - Identification of mass disaster victims
 - Paternity testing
- Y-STR specific
 - Kinship analysis
 - Human migration and evolutionary studies
 - Historical and genealogical research

Publications on new loci

- Coble, M.D. and Butler, J.M. (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. *J. Forensic Sci.* 50: 43-53.
- Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250-260.
- Coble, M.D., Hill, C.R., Vallone, P.M., Butler, J.M. (2006) Characterization and performance of new miniSTR loci for typing degraded samples. *Progress in Forensic Genetics 11*, Elsevier Science: Amsterdam, The Netherlands, International Congress Series 1288, 504-506.
- Decker, A.E., Kline, M.C., Vallone, P.M., Butler, J.M. (2007) The impact of additional Y-STR loci on resolving common haplotypes and closely related individuals. *FSI Genetics*, in press.
- Hill, C.R., Coble, M.D., Butler, J.M. (2007) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *submitted*.



Review Article on Core STR Loci

J Forensic Sci, March 2006, Vol. 51, No. 2
 doi: 10.1111/j.1556-4029.2006.00066.x
 Available online at: www.blackwell-synergy.com

John M. Butler,¹ Ph.D.

Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing

Journal of Forensic Sciences 2006, 51(2): 253-265

- Reviews STR kits, genomic locations, mutation rates, potential genetic linkage, and known variant alleles for autosomal STR and Y-STR loci
- **Covers characteristics of 18 autosomal loci (13 core CODIS loci, D2, D19, Penta D, Penta E, SE33) and 11 SWGDAM-recommended Y-STR loci**

STRBase

Short Tandem Repeat DNA Internet Database

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

<p><u>General Information</u></p> <ul style="list-style-type: none"> •Intro to STRs (downloadable PowerPoint) •STR Fact Sheets •Sequence Information •Multiplex STR Kits •Variant Allele Reports •Training Slides 	<p><u>Forensic Interest Data</u></p> <ul style="list-style-type: none"> •FBI CODIS Core Loci •DAB Standards •NIST SRMs 2391 •Published PCR Primers •Y-Chromosome STRs •Population Data •Validation Studies •miniSTRs 	<p><u>Supplemental Info</u></p> <p style="border: 1px solid red; padding: 2px;">Reference List >2500</p> <ul style="list-style-type: none"> •Technology Review •Addresses for Scientists •Links to Other Web Sites •DNA Quantitation •mtDNA •New STRs •Forensic SNPs
---	--	---

New information is added regularly...

The Future

- More Robotics**
- Expert Systems**
- Animal & Plant DNA**
- Physical Characteristics**
- Ethnicity Estimation**

Acknowledgments

Funding from interagency agreement 2003-IJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards

NIST Human Identity Project Team – Leading the Way in Forensic DNA...

John Butler

Margaret Kline

Pete Vallone

Jan Redman

Amy Decker

Becky Hill

Dave Duewer

Past and Present Collaborators (also funded by NIJ):
 Mike Hammer and Alan Redd (U. AZ) for Y-chromosome studies
 Tom Parsons, Rebecca Just, Jodi Irwin (AFDIL) for mtDNA coding SNP work
 Sandy Calloway (Roche) for mtDNA LINEAR ARRAYS
 Bruce McCord and students (FL Int. U.) for miniSTR work
 Marilyn Raymond and Victor David (NCI-Frederick) for cat STR work
 Artie Eisenberg and John Planz (U. North Texas) for miniSTR testing on bones
 Murray Brilliant (U. AZ) for phenotype markers
 Ken Kidd (Yale U.) for SNP typing population samples
 Sree Kanithaswamy (UC Davis) for dog STR multiplex assay
 Tom Reid (DNA Diagnostics Center) for father-son samples

Disclaimer: Points of view are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Thank you for your attention...

Our team publications and presentations are available at:
<http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm>

Questions?

See also <http://www.dna.gov/research/nist>
<http://www.cstl.nist.gov/biotech/strbase>
john.butler@nist.gov